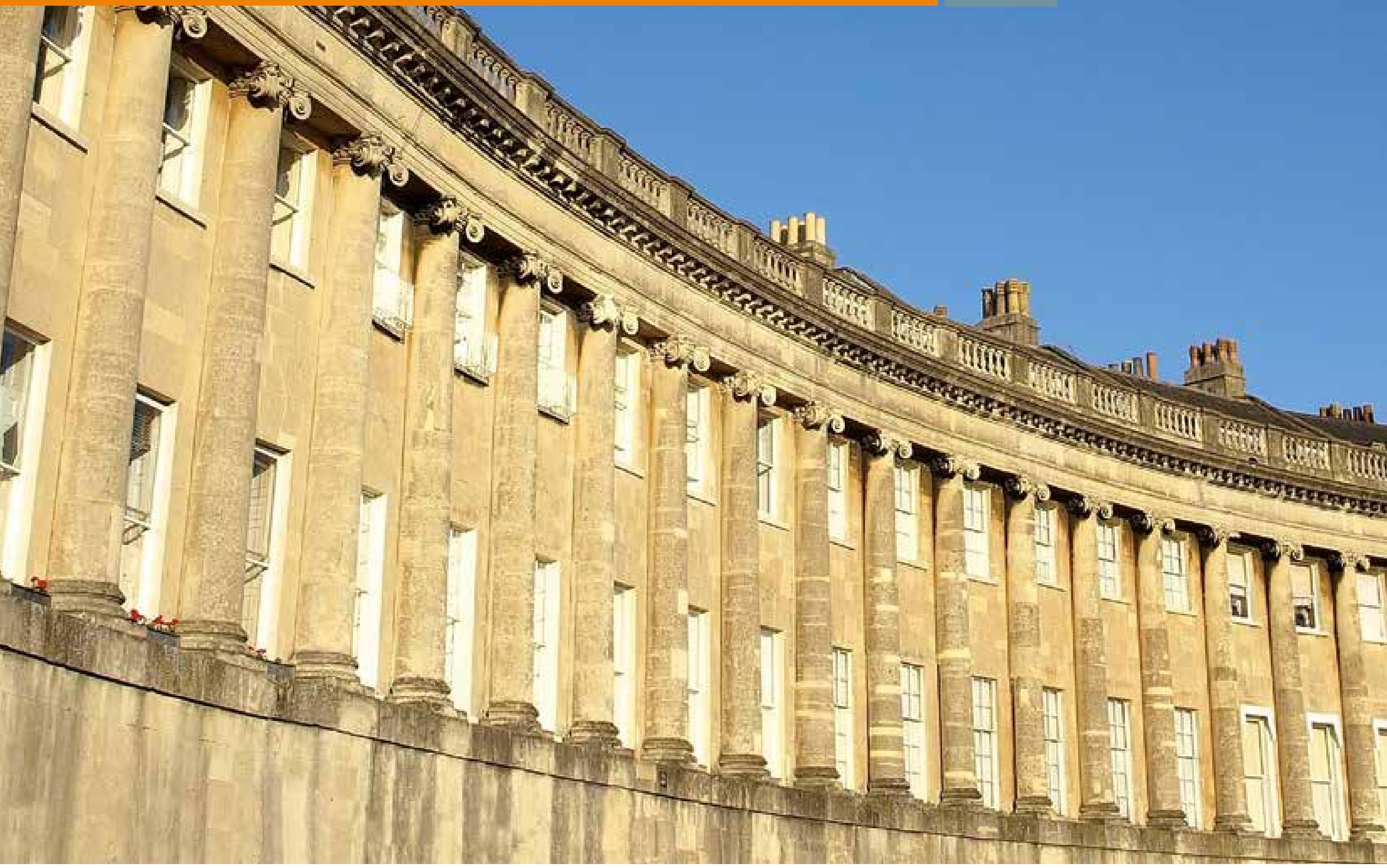
NATURALISTIC DECISION MAKING AND UNCERTAINTY

Proceedings of the 13th bi-annual international conference on Naturalistic Decision Making

Edited by Julie Gore and Paul Ward

NDM13



ndm
BATH 2017

20-23 June 2017
University of Bath, UK

University of Bath, Established 1966
University of Bath School of Management
University of Huddersfield
University of Sussex
E·S·R·C Economic & Social Research Council
CREST Centre for Research and Evidence on Security Threats
Research Councils UK
M2D Models to Decisions
VALCRI Visual Analytics for Sense-Making in Criminal Intelligence Analysis

NDM13 Naturalistic Decision Making and Uncertainty

Proceedings of the 13th International Conference on Naturalistic Decision Making

20-23 June 2017, Bath, UK

Edited by

Julie Gore and Paul Ward

The University of Bath,
School of Management Research Office,
Claverton Down,
Bath
BA2 7AY

ISBN 978-0-86197-194-7

WELCOME FROM THE CHAIRS

Welcome to NDM13 and to the historic city of Bath in the UK.

We are delighted to be hosting the 13th International Naturalistic Decision Making conference, which began 28 years ago in 1989. This year's conference theme – Decision Making Under Uncertainty – was intended to recapture the original focus of our self-organising community of practice, which commenced with a path breaking scientific curiosity for exploring cognition in challenging and complex environments. It also reflects our sponsors proposals and nascent lines of research inquiry.

The papers in the conference continue to advance many of the original NDM concepts, as well as research in contemporary NDM domains, including health, emergency services, and defence, to name but a few. Methodological refinements are also evident as researchers continue to innovate in their use of cognitive task analysis, visual analytics and technological integration. Attention to sensemaking, situation assessment and further unpacking of theoretical metacognitive developments are also highlighted alongside research in new areas, such as cybersecurity, intelligence analysis, sensitive policing, and deception.

In times of austerity it is fantastic to report that we received 70 submissions this year, all of which were rigorously reviewed. We had many more paper submission than we could accommodate, which has kept the quality at a high level. The final programme features 25 long papers, 25 posters, 3 panels, 8 invited/key note speakers, and 10 Phd papers.

We would like to thank our sponsors and supporters, which include the Centre for Research and Evidence on Security Threats (CREST), UK Economic and Social Research Council (ESRC), University of Bath's 50th Anniversary fund & School of Management, the Visual Analytics for Sensemaking in Criminal Intelligence Analysis (VALCRI) system, University of Huddersfield, University of Sussex, the US Army Research Laboratory (ARL), and two Decision Making Under Uncertainty Networks funded by Research Council United Kingdom (RCUK): Models to Decisions (M2D) and Challenging Radical Uncertainty in Society, Science and the Environment (CRUISSE). We also want to thank our international panel of reviewers and advisors whose knowledge of past NDM meetings was invaluable. Thanks also go to the team who brought this conference together: Tom Ormerod, Neville Stanton, William Wong, Chris Baber, Nikki Power, Joel Suss & Amanda Willmott.

While in Bath we hope you will take some time out to visit the Roman Baths and experience Austen country. And, of course, we hope you enjoy the conference.

On behalf of the NDM13 Committee, Welcome!

Julie Gore and Paul Ward
Co-Chairs, NDM13

NDM13 COMMITTEE

Co-Chairs

Julie Gore, Reader / Associate Professor in Organizational Psychology,
School of Management, University of Bath, UK
Paul Ward, Professor of Applied Cognitive Psychology,
School of Human and Health Sciences, University of Huddersfield, UK

Scientific Programme Chairs

Julie Gore, Reader / Associate Professor in Organizational Psychology,
School of Management, University of Bath, UK
Paul Ward, Professor of Applied Cognitive Psychology,
School of Human and Health Sciences, University of Huddersfield, UK
Thomas Ormerod, Professor of Psychology,
School of Psychology, University of Sussex, UK
Neville Stanton, Professor of Human Factors in Transport,
Engineering and the Environment, University of Southampton, UK
William Wong, Professor of Human-Computer Interaction,
School of Science and Technology, University of Middlesex, UK

Conference Secretariat

Amanda Willmott, Research Manager,
School of Management, University of Bath, UK

International Scientific Advisory Committee:

Emma Barrett, Research Fellow in Psychology,
Centre for Research & Evidence on Security Threats, University of Lancaster, UK
Gareth Conway, Principal Psychologist / Analyst,
Defence Science and Technology Laboratory, Ministry of Defence, UK
Stephen M. Fiore, Director of the Cognitive Science Lab,
University of Central Florida, USA
Simon Henderson,
Centre for Cyber Security & Information Systems, Cranfield University, UK
Robert R. Hoffman, Senior Research Scientist,
Florida Institute for Human and Machine Cognition, USA
Erik Hollnagel,
Center for Kvalitet, Institute of Regional Health Research, University of Southern Denmark
Rob Hutton, TriMetis Ltd Director / Research Scientist, UK
Gary A. Klein, Senior Scientist,
MacroCognition LLC, USA
Gary L. Klein, Prinicipal Scientist,
The MITRE Corporation, USA
Laura Militello, Senior Principal Scientist,
Applied Decision Science, Ohio, USA
Kathleen Mosier,
Department of Psychology, College of Science and Engineering, San Francisco State University, USA
Emilie Roth, Roth Cognitive Engineering, USA
Jan Maarten Schraagen, Professor of Cognitive Psychology,
University of Twente, Netherlands

Doctoral Consortium:

Nicola Power, Lecturer in Psychology and Security,
University of Lancaster, UK
Joel Suss, Assistant Professor of Psychology,
Department of Psychology, Wichita State University, USA
William Wong, Professor of Human-Computer Interaction,
School of Science and Technology, University of Middlesex, UK
Chris Baber, Department of Electronic, Electrical and Systems Engineering,
University of Birmingham, UK

Student Volunteers:

Matylda Gerber, Department of Computer Science, Middlesex University, UK

LIST OF REVIEWERS

We gratefully acknowledge the time and effort that our reviewers gave to help make this conference a success.

Chris Baber
Emma Barrett
Gareth Conway
Rhona Flin
Julie Gore
Simon Henderson
Robert Hoffman
Robert Hutton
Gary A. Klein
Gary L. Klein
Laura Militello
Kathy Mosier
Thomas Ormerod
Nicola Power
Jan Maarten Schraagen
Neville Stanton
Joel Suss
Paul Ward
William Wong

CONTENTS

Keynotes and Invited Speakers

Paper Presentations

Poster Presentations

Panels

Doctoral Consortium Presentations

KEYNOTE AND INVITED PRESENTATIONS

Expertise and Post-Truth

Harry COLLINS
Keynote Speaker

School of Social Science, Cardiff University, UK

ABSTRACT
This paper describes ‘Studies of Expertise and Experience’ (SEE) providing alternative insights to an NDM approach to expertise. The importance of socialisation, discourse and ‘interactional expertise’ will be examined as well as the implications of time to determine the substance and nature of the scientific consensus.

KEYWORDS
‘Studies of Expertise and Experience’ (SEE); post-truth

STUDIES OF EXPERTISE AND EXPERIENCE

Under the programme known as ‘Studies of Expertise and Experience’ (SEE), technical decisions are better if more weight is given to the views of those who ‘know what they are talking about’. A better decision does not mean that the ‘right’ decision has been made but to choose anything else – giving priority or equality to those who do not know what they are talking about – is dystopian. Knowing what you are talking about does not mean knowing the truth of the matter; it means having spent time studying the matter and making observations that pertain to it, usually in interaction with others, thus creating a domain of expertise.

Individuals acquire expertise through socialisation into a domain of expertise. Socialisation generally begins with deep immersion in the spoken discourse of the domain, thus acquiring interactional expertise, and, in the case of a practical domain, becoming a contributory expert by building up somatic tacit knowledge through sharing domain practices. Immersion in spoken discourse alone also leads to the acquisition of considerable tacit knowledge and can lead to an understanding of the practical domain which is good enough to make practical judgements indistinguishable from those made by practitioners themselves (as demonstrated by the ‘Imitation Game’). Were it not so, societies would not work.

Under this model, expertise is not defined according to whether its possessors know more true things than others but by whether the tacit knowledge of an expert domain has been acquired. Expertise is sometimes esoteric and sometimes ubiquitous. Ubiquitous domains include expertises needed to live in one’s society, such as native language speaking. *A new thought: democracy could be said to be ruled by experts at living in the society in question.* The model therefore avoids the problems associated with the different knowledges of the past and the future and the problem of disagreement among experts: experts may disagree violently about the truth of the matter, as they often do in the sciences, without any of them being less expert than the others – though, in the long term, some or all of them will turn out to be wrong. *It could also avoid the supposed problem of the clash between expertise and democracy.*

Domains of expertise can be large or small – from experts living in society to eccentric groups of hobbyists. These domains are embedded within one another and overlap in complex ways. This is the ‘fractal model’.
SEE is compatible with the minimal claim of the ‘Second Wave of Science Studies’: the truth of the matter generally takes longer to discover than is useful for political decision-making. SEE is even compatible with more radical claims such as that there isn’t an a-social truth of the matter. The approach sets truth on one side and settles for the *best* decisions rather than the *right* decisions.

Where it is relevant the best decision will take into account the current scientific consensus in respect of the natural or social world and will include relevant experience-based expertise. Science is favoured in these circumstances because its values overlap with the values of democracy and have more chance of resisting erosion by free-market capitalism than is the case for most other institutions (once more, utilitarian justifications are avoided). Under this model, sciences that are notably unsuccessful, such as econometric-forecasting or long-term weather forecasting are still favoured over tea-leaf reading, astrology, and the like even though these are domains of expertise.

A scientific consensus may seem to favour one policy decision rather than another but political decision-making is always a matter of politics and, so long as the substance and nature (e.g., strength) of the scientific consensus is presented honestly, politicians may act in opposition to it and take their chance with the electorate. In the technical part of the decision more weight will be given to science but, so long as it is dealt with openly and honestly, the technical decision is always subservient to the political decision. Thus technocracy and 'epistocracy' are rejected.

The substance and nature of a scientific consensus is a social fact: it may include input from domains of expertise which are primarily experience-based; it may depend on a sophisticated understanding of the organisation of science and its relationship to society; and it will depend on an estimate of the levels of agreement and disagreement among the experts. For these reasons it is best explored and reported by natural and social scientists working together; a committee or committees called 'The Owls', is proposed. To repeat, their job is not to make policy but to report on the scientific consensus.

Under SEE, time is needed to determine the substance and nature of the scientific consensus, to consider its policy implications, if any, and to decide whether those implications should be followed or rejected. Therefore, this model is opposed to all forms of populism and favours representative democracy or some such. It is, therefore, opposed to 'post-truth' and all that goes with it even though it does not reach for a utilitarian justification for science.

REFERENCES

Collins, H (2009). *Rethinking Expertise.* University of Chicago Press

Collins, H. (2010). *Tacit and Explicit Knowledge.* University of Chicago Press

Collins, H. & Evans, R. (2017). *Why Democracies Need Science.* Wiley

Intra-Operative Decision Making

Rhona FLIN
Key note Speaker

Aberdeen Business School, Robert Gordon University

ABSTRACT

The models and methods from Naturalistic Decision Making (NDM) research are now applied in all manner of work environments, extending far beyond their original development domains. The hospital operating theatre is one professional setting which had been rarely accessed by cognitive psychologists before the turn of the century. Surgeons report that most of their decision making is conducted in the pre-operative phase. This involves diagnosing the condition, deciding whether to operate, devising the procedure and estimating risk and time. However, surgeons also have to take intra-operative decisions, when conditions are not as anticipated or an unexpected event occurs. Emergency and trauma surgeons may have to operate without the benefit of a detailed planning phase. This presentation will examine what is known about surgeons' intra-operative decision making and will consider how NDM approaches are being used to understand and develop decision making skills at the cutting edge of surgery and emergency medicine.

KEYWORDS

Surgery; intra-operative; decision making; risk; situation awareness

REFERENCES

Flin, R., Youngson, G. & Yule S. (2015) (Eds.) Enhancing Surgical Performance. A Primer in Non-Technical Skills. London: CRC Press.

Pauley, K., Flin, R., Yule, S. & Youngson, G. (2011) Surgeons' intra-operative decision making and risk management. American Journal of Surgery, 202, 375-381.

Zilbert, N., Murnaghan, M., Gallinger, S., Regehr, G., & Moulton, C. (2015). Taking a chance or playing it safe: Reframing risk assessment within the surgeon's comfort zone. Annals of Surgery, 262, 253-259.

Information and Networks at War: Changing barriers of emergent warfare

David GALBREATH
Invited Speaker

Director, Centre for War and Technology,
Dean of the Faculty of Humanities & Social Sciences, University of Bath

ABSTRACT

This paper addresses the philosophical and empirical examinations of the changing character, if not nature, of warfare through the constitutionalising effect of technology. For instance, how has Information and Communication Technologies shaped doctrine, concepts and operations? The paper also examines the differing approaches to information warfare and how these heuristics are used by defence planners and combatant/operational commanders to think about emergent and future war. The aim of the paper is to set out a research programme on information warfare as it engages with, and in some cases, replaces kinetic warfare.

KEYWORDS

Technology; Information Warfare; Heuristics, Change

REFERENCES

Galbreath, D., 2015. Western European Armed Forces and the Modernisation Agenda. *In*: Eddie, P. and Rees, W., eds. *The Evolution of Military Power in the West and Asia.* Abingdon: Routledge, p. 79. (Routledge Global Security Studies)

Cognitive Skills Training

Gary KLEIN[a, b], Joseph BORDERS[a], Emily NEWSOME[a], Laura MILITELLO[a], and Helen Altman KLEIN[b]
[a]*ShadowBox LLC*
[b]*MacroCognition LLC*

ABSTRACT
This paper describes lessons we have learned about presenting cognitive skills training. We have used ShadowBox as our training approach, but the lessons apply regardless of specific techniques employed. We analyze key takeaways and lessons learned throughout the course of multiple ShadowBox projects. We explain how the original ShadowBox mission statement has evolved based on these lessons learned. Recommendations are offered for others who are engaged in cognitive skills training.

KEYWORDS
Learning and Training; Lessons Learned; Expertise; Decision Making.

INTRODUCTION

The purpose of this paper is to describe what our ShadowBox® team has learned about Cognitive Skills Training during the past few years. We appreciate that other researchers and practitioners have wrestled with these issues and described powerful methodologies — for example, Sherrie Gott and her team developed the PARI method for cognitive skills training (1988). It is beyond the scope of this paper to review the work of other cognitive training research programs. Our focus is on discoveries that emerged as we transitioned from research-based recommendations and demonstration projects to developing and delivering fielded cognitive skills training. We simply want to compile the lessons that we have painfully acquired using ShadowBox because our experiences may be useful for others who are engaged in training cognitive skills.

ShadowBox Training

ShadowBox is a way for people to see the world through the eyes of experts, without the experts being there. It is a scenario-based approach. The trainee is given a scenario, with decision points interspersed. Each decision point presents a small number of options. The decisions can be about which action to take, which cues to monitor most closely, which goals have the highest importance, etc. The trainee ranks the options from best to worst and writes the rationale for the ranking. As part of the training development, a small panel of experts also read the scenario, ranked the options, and provided their rationale. Their rankings and rationale statements have been synthesized so that once trainees complete a decision point they are shown what the experts ranked and why. Trainees are eager to match the expert ranking but the real learning occurs when they read the experts' reasons and appreciate what the experts have noticed. The final step is for the trainees to identify their biggest takeaways from that decision point — what have they learned from the experts.

We have applied ShadowBox to a variety of domains — with law enforcement, military, petrochemical operators, child protective services caseworkers, and helicopter rescue crews.

Our Initial Mission

When ShadowBox LLC was stood up 1 August 2014, the mission was very straightforward: Use the ShadowBox strategy to provide *cognitive skills training*, using *expert* feedback, and by building scenarios based on a front-end *Cognitive Task Analysis* (CTA). We developed an *electronic* version of ShadowBox to enable *individuals* to train on their own time. We evaluated training success in terms of the trainees' *match to the expert rankings*. And we achieved *quality control* of ShadowBox scenarios by carefully reviewing all materials generated by our clients.

If you like this mission statement, you shouldn't. It's a minefield. To our surprise, many of the key assertions in this mission statement turned out to be misleading and problematic. Part of the difficulty was terminology, and terminology counts if potential clients become unnecessarily confused or discouraged. But we also had more serious, substantial problems because in some ways we had the wrong mission.

Information for lead author: Gary Klein, Ph.D., Senior Scientist, MacroCognition LLC, 3601 Connecticut Ave NW, #110, Washington, DC 20008. Telephone: 937/238-8281. Email: gary@macrocognition.com

UNPLEASANT DISCOVERIES AND LESSONS LEARNED

Unpacking the Phrase "Cognitive Skills Training"

Cognitive

The term "cognitive" is a problem for us, and for the NDM community. It is jargon. Potential clients often don't know what it means. And we suspect most professionals would have some disagreements over its meaning. It is too often a confusing term.

Lessons: We still use the term "cognitive" because it distinguishes our approach from "procedural" or "rule-based," and some clients do resonate to it. However when we use it these days we are quick to unpack it, explaining that it covers the following activities: making decisions, making sense of situations, detecting and diagnosing problems, prioritizing and trading off goals, managing attention, anticipating future states, and performing workarounds. See Table 1 for definitions of each cognitive activity. Our clients can resonate to these kinds of outcomes, in a way they can't to "cognitive."

Table 1. Defining cognitive activities.

Cognitive activity	Definition
Decision making	Making accurate and timely decisions about courses of action to take.
Sensemaking	Quickly sizing up situations.
Problem detection and diagnosis	Noticing anomalies and spotting inconsistencies in data, as well as figuring out underlying problems.
Identifying tradeoffs and priorities	Identifying most important priorities within complex situations with competing demands.
Attention management	Recognizing and monitoring critical pieces of information or important situational features.
Anticipating future states	Forecasting what may happen in the future, as well as thinking of implications of decisions and problems.
Performing workarounds	Thinking beyond the scope of rules and procedures to decipher how to manage the situation.

We find even more success when we can provide some relevant examples of what these specific cognitive activities actually look like. For example, in our work with police we conducted a cognitive task analysis and used the interview materials to explain that, e.g., for decision making, a police officer might have to decide whether to pursue an assailant or stay with a victim to provide aid.

Skills

Next, we move on to the term "Skills" – what does that mean? It sounds very procedural, just the opposite of cognitive. There is danger in creating training to target a laundry list of skills and sub-skills (e.g., competencies) - they lead to disjointed and stove-piped training.

Lessons: Instead of trying to address a laundry list of skills, we seek to shift the trainee's mindset — to think more like the experts. We are more interested in developing expertise than in training specific skills.

The large research team performing the DARPA Good Strangers project[1] diligently assembled a list of motherhood virtues that would make warfighters more successful at interacting with civilians, e.g., showing respect, perspective taking, gaining rapport, showing empathy, etc. It was a long list without a clear focus. However, our CTA work had identified an overarching shift in mindset that seemed to organize all the more specific items: trying to get the civilians to trust you more at the end of the encounter than at the beginning (Klein, Klein, Borders, & Whitacre, 2015). So for the ShadowBox training we tried to move the Marines and soldiers from an authoritarian mindset to a trust-building mindset.

The lesson we have learned is that we want to help people develop richer mental models, more powerful mindsets, more tacit knowledge (i.e., knowledge that is difficult to put into words, such as perceptual discriminations, pattern recognition, recognition of familiarity, and detection of anomalies). The focus on mindsets can be more powerful and efficient than addressing a greater variety of skills. We find that addressing mindsets allows for more efficient training but the benefit goes beyond efficiency. Several of our clients are attracted to ShadowBox because they appreciate the need for radical shifts in the mindsets of their staff members, and ShadowBox is unique in the way it directly tackles mindset shifts.

[1] The official name was SSIM: Strategic Social Interaction Modules

We want each ShadowBox scenario to result in an “aha!” moment, a discovery stemming from a mindset shift and a revision of a mental model. On one occasion we have actually heard gasps from the group we were training as they suddenly appreciated how they needed to adjust their mindset and mental model. This discovery process is different from training specific skills. We are still learning how we might achieve these discoveries as we examine which scenarios promote an “aha!” reaction or a revised mental model.

However, these mindset shifts are not always straightforward. Take the challenge of shifting workers from a procedural to a problem solving mindset. We had to be careful not to present this shift as good (problem solving) versus bad (procedural). Workers still need to master the procedures, and the more entries in their playbook, the better. So the ideal is ‘problem solving + procedures,’ not ‘problem solving versus procedures.’ By not respecting the importance of procedures we were provoking resistence.

We are not just seeking to alter mindsets and initiate “aha” moments. ShadowBox training can help people build their tradecraft — acquire a more complete playbook of procedures for getting things done, as well as helping trainees gain a more nuanced understanding of how to adapt the procedures in their playbook.

Training

Now we move to “Training.” The notion of training seemed so straightforward until a potential client explained that his organization never did any training! Sure, they occasionally needed to bring new people up to speed. But for them, “training” meant formal training programs, lesson plans, platform instructors, and so forth. And they never did any of that. So the term “training” can be ambiguous and misleading. Then another client complained that we were describing ShadowBox as a training tool, whereas they also wanted a tool for practice and supervision.

Lessons: Appraise the needs and the capabilities of the clients to sort out the kind of solution the client needs and whether we can adapt ShadowBox to fit those needs. For example, do they have a system in place with formal lesson plans and instructors? Perhaps we should *cognitize* their existing training content, taking the scenarios they already have and injecting cognitive challenges.

One client in the petrochemical industry already had a full-mission simulator, so we expanded on the scenarios to emphasize the mindset shifts that trainees needed to make, working within the context of the simulation rather than our own software.

If facilitators are a limited commodity, perhaps the client needs a personalized, electronic version of the scenarios including the expert feedback. If trainees are far-flung, ShadowBox may be useful as a pedagogy for distance learning.

People think training is about learning rules and procedures. We see training as an opportunity to give people other experiences, and to provide an opportunity for them to have “aha” moments as trainees make discoveries and revise their mental models.

Another lesson we have learned is to be careful with clients who want to use ShadowBox to evaluate workers. If ShadowBox gets used for evaluation, it stops being effective for training because the workers will no longer enter into the experience with curiosity and an eagerness to explore.

The Problematic Term “Expert”

Next, consider the term “Expert.” No one likes this term. People we consider experts are uncomfortable being labeled as such. People who are comparing themselves to experts keep asking, “Who are these experts?” when the comparison ruffles them. In one study (Klein & Borders, 2016) we found that many of the military officers nominated as experts lacked the mindset we were trying to develop — they relied on an intimidation mindset in dealing with civilians. Selecting experts on the basis of rank or years of experience can be counter-productive if it perpetuates mindsets we are seeking to change. We had to more carefully vet the experts on the panel to ensure that they had been successful in gaining the cooperation of civilians in foreign countries and cultures.

We have known from the beginning that groups of experts don’t completely agree with one another, which is why we include the potential for a minority view. We make it clear that the experts are not perfect, and their rankings should not be considered as ground truth. Yet in taking this position we are raising questions about what it means to be an expert.

We would like to replace the term “experts” but haven’t yet found a suitable replacement. We’ve considered “respected practitioners,” “skilled practitioners,” “proficient practitioners,” and so forth. Perhaps instead of SME (Subject Matter Expert) we should use HRP (Highly Regarded Practitioner).

And, unfortunately, not everyone nominated to be on the panel of SMEs is actually very competent. In our Good Strangers project we struggled with so-called experts who were providing rankings and rationale that didn't square with being a good stranger — we had to discard a number of these because they showed no sign of being skilled at de-escalating situations and gaining trust. People are often nominated as experts because of years of experience, rather than their skill level.

We are also learning how to do a better job of synthesizing the expert feedback and describing it clearly and succinctly for the trainees because this is the window into the expert's head -- the way the expert sees the world as reflected in the scenario. Previously, we just bundled the different comments from the panel members without giving enough attention to the clarity and cohesion of the material.

Additionally, we have found it necessary to connect the expert feedback to consensus-based best practices from research and policy, especially in domains in which decisions are frequently subjective and depend on a practitioner's style or personal preference (e.g., child protective services, law enforcement). It is important to give experts more credibility in these domains, because it is easy for trainees to write the experts off by explaining away their decisions as a matter of preference or opinion.

Front-end Cognitive Task Analysis

The notion of a front-end CTA has often turned out to be impractical. It is too time-consuming and expensive. Few training departments can afford a front-end CTA, which can take several months to plan, conduct, and analyze the interviews.

Lessons: We are exploring ways to fold the CTA into the scenario construction process, basically using the simulation interview strategy described as part of Applied Cognitive Task Analysis (ACTA, Militello and Hutton, 1998; Klein and Militello, 2004)). The simulation interview presents the interviewee with a challenging incident, followed by questions about tough decisions, shifts in the way the situation was understood, critical cues, and so forth. We have also had success with a hybrid critical decision method that combines elements of ACTA and the Critcal Decision Method – knowledge audit interview (Borders & Klein, 2017). Why not use these streamlined CTA methods in conjunction with building ShadowBox scenarios? In this way, we can collect the cognitive data while at the same time constructing the scenarios.

Training Delivery

Our goal of using an electronic version of ShadowBox to allow individualized training is still active, but we found that many of our clients value the group discussions. And these can usually be achieved with paper-and-pen versions of ShadowBox. However, these group discussions create a need we hadn't anticipated — to train facilitators at each site.

Lesson: We developed a facilitator training program for social workers, and in a pilot project these facilitators have done very well. We are developing facilitator training for petrochemical plant controllers. One petrochemical plant is using the scenarios we created (and new ones they've created on their own) for training in group settings. They project the scenario to a group of trainees and at each decision point they use a clicker survey to poll the group. After each decision point the facilitator leads the discussion about their rankings and selections; after the scenario, they do an after action discussion and look at trends to see how the upset developed and what they could have done to prevent it (if possible).

We have also come to appreciate the importance of ensuring the quality and consistency of facilitation. Not all practitioners can effectively facilitate ShadowBox scenarios -- it requires curiosity, the ability to think on one's feet, and the willingness to challenge flawed beliefs in a non-confrontational way. Effective facilitators will stimulate fruitful discussions that generate new insights and build richer and more accurate mental models. We recommend careful vetting of facilitators, but are also exploring the use of scenario-specific facilitation guides that present key themes and indicators of mindsets to help ShadowBox facilitators.

We believe ShadowBox training is best suited for short, distributed sessions (ideally one but no more than two scenarios per session) over an extended period of time. The training scenarios are designed to introduce complex challenges and augment on-the-job experiences, without introducing the dangers often associated with such. Using ShadowBox, we can present a wide range of situations that they trainee may otherwise never experience. And through repeated exposure to the expert model in the form of expert feedback, trainees are encouraged to make new discoveries and restructure their own mental models. Unfortunately, in most of the evaluation studies we've conducted at this point, logistical constraints have forced us to introduce all of the scenarios, usually four and sometimes six, during one training session. This procedure is not recommended because each scenario provides a cognitive workout for the trainee, and completing more than one or two during a session can be exhausting and possibly limit insights and knowledge retention.

Evaluating Success

ShadowBox has a built-in evaluation measure, the match between the trainee rankings/selections to those of the expert panel. That measure still works for us. But our initial expectation that with more scenarios the trainees would match the experts more closely was wrong. Each scenario might have its own unique dynamics, and there is no reason to believe that the discoveries made on one scenario would translate to the next.

Lessons: For a fair comparison, we match two scenarios that revolve around the same issues and present one at the beginning of the training program and the other at the end — counterbalancing, of course. That way we can more powerfully determine how much the trainee has learned.

Some issues do cut across scenarios, such as shifts in mindset, so we do expect some improvement with practice, but even here we have learned that we need to fashion the scenarios and decision points to reflect the mindset shifts of interest. We have found this works best when situation-specific indicators of each mindset are articulated before developing decision options. For example, skilled panel operators possess an active mindset (versus a passive mindset) that drives continuous exploration and investigatation of the system and its interactions. Thus, when designing a process control scenario and corresponding decision points, we must clearly define how this active mindset is manifested in the context of the specific scneario. Decision point options should also accurately reflect the qualities of the mindeset, which will improve the validity of our evaluations and conclusions. Furthermore, we are now more careful about the distractor items for the decision points, using these to present flawed beliefs and to reflect mindsets we are trying to alter. In this way, ShadowBox can serve diagnostic purposes by illustrating the weaknesses in the trainees' mental models.

Many training directors want to go further than match to experts or mindset shifts — they want to see improvements in performance. And so do we. But we run into the problem that our clients cannot identify who is doing well or poorly; there are rarely any clear and objective, job-based performance indicators. So there is no easy way for us to demonstrate performance improvements. The best we have come up with is to gather supervisor ratings pre and post training, or for those with ShadowBox training and those who haven't yet received it.

Scenario Quality Control

We initially tried to ensure the quality of scenarios by reviewing all scenarios generated by our clients. We worried that if we let clients make up their own scenarios they might not generate very good ones and the ShadowBox program would get a poor reputation simply because of the low-quality scenarios produced by organizations with little background in cognitive skills training. Therefore, we decided that only scenarios developed with our team, or at least reviewed by our team, would count as ShadowBox.

This policy made a lot of sense from a quality control perspective. It made very little sense from a business perspective. Current clients were frustrated because they didn't want to be tied to us forever. Potential clients were turned off for the same reason. The impetus of ShadowBox is that it provides a workaround for the training bottleneck imposed by unavailable or limited subject matter experts, but our policy ironically made us into the bottleneck, needing to review every new scenario.

Lessons: We abandoned that policy and now encourage clients to build their own ShadowBox scenarios. We have also developed a training program, that we are continuing to refine, to teach clients how to generate effective scenarios. However, we have also gained a great deal of humility about the difficulty of crafting good scenarios. One trap we have sometimes stumbled into was to craft decision point options that made good sense to us because we were so familiar with the scenario, not realizing that trainees would interpret the option differently than we expected. We needed to pre-test the items.

CONCLUSION

We think we have learned a great deal by trying to implement ShadowBox training. This paper is only incidentally about the ShadowBox approach. The goal of this paper is to use our experiences to convey lessons about presenting cognitive skills training, regardless of the techniques employed.

Despite all the false starts, we are more enthusiastic about cognitive skills training than we were at the start. Our clients had not even considered cognitive issues prior to interacting with us. For them, training was about teaching rules and facts and procedures. The opportunity to address cognitive skills opens up possibilities that they find very exciting. Several use the phrase "game changer."

Our new mission statement is to use ShadowBox to *shift mindsets* using *'aha' moments* by presenting scenarios based on *tough cases.* We will see how long this version lasts.

ACKNOWLEDGMENTS

We thank our project sponsors highlighted in this article: Defense Advanced Research Projects Agency, California Peace Officer Standards and Training, The Center for Operator Performance, the Annie E. Casey Foundation, and Stottler-Henke Associates, Inc. The views expressed in this article are those of the authors alone, and do not necessarily reflect the views of our project sponsors.

REFERENCES

Borders, J., & Klein, G. (2017). The Critical Decision Audit: Blending the Critical Decision Method & the Knowledge Audit. Paper presented at the 13th International Conference on Naturalistic Decision Making, Bath, UK, 2017.

Klein, G., & Borders, J. (2016). The ShadowBox Approach to Cognitive Skills Training: An Empirical Evaluation. *Journal of Cognitive Engineering and Decision Making*, *10*(3), 268-280.

Klein, G., & Militello, L. (2004). The knowledge audit as a method for cognitive task analysis. In H. Montgomery, R. Lipshitz & B. Brehmer (Eds.), *How professionals make decisions*. Mahwah, NJ: Erlbaum.

Klein, G., Klein, H. A., Borders, J., & Whitacre J. C. (2015). Police and military as good strangers. *Journal of Occupational and Organizational Psychology*, *88*(2), 231-250.

Means, B. & Gott, S. (1988). Cognitive task analysis as a basis for tutor development: Articulating abstract knowledge representations. In J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), *Intelligent Tutoring Systems: Lessons learned* (pp. 35-37). Mahwah, NJ: Erlbaum.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618-1641.

The Evolution of Analysis; Changing Expectations and attitudes…

Esther MARTIN
Invited Speaker
West Midlands Police, UK

ABSTRACT

When I joined West Midlands Police in 1997 there were just 30 analysts and our role largely focused around assisting homicides and kidnaps, often simply to prepare visualisations of information collated by investigators. These numbers have more than doubled as the role has developed and the ‘ask’ has become more diverse. Both analysts and senior leaders find themselves in a new environment with wicked problems to ‘solve’, no quick fixes and a strange paradox in the technology era of information overload coupled with a knowledge deficit.

Throughout the last 20 years I have watched senior leaders struggle with this relatively new phenomenon of an analyst… a civilian that can unpick information and make recommendations for police action, and now policing also needs analysts to drive the organisational thinking and increase the understanding of why crime happens, analyse demands on policing and other agencies and identify solutions… but the journey has been turbulent, cultural norms and biases have had to be challenged, and analysts have had to build personal resilience and have confidence that their skillset can deliver creative and critical analysis.

The presentation will be delivered from a practitioner perspective. It will focus on examples from West Midlands Police where relevant to illustrate the narrative.

Further the presentation will address the technological constraints to analysis within law enforcement and the role that VALCRI and systems similar to this have in starting the creative process of crime analysis to create time and space for analysts to apply the range of techniques available to them to provide operational and strategic direction to senior investigators and executive teams

KEYWORDS

Police, Visual Methods, Criminal Intelligence Analysts

BIO SKETCH

Esther Martin is Head of Strategy and Direction, having previously held the leadership roles of Head of Strategic Intelligence and Head of Analysis at West Midlands Police. An Intelligence analyst by profession she has worked for law enforcement since 1997. Esther has responsibility for strategic direction and planning, organisational risk and learning, academic research, performance management and the professional development of criminal intelligence analysts.

Her current projects include redesigning force governance, the review and restructure of the WMP intelligence function and the introduction of an active citizen funding programme to build social capital. She is currently working with Middlesex University and a European consortium of academics and developers to design a Comparative Case Analysis and data visualisation tool for criminal intelligence analysis (VALCRI).

Integrating and Extending Models and Concepts of Macrocognition

Robert R. HOFFMAN
Invited Speaker

Institute for Human and Machine Cognition

ABSTRACT

Much has been accomplished with regard to the foundations of the science of macrocognition. First, we know there is a rich history to the approach, in the empirical psychology of Franz Brentano (late 1800s), Act Psychology (early 1900s), and certain works of more recent cognitivists, especially George Miller and Alan Newell. Second, we have robust methods for cognitive task and work analysis. Third, we have some very useful descriptive models of how cognition adapts to complexity, such as the Data/Frame model of sensemaking and the Flexecution model of re-planning. But more needs to be accomplished. Discussions of macrognnition refer to what has become known as the "cheese wheel." This is a roster of what are felt to be the most important macrocognitive processes, including sensemaking, mental projection to the future, collaborating, managing attention, and others. Obviously, this wheel diagram is really just a list, although it is useful since it is readily comprehensible to laypersons and potential sponsors. But the distinction between processes and functions has not been made entirely clear. In this presentation I will propose a more principled scheme for conceiving of the processes and functions. Furthermore, I propose two new descriptive models that are patterned after Data/Frame. One is the Re-projection model of causal reasoning. The second is the Re-grounding model of collaboration. Most important perhaps is the notion that the people who work in macrocognitive work systems have to (1) sensemake their observed and controlled world *and* (2) flexecute with regard to their actions with regard to their observed or controlled world, but also have to (3) sensemake and flexecute with regard to their technology, and (4) sensemake and flexecute with regard to their team, and even their organization. This permits the modular combination of the four core models, for use depending on the focus of the analysis. This scheme opens significant opportunities to richly and more completely describe the complexities of macrocognitive work systems, such as the dynamics of trusting. It is directly suggestive of a method for cognitive task analysis that goes well beyond previous methods for categorizing interview statements, and is also directly suggestive of requirements for information visualization and decision-aiding systems.

KEYWORDS

Data/Frame, sensemaking, Flexecution, mental projection to the future, teaming, collaboration, integrated macrocognitive model

REFERENCES

Klein, G. (2007, September/October). Flexecution as a paradigm for replanning, Part 1: *IEEE Intelligent Systems*, pp. 79-83.

Klein, G. (2007m, November/December). Flexecution, Part 2: Understanding and supporting flexible execution: *IEEE Intelligent Systems*, pp. 108-112.

Klein, G., Moon, B., and Hoffman, R.R. (2006, July/August). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 22-26.

Klein, G., Moon, B. and Hoffman, R.R. (2006, November/December). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, pp. 88-92.

Deception

Apollo ROBBINS
CREST/ESRC Invited Speaker

ABSTRACT
Using your experience and expertise to guide your attention to relevant information and to make sound inferences about it. Building accurate and comprehensive mental models to understand the situation you face. Thinking critically about potential next steps and their likely consequences. All important components of effective expert decision making in the real world, high-risk scenarios studied in Naturalistic Decision Making research. Yet we frequently underestimate how often and easily we are distracted, ignore important information, and come to erroneous conclusions - and so make unwise decisions. In security-relevant settings, sophisticated adversaries deliberately seek to exploit, manipulate, and undermine decision making - for instance, when planning military action, in information warfare, during acts of espionage, and in cyber-attacks.

Apollo Robbins is an internationally renowned performer, speaker, and consultant. A student of human nature, he is a pioneer in the application of misdirection to operational environments. Forbes has called him “an artful manipulator of awareness”. Apollo uses entertaining and educational acts of illusion - pick pocketing and sleight of hand - to demonstrate perception management and self-deception, helping us become more aware of how we perceive, how we understand, and how we draw our own conclusions.

Apollo will demonstrate how our decision-making can be undermined - through self-deception and through deliberate deception by others. He will explore innovative methods for raising awareness of the ways in which we are deceived, and help us understand what we can do to equip ourselves, our staff, and our organisations with effective defences.

Interpersonal Sensemaking in Law Enforcement

Paul J TAYLOR
Keynote Speaker
Department of Psychology, Lancaster University

ABSTRACT

Interpersonal sensemaking is the metacognitive process of understanding, predicting and responding to the actions and inferred beliefs of others. Current psychological models of this process differ on whether they emphasize theorizing—sensemaking as a function of a person's rule-based inferences—or simulation—sensemaking as a function of a person's ability to 'wear the other's shoes.' Using examples from studies of law enforcement interactions, this presentation will argue that effective interpersonal sensemaking combines both processes. Evidence from field studies show that experts who engage in adaptive, flexible theorizing educe cooperation and concessions from suspects. Evidence from experiments of cross-cultural interactions demonstrate the importance of simulation, but show how strategic behavior can enable other cultures to simulate effectively and thus educe greater cooperation. Recent research has utilized novel methodologies for examining the processes of interpersonal sensemaking and the resulting common ground that sensemaking facilitates. This presentation will review these methodologies and the emerging evidence that interpersonal sensemaking is principally a bottom-up process in which behavioral alignment causes rather than results from high-level cognitive alignment.

KEYWORDS

Law-enforcement; sensemaking; situation awareness

REFERENCES

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B, 367,* 2213-2223.

Giebels, E., Oostinga, M. S. D., Taylor, P. J., & Curtis, J. (2017). The cultural dimension of uncertainty avoidance impacts police-civilian interaction. *Law and Human Behavior, 41,* 98-102.

Taylor, P. J. (2002). A cylindrical model of communication behavior in crisis negotiations. *Human Communication Research, 28*, 7-48.

PAPER PRESENTATIONS

Measuring the Meta and Cognitive abilities of Air Defence Operators

Jade ADAMS-WHITE, Jacqueline M. WHEATCROFT, and Michael JUMP
The University of Liverpool
School of Engineering
Liverpool
L69 3GH
j.adams-white@liverpool.ac.uk

ABSTRACT

Introduction: This study aimed to understand more fully some factors that influence decision confidence and accuracy related to air defence. To investigate the metacognitive abilities of air defence operators a Within-Subjects Confidence-Accuracy (W-S C-A) measure was used. Specifically, therefore, this study investigated the impact of Decision Criticality (DC) and Task Stress (TS) on decision making, measures of confidence, accuracy, and the W-S C-A relation. Personality constructs, workload and situation awareness were also included. **Method:** Participants were allocated to either a high, moderate or low task stress condition. Each participant then took part in a computer generated air defence scenario where they were required to make various decisions and provide a confidence rating for each of those decisions. Confidence, accuracy and W-S C-A were calculated. **Results & Discussion:** DC impacted both on decision confidence and accuracy, with low DC increasing confidence in decisions and high DC increasing accuracy in decisions.

KEYWORDS

Decision Making; Command and Control; Situation Awareness/Situation Assessment; Military

INTRODUCTION

Air defence operators are an integral part of any warship. Operators must detect, locate and identify potential air threats, making complex and cognitively demanding decisions in dynamic environments. The aim of this paper is to introduce a metacognitive methodology to increase understanding of air defence decision making. Previous naturalistic decision making (NDM) research examining metacognition has used less numerically-based methods, such as think aloud protocols (Cohen, Freeman & Wolfe, 1996; Fyre & Wearing, 2013). However, more experimentally-based methods may be of benefit to NDM research (Lipshitz, Klein, Orasanu & Salas, 2001).The method proposed in this paper uses realistic decision-making scenarios and provides a combination of subjective measures of confidence alongside objective scores of accuracy to investigate the metacognitive abilities of air defence operators.

Decisions made in warfare are characteristically made under high levels of uncertainty and time pressure (Jenkins, Stanton, Salmon, Walker & Rafferty, 2010). In a naval warship, this is combined with the complex and knowledge-rich environment of a ship's Operations Room (OR). The OR is the focal point of the ship with significant amounts of incoming information from various data sources that must be processed and attended to by OR personnel in order to make both tactical and strategic war-fighting decisions. This study aims to examine metacognitive decision making, in light of both internal and external influences surrounding air defence.

The term metacognition refers to an awareness of ones' performance, and the ability and willingness to reflect on ones' thinking processes (Parker & Stone, 2014). It has been argued that metacognitive confidence should be included in the study of decision making because it is an important indicator of real-world outcomes (Jackson & Kleitman, 2014) and critical to performance (Rousseau, Tremblay, Banbury, Breton & Guitouni, 2010). Further, confidence in one's own ability plays an important role in the decision made (Griffin & Tversky, 1992) and assessments of confidence can be used to guide current and future decisions (Kepecs & Mainen, 2012). Ensuring confidence is correctly placed has important implications. For example, over (too much) confidence has been linked to underestimation of risk which could have a direct impact on the evaluation of future events (Lovallo &

Kahneman, 2003). However, it is not only how confident one is in a decision (i.e., decision confidence) but the corresponding accuracy (i.e., whether a response is correct or incorrect) of a decision that is relevant. Strong positive relationships between confidence and accuracy (i.e., the more confidence expressed in accurate decisions) are highly beneficial as they demonstrate an individual's ability to weight information and subsequent decisions appropriately (Stichman, 1962).

In light of this, metacognition sensitivity can be assessed by using decision confidence. Fleming and Lau (2014) argue that the relationship between decision confidence and accuracy can provide a quantitative measure of metacognition. Hence, one metacognitive measure which has been used to assess the relationship between confidence and accuracy in decision making is the Within-Subject's Confidence-Accuracy (W-S C-A) measure. W-S C-A has been used successfully in domains such as forensic, investigative and legal psychology (Wheatcroft & Woods, 2010; Wheatcroft, Kebbell & Wagstaff, 2004; Wheatcroft, Wagstaff & Manarin, 2015), perceptual tasks (Koriat, 2011), and general knowledge tasks (Buratti, Allwood & Kleitman, 2013). More specifically, W-S C-A is a measure of metacognitive sensitivity and has been defined as a "calculation which enables expression of individual confidence in each incorrect or correct response made" (Wheatcroft & Woods, 2010; p.195). Thus it can provide a method to assess individual awareness of the accuracy of decisions made and can also be used to assess group responses. Put simply, the method is able to calculate the statistical relationship between the levels of confidence individuals might place in responses given relative to their corresponding correctness. A positive relation between the two means that individuals are more confident in correct decisions than their incorrect counterparts. Whilst a subjective metacognitive measure, it has real and critical potential to affect the amount of resources applied to an action (Bingi, Turnipseed & Kasper, 1999) which are crucial in air defence environments.

Prior research has demonstrated external factors which may be influential to this relationship. For example, the difficulty of a decision (Wheatcroft, Wagstaff & Manarin, 2015) have been shown to impact on the W-S C-A relationship. Such research highlights the potential for the W-S C-A relationship measure to be beneficial in adding to understandings of the external factors which influence the decision maker – such as the criticality of the decision to be made and the level of stress experienced during a situation. Both these are crucial factors on board a warship as operators must be able to cope with varying levels of decision criticality and stressful environments effectively and respond accurately to the presenting situation.

Moreover, research is required to increase awareness of the individual differences that impact on air defence decision making and highlight the internal factors that influence those decisions to ensure that the decisions taken are effective. Individual differences are concerned with how individuals differ from one another and research has suggested that is plays a key role in decision making (Jackson & Kleitman, 2014). One particular individual difference which has been considered when assessing confidence and accuracy in decision making is personality. Personality is important to decision making as it can influence how people think, feel and behave (Roberts, 2009) between and within contexts. Similarly, this study is also interested in the role of decision style, ambiguity and decisiveness. For example, in terms of ambiguity, in the context of critical decision making, a low acceptance of uncertainty may be psychologically advantageous in that decisions may be made which are less influenced by the need to reduce uncomfortable feelings in complex circumstances and decision making contexts.

In summary, to begin to uncover some of the factors related to decision making in an OR air defence role and their implications on confidence, accuracy and W-S C- A, this study investigates the impact of task stress and decision criticality on confidence, accuracy and W-S C-A. Individual differences in personality and decision related tendencies are also considered. Additionally, this study aims to establish how W-S C-A fits into the wider measurements currently used in the human-machine interaction decision making literature such as Workload and Situational Awareness.

METHOD

Participants

60 participants were recruited through opportunity sampling from The University of Liverpool. The participants consisted of 30 females and 30 males with a mean age of 26, ranging from 18-27. None of the participants had any prior experience in naval warfare. The study received approval from the University of Liverpool's Institute of Psychology Health and Society Ethics Committee, and a favourable opinion from the Ministry of Defence Research Ethics Committee.

Design

A mixed measures quasi-experimental design was employed. 3 (Task Stress: Low, Moderate, High) X 3 (Decision Criticality: Low, Medium, High); with repeated measures for the Decision Criticality independent variable.

The independent variables (IV) were Group, Task Stress and Decision Criticality. The dependent variables (DV) were confidence, accuracy, W-S C-A, personality (NEO-PI, Costa & McCrae, 1992) decision tendencies

(Tolerance to ambiguity, Budner, 1961, Decision style Roets & Van Hiels, 2007), workload (NASA TLX, Hart, 2006) and situational awareness (SART, Taylor, 1990).

Materials

Decision logs

To ensure as high ecological validity as possible in a quasi-experimental design an air defence scenario was created with the guidance and assistance of subject matter experts (SMEs). The scenario depicts a realistic set of events using a Peace Enforcement (PE) environment. A series of events and associated event decision logs were also created and agreed by SMEs. The event decision logs specify three decision options of reasonable equivalence for each event presented to the operator. SMEs have agreed one option per decision made as the 'optimal/best' decision option given the current situation.

Computer Scenario

The visual display used as the stimulus for the experiment was created using VAPS XT (Virtual Avionics Prototyping Software) software. The screen depicted a pseudo-realistic radar screen which included an airlane, a No Fly Zone (NFZ), a coastline and a border. A textbox to display additional information to assist in the decision making and a timer which counted down from 20 seconds at each decision event was also included (see Figure 1). The algorithms used to animate the visual display symbols were created using Matlab/Simulink. The symbology used is as specified by APP-6c (NATO, 2008). Microsoft Movie Maker was used to edit the video (e.g., to apply timer).The display was verified as being sufficiently realistic by the SMEs.

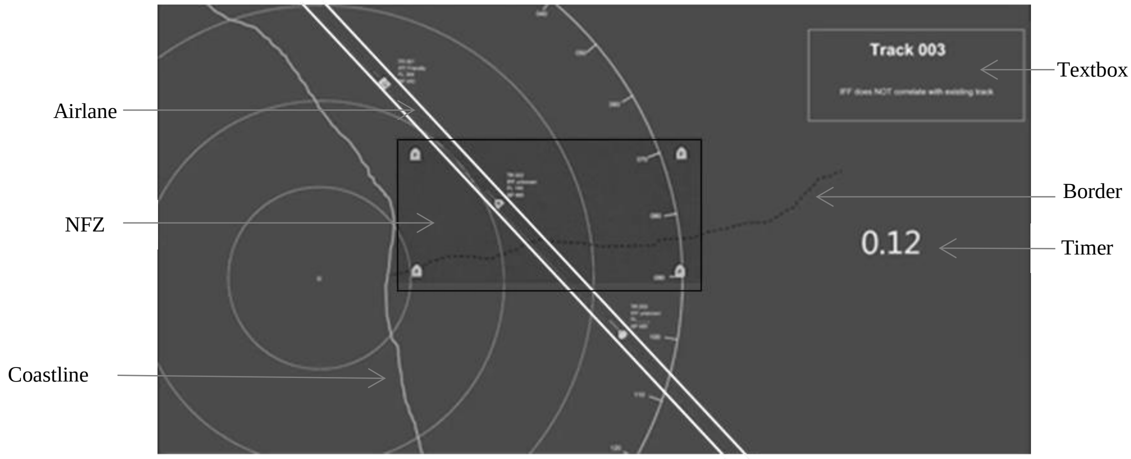


Figure 1. Visual display

Questionnaires

Situation Awareness Rating Technique (SART; Taylor 1990) was used to measure SA. To measure WL the NASA-TLX (Hart, 2006) will be utilised. NASA-TLX is a subjective workload assessment tool. Personality was assessed by the NEO-PIR (Costa & McCrae, 1992).

PROCEDURE

Participants were randomly allocated to a High, Moderate, or Low stress condition. Participants first completed participant demographic forms which collected data on age, gender and occupation. Participants were also asked to complete paper-based questionnaires to gauge the relevance of a number of measures across groups (e.g., general personality constructs, thinking and reasoning) where they may be relevant to particular questions. Following this, participants were provided with the task booklet to read. The task booklet provided the participants with information needed to assist them in the decision making task, including air defence terminology and symbols. Once they had read the booklet, participants undertook a practice trial. The questionnaire booklet presented three (3) separate decision options based on the events of the scenario. One choice was required to be selected by placing a tick by the option they believed to be the 'best option given the current situation'. Participants were then required to rate how confident they were in the options chosen on a Likert scale, where 0 = not at all confident to 5 = extremely confident. After twenty (20) seconds, the screen was blanked out to signal to the participants that the allocated decision time has ended. All participants then undertook the experimental air defence scenario, following the same procedure as described for the practice.

Thirty (30) decision events were presented during the experimental simulation. A decision event was defined as an occasion where a decision may need to be made by an operator. For example, an unknown data link track appears on the screen. The Decision Criticality was varied across the decision events presented (i.e., 10 high, 10 medium, and 10 low, DC) and the event occurrences varied depending on Task Stress condition. The scenario video ran for 20 minutes, 30 minutes, or 45 minutes for the High, Moderate and Low conditions, respectively. Once completed participants completed Situational Awareness and Workload questionnaires. Participants were fully debriefed to ensure each understood the nature of the study and given the opportunity to ask further questions.

RESULTS

A number of statistical analyses were performed on the data for Accuracy, Confidence and W-S C-A using Analysis of Variance (ANOVA).

Accuracy

A 3 x 3 mixed measures ANOVA was also carried out to assess the relationship between task stress and DC on individuals' decision accuracy.

A main effect of DC was found F (2,114) =16.71, $p<0.01$, $\eta^2_p = 0.23$. Bonferroni corrected post hoc tests showed participants were more accurate in high DC decisions (M=5.3 SD=2.0) than low DC decisions (M= 3.5, SD=2.0). Additionally, participants were more accurate in medium DC decisions (M=4.9, SD= 1.8) than low DC decisions both $p<0.01$.

However, no main effect of task stress F (2, 57) =2.03, $p>0.05$ and no interaction effect was observed F (4,114) =1.77, $p>0.05$ (see Table 1).

Table 1. Means and Standard Deviations for Accuracy

CONDITION	Overall	High DC	Medium DC	Low DC
High	13.8 (4.2)	5.2(2.3)	4.6(1.8)	4.1(2.2)
Moderate	12.3(3.6)	4.5(1.6)	4.7(1.8)	3.2(2.0)
Low	14.9(3.8)	6.1(1.7)	5.4(1.7)	3.4(1.7)
TOTAL	13.7(3.9)	5.3(2.0)	4.9(1.8)	3.5(2.0)

Standard Deviations are in parenthesis

Confidence

Again, a 3 x 3 mixed measures ANOVA was carried out to assess the impact of Task Stress and Decision Criticality (DC) on decision confidence. As Mauchly's test of sphericity was found to be significant, Greenhouse-Geisser was used.

A main effect of DC was found F (2, 88) =3.29, $p<.05$, $\eta 2p = 0.55$. A Bonferroni corrected post hoc test showed that participants were significantly more confident in Low DC decisions (M=37.3, SD=9.6) than medium DC decisions (M=35.1, SD= 7.2) $p=0.02$. No significant differences were found between high DC and low DC or medium DC and high DC.

No main effect of the Task Stress condition was found, F (2, 57) = 1.32, $p>0.05$) and no interaction effect was observed F (4, 88) =2.13, $p>0.05$ (see Table 2).

Table 2. Means and Standard Deviations for Confidence

CONDITION	Overall	High DC	Medium DC	Low DC
High	102(26.8)	32.5(11.9)	32.9(8.8)	36.7(13.7)
Moderate	111(21.7)	38.8(7.4)	36.0(7.0)	37.4(8.6)
Low	113(14.9)	38.6(5.6)	36.3(5.4)	37.7(5.1)
TOTAL	109(21.9)	36.6(9.1)	35.1(7.2)	37.3(9.6)

Standard Deviations are in parenthesis

W-S C-A

As before, a 3 x 3 mixed measures ANOVA was performed on the relationship between task stress and DC on individuals within-subjects confidence-accuracy (W-S C-A).

There was no main effect of DC F (2, 98) =0.62, $p>0.05$ and no main effect of task stress F (2, 49), 1.61, $p>0.05$ found. No interaction was observed F (4, 98) =0.61, $p>0.05$. An observation of the descriptive statistics shows that individual W-S C-A was found to be lowest between subjects in moderate task stress and within subjects in medium DC (see Table 3). W-S C-A was found to be highest between subjects in low task stress condition and within- subject in high DC. Overall W-S C-A scores were very low and not negative (M=0.02).

Table 3. Means and Standard Deviations for W-S C-A

CONDITION	Overall	High DC	Medium DC	Low DC
High	0.02 (0.2)	0.04(0.4)	-0.05(0.4)	0.05(0.3)
Moderate	-0.04 (0.7)	0.09(0.3)	-0.08(0.3)	-0.03(0.4)
Low	0.08 (0.2)	0.06(0.3)	0.10(0.2)	0.08(0.4)
TOTAL	0.02 (0.2)	0.06(0.3)	-0.00(0.3)	0.03(0.4)

Standard Deviations are in parenthesis

WORKLOAD (WL) AND SITUATIONAL AWARENESS (SA)

To assess the relationship between workload and SA a series of Pearson's correlations were calculated.

A significant negative relationship was found between SA and WL was found r= -0.53, p<.0.01. Higher levels of reported WL were related to lower feelings of SA.

A one way ANOVA was conducted to assess the relationship between SA and task stress. There was a significant effect of task stress condition on SA F (2, 57) = 6.44, p<0.01. Participants in the low task stress condition reported higher levels of subjective SA (M=21.4, SD=4.7) than participants in the high task stress (M=14.3, SD=5.4).p<0.01. See Table 4.

No significant relationship was found between WL and task stress F (2, 57) =3.00, p=0.06.

Table 4. Means and Standard Deviations for SA and WL

CONDITION	Overall SA	Overall WL
High	14.3 (5.4)	64(14.4)
Mod	18.7(8.3)	60.8(15.4)
Low	21.4(4.6)	53.3(12.4)
TOTAL	18.1(6.9)	59.4(14.1)

Standard Deviations are in parenthesis

Relationships between WL, SA and Accuracy, Confidence and W-S C-A

A significant moderate negative relationship was found between overall WL and confidence, r= -0.42, p< 0.01. As subjective measures of workload increased, confidence in decisions decreased.

A significant strong positive relationship was found between overall SA and confidence, r=0.63, p=<0.01. Higher scores in subjective SA were related to higher scores of confidence in decisions.

No significant relationships were found between SA, WL and W-S C-A. No significant relationships were found between SA and accuracy or WL and accuracy in decisions, all comparisons, p>.05. No significant relationship was found between-subjects confidence and accuracy, p>0.05.

Personality Constructs

Pearson's correlations were conducted to establish whether accuracy, confidence, and W-S C-A were related to the psychometric scores.

A significant negative relationship was found between Tolerance to Ambiguity (A) and Accuracy r = -0.34, p<0.01. Those who scored higher on the tolerance to ambiguity scale (i.e., less tolerant) were less accurate.

A significant negative relationship was also found between Decision Style and Accuracy r = -0.35, p<0.01. High scorers on the decision style scale were less accurate. Decision style explicitly probes the need for quick and unambiguous answers.

No other relationships were found to be significant, p>.05.

DISCUSSION

This study investigated operators' decision making during an air defence scenario. The aim was to assess the impact of Decision Criticality (DC) and Task Stress on measures of confidence, accuracy, and W-S C-A. Personality constructs, workload (WL) and situation awareness (SA) were also included.

The results from this study show that DC impacts on the accuracy of a decision. In accordance with the SME's validated correct decisions, participants made more accurate decisions in high DC than both low DC and medium DC, suggesting that accuracy increases with DC. The outcome supports previous literature in that participants make fewer errors in highly critical scenarios (Hanson, Bliss, Harden & Papelis, 2014). Furthermore, research has shown that performance in a task increases when participants find the task more important (Kliegel, Martin, McDaniel, & Einstein, 2004). These results may therefore indicate that individuals believed that the high DC decisions were more important in the context in which the task was operating. Additionally, the findings suggest that some measures of cognitive ability are not necessarily impaired when making critical decisions.

The work demonstrates that the criticality of the decisions did influence confidence. In particular, individuals were significantly more confident in low DC decisions than medium DC decisions. This lends some support to previous literature that has demonstrated that as difficulty increases confidence decreases (Chung & Monroe, 2000; Kebbell, Wagstaff & Covey, 1996). Nevertheless, it is the corresponding confidence which relates to an individual's awareness of the accuracy of these decisions that is important. W-S C-A remained unaffected, with no significant differences found in W-S C-A across Task Stress and DC. However, research has shown that training and experience improves calibration (Lichstenstein & Fischhoff, 1977). It would therefore be beneficial to conduct further experiments with those with particular cognitive skills, and relevant naval and air defence participants with appropriate experience.

An interesting finding emergent from this study is that although no significant differences were found in confidence, accuracy and W-S C-A in the different task stress conditions, individuals did report differences in subjective feelings of WL and SA across the different task stress conditions. In support of this finding, previous literature has found that stress impacts on WL and SA by reducing attentional resources and working memory (Endsley, 1995).

The broad personality constructs (i.e., neuroticism, extraversion, openness to experience, agreeableness, conscientiousness) using NEO-PIR were not related to confidence, accuracy, W-S-C-A, WL or SA. Nevertheless, this study did however find relationships existed with other cognitive constructs. Individuals who were less tolerant to ambiguity were less accurate in their decisions and high scorers on the decision style scale were also less accurate. The latter findings have implication for air defence personnel as the OR environment is both complex and at times ambiguous meaning operators need to be able to deal with such situations. Consequently, it has been demonstrated here that tolerance to ambiguity is a measurable skill which is required to increase the possibility of making accurate decisions in air defence.

CONCLUSIONS, IMPLICATIONS, LIMITATIONS AND FUTURE WORK

The aim of this paper was to introduce the methodology of W-S C-A to measure an element of the metacognitive abilities of air defence operators in a realistic decision making scenario. It has been previously argued that NDM research should use a mixture of measures to reduce the limitations of using a single methodology (Lipshiz *et al.*, 2001). Overall, the study combined objective measures alongside subjective measures in order to measure metacognitive abilities. As such, it is envisaged that the proposed method will provide a wider view of metacognition in critical decision making environments.

One key finding from this study is that DC had a significant impact on both decision accuracy and confidence. Although previous research has shown that DC plays a role in decision making in a business setting (Dunegan, Duchon & Barton, 1992), there has been a dearth of research into the effects of DC in complex decision making environments. However, as this work demonstrates, DC can contribute to both decision confidence and decision accuracy; hence, future work should certainly consider the impact of the criticality of decision.

No research is without its limitations. One such limitation was the use of novice participants rather than experts. Nevertheless, it is envisaged that future research will use experts to further validate the work. Further, participants were allocated 20 seconds to make a decision. It is possible the timeframe could affect the processes individuals use to make their decisions; though due to the nature of air defence decision making it is realistic to expect operators to be under some time pressure during the circumstances that surround these types of decisions.

In summary, this study provides a sound basis for future research, the aim of which is to investigate the internal and external factors that are involved in the meta and cognitive abilities in air defence decision making. The results from this study will later be compared with other populations, including military personnel, thus comparing experts and novices. The overall rationale is to uncover the skill sets which are of benefit to metacognitive abilities in air defence.

More broadly, the implications of this research include the potential for the approach and outcomes to be used to prioritize training, individual needs, and selection, in order to improve the effectiveness of decision making in air defence.

ACKNOWLEDGMENTS

This research was supported by the Defence Science Technology Laboratory contract no. DSTLX-1000092181. The authors would also like to thank Systems Engineering and Assessment Ltd. for their valuable help and support during the design of the study reported in this paper.

REFERENCES

Bingi, P., Turnipseed, D., & Kasper, G. (2001). The best laid plans of mice and men: The role of decision confidence in outcome success. *North American Journal of Psychology*, *3*(1), 91-108.

Budner, S. (1961). An Investigation of Intolerance to Ambiguity. *Nursing Research*, *10*(3), 179.

Buratti, S., Allwood, C. M., & Kleitman, S. (2013). First- and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition and Learning*, *8*(1), 79–102.

Chung, J., & Monroe, G. (2000). The effects of experience and task difficulty on accuracy and confidence assessments of auditors. *Accounting & Finance*, *40*(2), 135-151.

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Metarecognition in time-stressed decision making: Recognizing, critiquing, and correcting. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(2), 206-219.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.

Dunegan, K. J., Duchon, D., & Barton, S. L. (1992). Affect, risk, and decision criticality: Replication and extension in a business setting. *Organizational Behavior and Human Decision Processes*, *53*(3), 335-351.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Frye, L. M., & Wearing, A. J. (2013). A model of metacognition for bushfire fighters. In: Proceedings of the 11th international conference on naturalistic decision making, 22–24 May 2013, Marseille, France

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology*, *24*(3), 411-435.

Hanson, J. A., Bliss, J. P., Harden, J. W., & Papelis, Y. (2014). The effects of reliability and criticality on an IED interrogation task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2340-2344). Sage Publications.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage Publications.

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9*(1), 25-49.

Jenkins, D. P., Stanton, N. A., Salmon, P. M., Walker, G. H., & Rafferty, L. (2010). Using the decision-ladder to add a formative element to naturalistic decision-making research. *Intl. Journal of Human–Computer Interaction*, *26*(2-3), 132-146.

Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology*, *87*(4), 653-662.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322-1337.

Kliegel, M., Martin, M., McDaniel, M., & Einstein, G. (2004). Importance effects on performance in event-based prospective memory tasks. *Memory*, *12*(5), 553-561.

Koriat, A. (2011). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General*, *140*(1), 117.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In *Decision making and change in human affairs* (pp. 275-324). Springer Netherlands.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of behavioral decision making*, *14*(5), 331-352.

Lovallo, D., & Kahneman, D. (2003). Delusions of success. *Harvard business review*, *81*(7), 56-63.

NATO(2008).Department of defence interface standard Received from http://www.dtic.mil/doctrine//doctrine/other/ms_2525c.pdf.

Parker, A. M., & Stone, E. R. (2014). Identifying the effects of unjustified confidence versus overconfidence: Lessons learned from two analytic methods. *Journal of behavioral decision making*, *27*(2), 134-145.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137-145.

Roets, A., & Van Hiel, A. (2007). Separating ability from need: Clarifying the dimensional structure of the need for closure scale. *Personality and Social Psychology Bulletin*, *33*(2), 266-280.

Rousseau, R., Tremblay, S., Banbury, S., Breton, R., & Guitouni, A. (2010). The role of metacognition in the relationship between objective and subjective measures of situation awareness. *Theoretical Issues in Ergonomics Science*,*11*(1-2), 119-130.

Communications Personnel (No. BESRL-TRN-192). Army Behavoir and Systems Research Lab Arlington Va.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (AGARD-CP-478) (pp. 3/1–3/17). Neuilly Sur Seine, France: NATO - AGARD.

Wheatcroft, J. M., & Woods, S. (2010). Effectiveness of Witness Preparation and Cross-Examination Non-Directive and Directive Leading Question Styles on Witness Accuracy and Confidence. *The International Journal of Evidence & Proof*, *14*(3), 187-207.

Wheatcroft, J.M., Wagstaff, G.F & Kebbell, M.R. (2004). The influence of courtroom questioning style on actual and perceived eyewitness confidence and accuracy. *Legal & criminological psychology, 9*, 83-101.

Wheatcroft, J.M., Wagstaff, G.F., & Manarin, B. (online 3 April 2015). The influence of delay and item difficulty on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research* .

Elaborating the Frames of Data-Frame Theory

Simon ATTFIELD[a] and Chris BABER[b]
[a]*Middlesex University*
[b]*University of Birmingham*

ABSTRACT
As an explanation of sensemaking, data-frame theory has proven to be popular, influential and useful. Despite its strengths however, we propose some weaknesses in the way that the concept of a 'frame' could be interpreted. The weaknesses relate to a need to clearly contrast what we refer to as 'generic' vs. 'situation-specific' belief structures and the idea that multiple generic belief structures may be utilized in the construction of embedded situation-specific beliefs. Neither weakness is insurmountable, and we propose a model of sensemaking based on the idea of spreading activation through associative networks as a concept that provides a solution to this. We explore the application of this idea using the notion of activation to differentiate generic from situation specific beliefs.

KEYWORDS
Sensemaking.

INTRODUCTION

A good deal has been written about sensemaking, including various empirical studies and models. Notable accounts include Russell, Stefik, Pirolli and Card's (1993) Learning-Loop Complex; Weick's (1995) account of sensemaking in organisations; Pirolli and Card's model of sensemaking by intelligence analysts and Klein, Phillips, Rall and Peluso's (2007) Data-Frame theory. Each aims to capture something important about sensemaking with implications for how we might understand it better and perhaps how we might better enable it.

Inevitably, theories and methods for studying how people do sensemaking have tended to focus on analysis of post-hoc accounts of particular kinds of activity. While this has provided insight into the ways people develop understanding of complex situations, it may also limit it. And so we feel that there are gaps in the theoretical frameworks which result in methodological problems for analysis which are worthy of exploration. Any model of course, is a more or less useful approximation of what really goes on. Guided by the spirit philosophical pragmatism captured in Box's aphorism, "All models are wrong but some are useful" (1976) we aim to take a fresh look at sensemaking and develop on existing ways of thinking about it.

Our thinking references the data-frame model of sensemaking (Klein, Philips, Rall & Peluso, 2007), since this represents an important and influential model and one which attempts to engage with underlying cognitive processes of sensemaking, as we wish to. We will begin by developing an initial perspective on what sensemaking is, with an approach which has its roots in ordinary language philosophy. The approach is to develop a view of sensemaking based on considerations of the use of language surrounding sensemaking. From this we develop a view of sensemaking as seeking coherence between beliefs at different levels of abstraction about a domain. Following an overview of data-frame theory, we then introduce some empirical and theoretical reflections which may challenge our ways of thinking about sensemaking. From this we develop a new model which regards sensemaking as a process of developing interconnected networks of beliefs.

WHAT IS SENSEMAKING? - AN ORDINARY LANGUAGE APPROACH

To ask, 'What is sensemaking?' is to ask 'What is the meaning of the word 'sensemaking'? We consider this question from the perspective of ordinary language philosophy. In his Philosophical Investigations (2010) Wittgenstein said that, "For a large class of cases--though not for all--in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language." Thus, if wish to understand the meaning (of at least some) words, we must look at how the word functions in language. The motivation is to resist a temptation of an idealised yet nonetheless stipulated language, and to do this by examining the conceptual structures that are shown in language use. What this gives is a methodology for researching meaning through ordinary usage; an approach pioneered by ordinary language philosophers in the Oxford tradition such as Gilbert Ryle and J.L. Austin. For us, it also suggests that the concept of 'sense' has to be derived from the experiences and language of the people who we are studying and working with, rather than being imposed from some external source. While the

NDM community would find this point self-evident, it is worth pointing out because some of the representations used later in this paper could be misread as imposed (a posteriori by the researchers) rather than derived from the words, beliefs and experiences of subject matter experts.

Sensemaking seeks coherence

One immediate difficulty is that the verb ‘sensemaking’ is not part of common usage - it is itself a stipulated term. But we do talk about things ‘making sense’ and ‘not making sense’. We explore this with an example (which we treat here as data). In the book, 13 Things that Don’t Make Sense (Brooks, 2009), the author relays a discussion between Marie Curie, Hendrik Lorentz and Albert Einstein about how radioactive materials apparently defy the laws of conservation of energy and momentum. Brooks explains that, “Radioactivity was an anomaly; it didn’t *make sense*. The problem was eventually solved by the birth of quantum theory” (our italics). Without understanding the laws of conservation of energy and momentum or quantum theory, we can at least understand the initial problem as one in which there was a lack of consistency or coherence (we will stick with ‘coherence’) between beliefs about phenomena at two different levels of description. On the one hand, there is the description of the behaviour of radioactivity, whilst on the other hand, there are the laws of the conservation of energy and momentum. We can assume that the laws predict that radioactivity would behave in one way, whereas radioactivity actually behaves differently─hence, there is a lack of coherence between two beliefs. We can also assume that this lack of coherence is resolved once quantum theory is introduced.

Our next case is a thought experiment. Imagine you have a friend who you have learned enjoys museum visits - he visits many and talks about them frequently and you notice museum guides on his bookshelves. But, on a trip to a different city you see him turn down a trip to a local museum on the grounds that it sounds dull. This behaviour doesn’t make sense to you. Here, your theory about your friend is playing a similar role to the laws of the conservation of energy and momentum in the previous example and his behaviour is playing a similar role to the behaviour of radioactivity. The theory doesn’t predict the behaviour; it predicts something different. Again, the problem is one of coherence between beliefs at different levels of description. At one level we have your theory which is intended to subsume certain kinds of behaviour and at the other level an observation of behaviour which is inconsistent with it. There is a lack of coherence. We are likely to say that the behaviour doesn't make sense in the light of the current theory/model. It’s not until you realise that your friend’s interest is limited to natural history museums (the offer related to a cultural museum), that correspondence is resumed. His behaviour can now be said to ‘makes sense’, but notably, not because the behaviour changed.

We note that the two examples above deal with what we will refer to as retrospective sensemaking, that is to say, they relate to things that have been observed and the way that they can be accounted for or subsumed within prevailing theories that ought to apply. We suggest a working definition of ‘sensemaking’ as a process of reasoning about information to construct a view (belief, representation, understanding) about some situation and then testing this view for coherence with other beliefs that one holds at different levels of description. In brief, sensemaking is a quest for a coherent belief set. This is in contrast to any notion of truth. The question about whether something makes sense is not a question about whether it is true or not, although we assume that coherence here could be said to act as a proxy for truth, since incoherence can be assumed to indicate falsehood. However, in a more pragmatic sense, coherence provides the basis for selecting a course of action that feels appropriate for that situation. Thus, the burden of proof is much lighter than for a test of ‘truth’ per se.

DATA-FRAME THEORY

As an explanation of sensemaking, data-frame theory (Klein, Philips, Rall & Peluso, 2007; Klein, Moon and Hoffman, 2006a & b) has proven to be popular, influential and useful. It is a theory of how understanding (and also misunderstanding) of aspects of the world occurs, evolves and changes. It provides a high-level account of the types of process that sensemakers need to perform and illustrates how these processes interconnect. The model is informed by studies of military operations, navigation incidents, intensive care nurses, fire-fighters, weather forecasters and navy commanders (Klein, Philips, Rall & Peluso, 2007).

The theory distinguishes two kinds of entity which interact during sensemaking: data and frame. Data are aspects of the world which a sensemaker experiences. A frame is a representation in the mind of the sensemaker which stands for the situation, for example, a doctor’s beliefs about a patient’s medical condition, a pilot’s understanding of location and heading, or a warship captain’s beliefs about the position, heading and intent of an approaching aircraft. The frame acts as both interpretation and explanation of the data by accounting for it within a more integrated and complete picture. Importantly, the frame extends beyond the data, using background knowledge and expectations to fill gaps, or rather, it creates gaps in which data can be accommodated.

Klein, Philips, Rall & Peluso (2007) suggest that frames take a number of forms, including stories, explaining the chronological and causal relationships between events, maps describing locations and directions, or plans describing a sequence of actions. The term frame is intended as a synthesis of various concepts which have been used previously in accounts of human understanding, including frames by Goffman (1974), Minsky (1974), scripts by Schank and Abelson (1977) and schemata by Bartlett (1932), Neisser (1976), and Piaget (1952, 1955).

The theory presents sensemaking as a process of framing and re-framing in the light of data. As we encounter a situation a few key elements, or anchors, invoke a plausible frame as an interpretation of the situation. Active exploration guided by the frame then elaborates it or challenges it by revealing inconsistent data. By extending further than the data, a frame offers an economy on the data required for understanding, but also sets up expectations. Hence a frame can "direct" information search and in doing so reveal further data that changes the frame. An activated frame acts as an information filter, not only determining what information is subsequently sought, but also affecting what aspects of a situation will subsequently be noticed.

Important to the selection of a frame is the sensemaker's repertoire of frames and this underlies a distinction between experts and novices. It is argued that whilst experts and novices reason using the same procedures, experts have richer repertoires of frames that are better differentiated, allowing sense of a greater variety of situations, to be more precise about the situations, and to focus on fewer (but higher level) elements in a given situation (Lipshitz et al., 1997). Klein, Philips, Rall & Peluso (2007) also argue that frame activation depends upon the sensemaker's 'stance' including factors such as workload and motivation, and their current goals.

GENERIC AND SITUATION-SPECIFIC BELIEF STRUCTURES

The first issue we raise concerns a distinction we want to make between what we will refer to as generic and situation-specific belief structures. By generic belief structures, we mean something like a conceptual repertoire. Generic belief structures amount to a set of preformed and general 'understandings' that a sensemaker can bring to situations to help them make sense of them. For a doctor, this might include a set of medical conditions and their features; for a football coach it might include 'set plays' and strategies. This repertoire is key in determining what a sensemaker can 'see' in a situation and hence, in part, defines their expertise. It is a repertoire of categories (possibly with labels) and their associations that is an abstraction from experience, or perhaps acquired through training and acculturation. It is a theory about things that *can* happen.

The idea of situation-specific belief structures is of a set of non-generalised beliefs about a specific situation; it is a situation picture. Situation-specific belief structures occur through information or 'cues' from a situation combined with the application of generic belief structures to form an interpretation of a prevailing situation. This is 'comprehension' or interpretation with all the associations and expectations that the application of generic belief structures can invoke. A situation-specific belief structure is a theory about a prevailing situation with part of that theory being that the application is correct. In this sense, it has a truth value. It is an interpretation of a situation which can be true or false. In contrast, a generic belief structure is neither true or false in the same sense. A generic belief structure isn't a theory about a given situation but a theory about possible situations.

This distinction, of course, is not new. Data-frame theory has its roots in schema theory and here the distinction is made. Rummelhart (1980) (for example) makes it in his account of Schema Theory, describing schemata as generic data structures that represent generic concepts in memory, and distinguishing this from instantiations of schemas which occur when a situation is interpreted as an instance of a concept. He gives the example of buying something as an instantiation of a general purpose BUY schema. We prefer the term belief structure to schema since this avoids commitment to the idea of a simple mapping between situation-specific belief structures and generic belief structures. Situation-specific belief structures are complex. Below we develop the idea of multiple generic belief structures contributing to the construction of a given situation-specific belief structure. However, we argue that these two ideas are somewhat conflated in accounts of data-frame theory, and yet they are important to maintain as separate parts of a theory of sensemaking. One reason in particular that they are important to maintain, is that an account of the role of expectations and expertise in sensemaking depend upon it.

EMBEDDED BELIEF STRUCTURES

The distinction between generic and situation-specific belief structures helps us to develop the next idea, which is that the construction of situation-specific belief structures (or 'situation picture') can utilize multiple generic belief structures, applied opportunistically, to provide sense to a situation. Our suggestion arises from our own empirical studies and in part from experiences in using data-frame theory for analying qualitative protocols.

In one study, a group of university librarians performed an information task using a document information visualisation tool called INVISQUE (reported in Kodagoda et al., 2013). With INVISQUE, searches are submitted to a canvas style interface and results are presented as visual objects that resemble index cards. These appear in groups or 'clusters' that the user can re-group and sort as they wish, using both manual and automated sorting functions. The system could be described as a spatial hypertext (Marshall, Shipman & Coombs, 1994). The system was loaded with a set of ACM SIGCHI conference papers and participants were asked to identify at least three influential authors in the field of information visualization (for them, an unfamiliar domain). The purpose of the study was to observe how users might appropriate the tool to conduct the task.

In a later paper (Kodagoda et al. 2016), we used data from this study to experiment with using data-mining for inferring users' sensemaking actions from system log data. We coded think-aloud protocols recorded during the task with events described by the data-frame theory (*e.g.* connect data to frame, elaborating the frame, questioning the frame etc.). Our initial attempts at coding presented a difficulty. We had asked participants to draw conclusions about authors based on information about papers. In order to do this, participants had initially to draw conclusions about papers, using cues such as citation count and publication date. They then used assessments of 'influential papers' to make inferences about the standing of their authors.

During the analysis, the following two think-aloud extracts were coded as 'connecting data with frame':

> "I probably have a look at the overall paper. Ah interesting... has a heavy citation count."

> "Back to Stuart K. Card again, Stuart K. Card seems to have a reasonable distribution. He is clearly highly cited. So we going to go with Card."

The frames, however, are of different types. The first appears to be an 'influential paper' frame, and the second appears to be an 'influential author' frame. They relate to different kinds of generic belief structure. This is not to say that one might not support inferences about the other, but that they correspond to different concepts.

In a study of military signals intelligence (Attfield et al., 2015; Wheat, Attfield & Fields, 2016), analysts were required to make inferences about the identity and locations of military units based on intercepted radio communications. They used features such as call frequency to make inferences about the radio model and combined this with communication content to make inferences about the kind of unit communicating. From this they were able to make inferences about their regiments and divisions. Each step appeared to be a sensemaking exercise in itself exploiting background knowledge (*i.e.* generic belief structures), and at each step the sense that was made provided a cue for the next, ultimately contributing to a complex situation picture (or situation-specific belief structure). Wheat, Attfield & Fields (2016) analyse these chains as 'inference trajectories', and consider the overall situation picture as a meta-frame consisting of a number of sub-frames.

In a study of corporate lawyers conducting e-discovery investigations on large email collections, Attfield & Blandford (2011) reported that the lawyers mapped out an investigated domain using extensive chronologies constituted out of episodes and events each of which may draw on generic belief structures at different levels of scale or granularity. These could range from the events and activities involved in an investigated company bidding for a contract to individual meetings with arrangement discussions, participants, locations and follow-up actions. Pennington and Hastie's 'story model' (1981) makes similar observations of jurors making sense of evidence in criminal cases by building situation-specific belief structures in narrative with these utilizing episode schemas. Selveraj, Attfield, Passmore & Wong, (2016) describe how a group of police crime analysts used 'think-steps' - a series of extensible templates that they use to decompose cases of different types (such as people trafficking and murder) into elements, and how these provided a structure for storing and visually representing data, generating requests for information, focussing research, structuring mental simulation, and reporting.

Finally, in work on intelligence analysis (Baber et al., 2015; 2016), suggested that analysts tend to alternate between broad and narrow focus, *i.e*., looking at several topics and then narrowing to a smaller number. Similar effects have been observed by Elm et al. (2005) and Roth et al. (2010). For this paper, the argument is that the broad/narrow focus represents effort toward the development of situation-specific beliefs (particularly through the use of 'working representations' that people construct to visualise the links between information). Coherence testing of situation-specific beliefs arises through the development of explanatory stories which allow generic beliefs to be applied to scrutinise the expectations and assumptions that the situation-specific beliefs imply.

The idea of comprehension as involving multiple schemas is described once again by Rummelhart (1980) who describes something akin to the paradox of the hermeneutic circle in the simultaneous interpretation of parts and wholes of objects. He explains this using the example of the interpretation of an image of a face. In the image there are marks that represent a nose, some lips, an eye, an ear etc., only they cannot be interpreted independent of the interpretation of face. Conversely, the interpretation of a face cannot occur without a certain interpretation of the parts. He points out that we presumably have a schema for each component (nose, lips, eye, ear etc.) and a schema for the whole (face) and that evidence for nose contributes to evidence for face and vice versa, and that these are all have a role through processes of inference in the interpretation of the picture.

Data-frame theory seems to say little about how frames might embed within each other in the construction of sense and how this might be a helpful way to think about the sensemaking process. Possibly, an explanation lies in the type of sensemaking situation that motivated that theory (*i.e.* military operations, navigation incidents, intensive care nurses, fire-fighters, weather forecasters and navy commanders) which we assume are more situated in action and time-critical. In more slower paced sensemaking tasks, such as intelligence analysis, e-discovery and crime analysis (*i.e.* the ones that motivated the issue here), sensemaking presumably extends over longer periods and involves greater use of external representations as aids for memory and collaboration, and therefore we assume, the construction of more complex situation-specific belief structures. We think it is helpful, in understanding these kinds of sensemaking scenario to think about situation specific belief structures as having the potential to instantiate or utilize multiple generic belief structure in their construction.

AN ASSOCIATIVE MODEL OF SENSEMAKING

Following these reflections, we develop our case further by proposing a way of understanding sensemaking as the application of associative networks. If one assumes beliefs to be held in a structure in which associations between different belief objects can be usefully represented as links between nodes, then one can assume that activation of one node could lead to activation which spreads through the network. This raises several questions which are relevant to our discussion. For example, how are nodes defined and how are they linked?

How is the network structured and where is the activation spreading?

In the original conception of spreading activation (*e.g.*, Collins and Loftus, 1975), 'knowledge' was described as stored in a semantic network. This provided a clearly defined structure in which properties of a category could be decomposed into subcategories. For example, <animal><mammal><dog><retriever><golden>. This implies that the 'meaning' of each word in the category could be related to other words by the semantic properties of the words themselves. The represented propositions were analytic, in the sense that they were true in virtue of meaning. When one of the words in the network is activated, *i.e.*, when its level of excitation rises from resting level because it has been read, heard or spoken, this rise in excitation spreads to words connected to it. The spread is limited by distance, so closer words receive higher excitation than words which are more distant.

Amongst the initial successes of this approach was a plausible explanation of phenomena such as semantic priming effects (in which seeing or hearing related words would reduce the recognition time of subsequent words). The semantic structure was replaced by the late 1970s with forms which were defined by synthetic conceptual relations *i.e.* things that just happen to be the case. This is the basis of the concept map and is illustrated by figure 1 (constructed using the CMap tool). This shows concepts related to a 'Point of Entry' (or means of breaking into a house) for a domestic burglary and could be a possible concept map held by a Scene of Crimes Officer attending a crime scene. In concept networks, activation can still be assumed to spread between connected nodes. Thus, in figure 1, if an <openable> point of entry has been <broken / smashed>, then there might be <blood, DNA, etc.> that can be recovered as <evidence>.

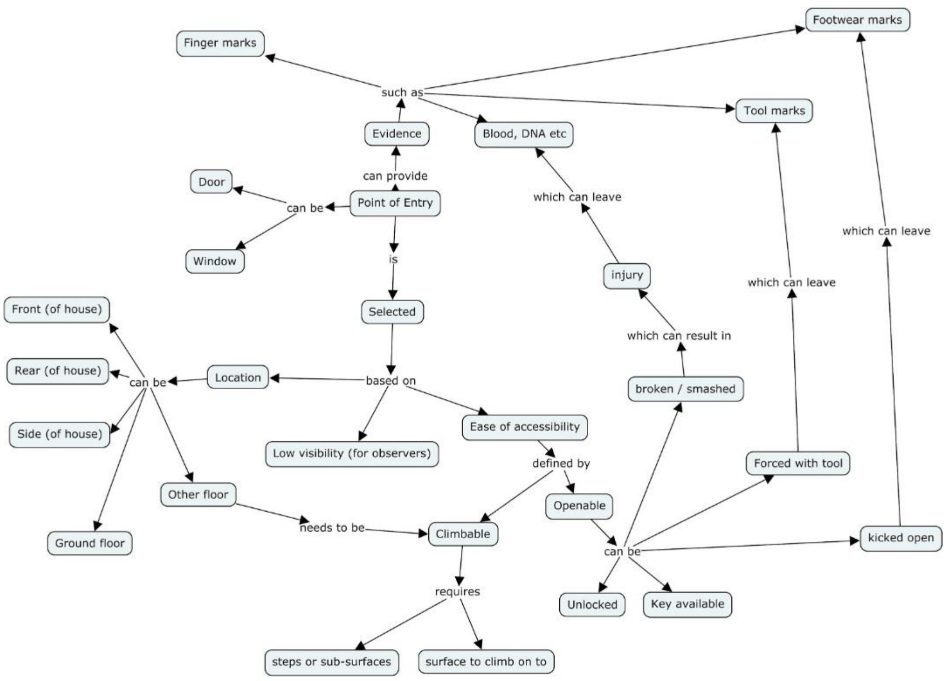


Figure1: A concept map as relevant to a 'Point of Entry' (or means of breaking into a house)

A concept map could represent the generic belief network of an individual. Furthermore, one could assume that each node could have differing levels of resting activation which would reflect prior experience and domain expertise. For example, figure 1 could be a concept map for a Crime Scene Examiner but could also be appropriate for an experienced burglar. Resting activation of specific nodes could differ between these individuals, *e.g.*, the burglar might have higher resting activation for nodes relating to 'low visibility' and 'ease of accessibility'. The CSE might have higher resting activation for nodes relating to 'evidence'. Thus, generic beliefs could be represented by an associative network with differing levels of resting activation on the nodes.

In terms of situation-specific beliefs, the concept map would be read as partially activated. In other words, assume that all of the nodes in the concept map have a level of excitation which was zero. In this case, the generic belief network would not be immediately accessible to conscious awareness. A would-be burglar, intent on breaking into a house, is walking down a street. In order to meet this intention, the would-be burglar would need to have 'point of entry' activated, which then leads to spreading activation to location, ease of accessibility and low visibility. So, these elements become a frame by which the situation can be judged. Whether these elements define a 'plan' for information search (*i.e.*, a conscious engagement with the activation of the elements) or whether they become a 'lens' for perception-action coupling (*i.e.*, a preconscious adaptation to environmental cues) is an important distinction in how sensemaking activity is performed. We feel that the NDM community tends to concentrate on the former while missing some essential aspects of the latter.

DISCUSSION

It feels credible to us that, rather than holding background knowledge in a single concept map, individuals would have a range of smaller, associative networks which are linked by 'weak ties'. In this case, the activation of nodes could be interpreted as expectations for more information. If all of the expectations are met, *i.e.*, if all of the associated nodes become active, then 'sense' is made. In this instance, coherence involves the confirmation that background knowledge is appropriate to the given situation. If their expectations are not confirmed then additional action is required. For us, this additional action takes the form of situation-specific beliefs. In this case, contextual features (which might be emotional, physical, kinaesthetic etc.) could become activated in response to a given situation. This could involve the gradual activation of nodes associated with a given context, perhaps through the sort of Q-learning mechanisms discussed in Baber *et al.* (2015). To this end, the spreading activation need not commit the person to a fixed structure or to the need to follow rules in order to amend the structure. Intriguingly, this suggests that classroom learning which reinforces semantic knowledge only partially develop concepts and that there is a need to put the person in situations in which the associated contextual information (including emotional response) could also be learned.

Sensemaking is the process of aligning situation-specific and generic belief networks. From this, discrepancy between generic and situation-specific beliefs requires effort to reconcile beliefs and this is where sensemaking occurs. This means that there is a continued need to apply coherence tests to the available information (in the situation-specific model) in terms of the belief networks that have been activated. For example, a smashed window could indicate an attempt to break into a building. However, if the glass is on the outside of the building, this would suggest that the break arose from actions inside the building (and could, perhaps, point to an attempt at staging a burglary for insurance purposes). In summary, we are proposing that a 'frame' can be considered as the activated portion of an associative network, and that it is likely that an individual will be using more than one associative network in a given situation. At the very least, the individual will have associative networks which represent generic beliefs, where the resting activation of nodes in these networks reflect prior experience of the individual, and another associative network that is constructed to reflect situation-specific beliefs.

REFERENCES

Abelson, R. and Schank, R.C., (1977). Scripts, plans, goals and understanding. An inquiry into human knowledge structures New Jersey, 10.

Attfield, S. and Blandford, A., (2011). Making sense of digital footprints in team-based legal investigations: The acquisition of focus. Human–Computer Interaction, 26(1-2), pp.38-71.

Attfield, S., Fields, B., Wheat, A., Hutton, R.J.B., Nixon, J., Leggatt, A. and Blackford, H., (2015), June. Distributed sensemaking: a case study of military analysis. In Proceedings of the 12th International Conference on Naturalistic Decision Making, McLean, VA, USA (pp. 9-12).

Baber, C., Attfield, S., Conway, G., Rooney, C. and Kodagoda, N., (2016). Collaborative sensemaking during Intelligence Analysis exercises, International Journal of Human Computer Studies, 86, 94-108.

Baber, C., Chen, X. and Howes, A., (2015). (Very) Rapid Decision Making: framing or filtering? 12th International Conference on Naturalistic Decision Making, McLean, VA.

Bartlett, F.C., (1932) Remembering: a study of experimental and social psychology, Cambridge : Cambridge University Press.

Bartlett, F.C. and Burt, C., (1933). Remembering: A study in experimental and social psychology. British Journal of Educational Psychology, 3(2), pp.187-192.

Box, G.E., (1976). Science and statistics. Journal of the American Statistical Association, 71(356), pp.791-799.

Brooks, M., (2009). 13 Things that Don't Make Sense: The Most Baffling Scientific Mysteries of Our Time. Vintage Books USA.

Collins, A.M. and Loftus, E.F., (1975). A spreading-activation theory of semantic processing. Psychological review, 82(6), p.407.

Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J. and Patterson, E., (2005). Finding decision support requirements for effective intelligence analysis tools, Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA: HFES, 297-301.

Goffman, E., (1974). Frame analysis: An essay on the organization of experience. Harvard University Press.

Klein, G.A., (1999). Sources of power: How people make decisions. MIT press.

Klein, G., Moon, B.M. and Hoffman, R.R., (2006a). Making Sense of Sense-making 1: Alternative Perspectives, IEEE Intelligent Systems, 21, 70-73.

Klein, G., Moon, B.M. and Hoffman, R.R. (2006b). Making sense of sense 2: A macrocognitive model, IEEE Intelligent Systems, 21, 88-92.

Klein, G., Phillips, J.K., Rall, E.L. and Peluso, D.A., (2007). A data-frame theory of sensemaking. In Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making (pp. 113-155). New York, NY, USA: Lawrence Erlbaum.

Kodagoda, N., Attfield, S., Wong, B.W., Rooney, C. and Choudhury, S., (2013). Using interactive visual reasoning to support sense-making: Implications for design. IEEE transactions on visualization and computer graphics, 19(12), pp.2217-2226.

Kodagoda, N., Pontis, S., Simmie, D., Attfield, S., Wong, B.W., Blandford, A. and Hankin, C., (2016). Using Machine Learning to Infer Reasoning Provenance From User Interaction Log Data Based on the Data/Frame Theory of Sensemaking. Journal of Cognitive Engineering and Decision Making.

Lipshitz, R., & Ben Shaul, O. (1997). Schemata and mental models in recognition-primed decision making. Naturalistic decision making, 293-303.

Marshall, C.C., Shipman III, F.M. and Coombs, J.H., (1994), September. VIKI: Spatial hypertext supporting emergent structure. In Proceedings of the 1994 ACM European conference on Hypermedia technology (pp. 13-23). ACM.

McClelland, J.L., Rumelhart, D.E. and PDP Research Group, (1987). Parallel distributed processing (Vol. 2). Cambridge, MA: MIT press.

Minsky, M., (1974). A framework for representing knowledge.

Neisser, U., (1976). Cognition and Reality, San Francisco, CA : W.H. Freeman

Pennington, N. and Hastie, R., (1991). A cognitive theory of juror decision making: The story model. Cardozo L. Rev., 13, p.519.

Piaget, J., (1952). The origins of intelligence in children (Vol. 8, No. 5, pp. 18-1952). New York: International Universities Press.

Pirolli, P. and Card, S., (2005), May. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proceedings of international conference on intelligence analysis (Vol. 5, pp. 2-4).

Piaget, J., (1955). The construction of reality in the child. Journal of Consulting Psychology, 19(1), p.77.

Roth, E.M., Pfautz, J.D., Mahoney, S.M., Powell, G.M., Carlson, E.C., Guarino, S.L., Fichtl, T.C. and Potter, S.S., (2010). Framing and contextualizing information requests: Problem formulation as part of the intelligence analysis process, Journal of Cognitive Engineering and Decision Making, 4, 210-230.

Rumelhart, D., (1980). Schemata: The building blocks of cognition. In. R. Spiro, B. Bruce and W. Brewer (eds.) Theoretical issues in reading comprehension.

Russell, D.M., Stefik, M.J., Pirolli, P. and Card, S.K., (1993), May. The cost structure of sensemaking. In Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems (pp. 269-276). ACM.

Selvaraj, N., Attfield, S., Passmore, P. and Wong, W., (2016). How Analysts Think: Think-steps as a Tool for Structuring Sensemaking in Criminal Intelligence Analysis. IEEE, European Intelligence and Security Informatics Conference.

Wheat, A., Attfield, S. and Fields, B., (2016), February. Developing a model of distributed sensemaking: a case study of military analysis. In Informatics (Vol. 3, No. 1, p. 1). Multidisciplinary Digital Publishing Institute.

Weick, K.E., (1995). Sensemaking in organizations (Vol. 3). Sage.

Wittgenstein, L., (2010). Philosophical investigations. John Wiley & Sons.

The Critical Decision Audit: Blending the Critical Decision Method & the Knowledge Audit

Joseph BORDERS[a], Gary KLEIN[a]
[a]ShadowBox LLC

ABSTRACT
The purpose of this article is to propose a hybrid Cognitive Task Analysis (CTA) approach, which merges elements of two cognitive interview techniques, the Critical Decision Method (CDM) and the Knowledge Audit (KA) to streamline how we collect critical incidents and uncover characteristics of expertise. The proprosed hybrid approach uses CDM principles to capture and outline a critical incident (Sweep I: Incident Identification, Sweep II: Incident Timeline), and introduces pre-defined KA probes to target cognitive aspects of expertise (tacit knowledge) involved in problem detection, sensemaking, and decision-making. Stemming from a recent project in the petrochemical domain, we have discovered this approach can expedite the transition from CTA interview transcript to scenario-based training exercise. This more formalized interview technique may also be more time efficient and approachable for practitioners with less experience in knowledge elicitation.

KEYWORDS
Guides; Expertise, General & Miscellaneous, Cognitive Field Research & Cognitive Task Analysis

INTRODUCTION

Cognitive scientists have developed a variety of qualitative methods to elicit knowledge that is otherwise difficult for people to articulate. These methods reveal *how* someone thinks about their job, including what information they attend to, how they prioritize goals, and their underlying assumptions that contribute to detecting problems and making decisions in the field. Cognitive Task Analysis (CTA) has become a catch-all term to describe these varying methods and practices, all of which offer different strategies and formats to elicit, capture, and represent cognitive processes (Crandall, Klein, Hoffman, 2006; Hoffman & Militello, 2008). As a result, investigators have a wide range of techniques to consider when choosing a cognitive interview approach. Seasoned CTA researchers carefully adapt and tailor methods to support data collection and project requirements, and leverage their individual interviewing strengths. They may also combine elements of different techniques to accommodate different research questions and make new discoveries. These nuanced shifts in interview strategy may emerge over the course of several interviews. Less-experienced cognitive researchers often struggle to flexibly apply CTA methods. The purpose of this article is to revisit two of the most common CTA techniques, the Critical Decision Method (CDM; Klein, Calderwood, MacGregor, 1989) and the Knowledge Audit (KA; Militello & Hutton, 1998; Klein & Militello, 2001), and propose an intuitive and accessible hybrid technique for streamlining how practitioners collect critical incidents and uncover characteristics of expertise. The Critical Decision Audit is designed to achieve the advantages of the CDM and the KA while increasing the efficiency of the CDM and the power of the KA.

The CDM and KA are popular knowledge elicitation strategies for uncovering how experts come to make effective judgments in complex and ill-defined work environments (Crandall & Getchell-Reiter, 1993). The CDM structures the majority of the interview around one critical incident, a non-routine case that highlights decision strategies and expert knowledge in naturalistic environments; the assumption is that subtle aspects of expertise are elicited in tough cases. The KA relies on a set of prompts about different aspects of expertise and is more accessible and easier to implement for novice practitioners than the CDM (Crandall, Hoffman, Klein, 2008). Our team regularly uses both strategies because they allow us to represent and transfer expert knowledge into effective training applications and job aids for organizations dealing with a retiring workforce and/or turnover.

Here, we describe a case study from a recent project, which led our team to a hybrid CTA approach combining elements of both the CDM and KA techniques. The goal of this article is to 1) describe the circumstances that led to the discovery, which may stimulate additional insights regarding CTA, and 2) discuss the hybrid approach, including its constituent parts and the appropriate context for researchers to implement this strategy in the future.

Background: Cognitive Task Analysis & ShadowBox™

Our focus is on applying CTA findings and insights to develop cognitive training requirements and build scenario-based training for organizations seeking to improve their knowledge management and training capabilities (i.e., ShadowBox training; Borders, Polander, Wright & Klein, 2015, Klein & Borders, 2016). ShadowBox is a scenario-based training exercise incorporating decision points that target situational cognitive requirements. Trainees respond to decisions by selecting or ranking pre-defined options and providing their reasoning, and then comparing their own responses to a panel of experts who completed the same exercise (Hintze, 2008, Klein, Hintze, & Saab, 2013). The idea is to bypass the training bottlenecks often imposed by limited availability and access to subject matter experts and bring novices up to speed faster. ShadowBox allows the trainee to see the scenario through the experts' eyes without the expert having to be present during the training. Thus, we regularly implement the CDM and/or KA techniques with subject matter experts at the start of ShadowBox projects.

Critical Decision Method (CDM). In the past, we have preferred to use the CDM as a front-end cognitive interview method with ShadowBox because it prompts interviewees to describe rich, complex stories of actual events (sweeps I and II), which we can later fine-tune to become the foundation for ShadowBox training exercises. Experts' stories serve as excellent training scenarios because they describe tough cases in which they had to draw from their experience to resolve an issue. Moreover, their stories include naturally occurring decision points that involve cognitive skills, such as prioritizing information, setting goals, anticipating future events, and taking actions. Sweep III of the CDM deepens on decision makers' assessment of the incident, which prompts them to articulate their tacit knowledge and the decision strategies that they relied on throughout the incident. This process systematically brings to light the priorities, goals, and expectancies they had during the incident, and the types of cues and/or anomalies they were noticing. Additional what-if queries (sweep IV) investigate potential expert-novice differences and errors, which can be combined with sweep III to generate scenario content, decision points, correct and incorrect response options, and expert-informed feedback within the scenario training.

Knowledge Audit (KA). The KA similarly highlights elements of expertise by identifying cognitive skills and capabilities. However, the KA applies pre-defined knowledge elicitation probes that cover a number of expertise dimensions (e.g., noticing, job smarts, past & future), which may result in the interviewee describing a variety of mini-cases. To start the KA interview, it is first necessary to identify a decision or task of interest. The task diagram is an effective method to generate parts of the job that require skilled judgment (i.e., tasks of interest), but it is not necessary. Next, the interviewer elicits elements of the interviewee's expertise using a set of pre-defined probes and follow-up questions. This information is recorded using a matrix (see Figure 1), which includes the columns: probe, cues and strategies, and why difficult. Consequently, this option may be advantageous under certain conditions, such as when the interviewer has limited domain knowledge, minimal interviewing experience, and/or restricted time with the interviewee. The KA is also useful when the interviewee cannot recall specific incidents, or cannot discuss them. The CDM requires ample time to exhaustively analyze an incident, therefore it can be challenging to conduct when there are time constraints and/or when the interviewee has difficulty thinking about an incident in fine detail. Because of these reasons, the CDM can be intimidating to novice interviewers.

In a recent project, we identified a novel way to combine the KA and CDM techniques into a hybrid method, the Critical Decision Audit. The idea was to deeply examine a single critical incident using a more formalized KA approach (e.g., generalized probes). Not only did this exercise yield fruitful interview data, it was easy to facilitate, which may circumvent some of the barriers associated with conducting full CDM interviews. The following describes how we discovered this approach, and the steps to implement it.

DISCOVERY

We developed the Critical Decision Audit on the last day of a recent CTA data collection visit with console operators, engineers, and training experts responsible for controlling a large ethylene reactor. Our main tasks were to identify console operators' cognitive requirements and use our CTA interviews to procure rich and challenging incidents, which we would later transform into scenario training exercises focusing on improving their problem detection and decision making capabilities. For the last interview, we decided to conduct a KA because we were content with the number of incidents we previously generated using the CDM and because we thought it may be possible to cover more areas of expertise using the various KA probes. We expected the KA to highlight a range of cognitive elements and uncover expert-novice differences regarding how an operator notices, reacts to, and recovers from process upsets in the reactor.

Early on during the interview process we noticed that the interviewee wanted to share a particular story that he had recently encountered. We had a basic understanding of the cognitive difficulties and complexities associated with the work of console operators, therefore we bypassed the task diagram portion of the interview. Instead, we began the session by describing the KA format, and presenting the first probe: *Is there a time when you were called into a situation and knew exactly how things got there and where they were headed?* The interviewee described a puzzling incident involving a faulty valve leaking dangerous materials to a flare, which caused an unstable pressure differential between a catalyst storage tank and the reactor. Unlike the novice operator at the console that day, the interviewee was able to quickly find the leaking valve and make the appropriate adjustments before the pressure differential became too critical and damaged the polyethylene reaction. Continuing the KA matrix (see Figure 1), we then asked about the specific cues and strategies he was using to notice the malfunctioning valve. He described two: first, he frequently sifted through the various panel displays out of curiosity; second, he closely monitored trend data throughout the plant. These tendencies led him to quickly spot the pressure differential in this situation. To complete this probe, we asked about why this was such a difficult task for the novice operator. He described how it is confusing and overwhelming for inexperienced operators to imagine how seemingly disparate parts of the plant interact.

As we introduced other KA probes (e.g., interactions, noticing, job smarts), the interviewee continued to describe aspects of the same incident. Rather than asking him to describe new cases, we delve further into the story about the leaking valve. Prior to continuing with the KA probes we asked the interviewee to broadly outline the incident so we had a rough timeline that we could use as a reference. After constructing a basic timeline with the start, end, and major events, we continued with the KA. Each additional probe uncovered new information about the incident, and allowed us to investigate the cues and strategies that led him to detect the problem and make effective decisions. For example, when asked about being able to attend to appropriate information and parsing through noisy data, he described the red flags and subtle warnings signs that popped out to him during this incident. We also continued to press for more cues and strategies that set him apart from his novice counterpart. He described the benefit of using trends and keying in on specific process variables, which afforded him the clues necessary to noticing the pressure differential.

This strategy was particularly effective because the pre-defined KA probes and corresponding matrix allowed us to deeply explore a single incident. During the two-hour interview, we surveyed five probes highlighting a wide range of the interviewee's expertise. Rather than working through the incident chronologically to deepen from one level of understanding to the next as directed by the CDM technique, we let the KA probes and follow-up questions guide the interviewee to describe the incident in the context of his decisions and actions. This approach bypassed all of the chronological deepening and instead just focused on the aspects of expertise that were involved in the incident. The follow-up questions allowed us to consolidate sweep III (deepening) and IV (hypotheticals) of the CDM technique within each probe. This approach made it easy for the interviewee to describe the cues and strategies he used, and why the situation was challenging because he was actively thinking about that moment and/or decision. By the end of the interview we had reconstructed the incident in detail, and the KA probes yielded a fruitful matrix of cognitive skills and strategies, and expert-novice contrasts that the interviewee had identified throughout the scenario.

We found this organic process to be productive and easy to navigate for both the interviewer and interviewee. Each additional probe simultaneously added new information to the incident timeline, while also revealing additional elements of expertise. The interviewee was able to associate the probes with the particular incident of interest, and he described the cognitive elements of the incident with ease. The result was a useful a very detailed incident, including a number of ways the interviewee used his expertise to make sense of the situation, detect problems, and make decisions. We believe this approach may be a more accessible interviewing technique for CTA novices. To our surprise, this interview yielded more productive insights than most of the CDM interviews we had previously conducted. The following section outlines recommendations for implementation.

A NEW HYBRID APPROACH: THE CRITICAL DECISION AUDIT

The CDM and KA, like most CTA methods, are highly adaptable. This is critical because they must equally accommodate various data collection environments (e.g., petrochemical plant, military/police environments) and be capable of making new discoveries. Often the KA is paired with the CDM when the interviewees' expertise is sparse, or when they have difficulty recalling specific incidents. The KA probes can be surveyed at the beginning of an interview to help trigger a specific incident, which can be explored in more depth using the CDM (Klein, Armstrong, 2004). Here, we are suggesting that we use the KA probes to survey the relevant expertise involved in a single incident.

Step 1: Incident identification

The first step of the interview involves identifying a complex incident that is rich enough to elicit discoveries about cognitive phenomena. We believe there are a number of ways to achieve this first step. The CDM interview begins by asking interviewees to think of a time when their skills were challenged. They are asked to briefly describe these incidents from start to finish, and the researcher selects what s/he believes will be the most fruitful incident to explore further. The KA interview begins with a generalized probe, which will ultimately prompt the interviewee to recall an incident. For less experienced interviewers, we recommend using the KA probe to start the interview. This approach may also be advantageous if the interviewee has difficulty recalling incidents that challenged their skills. The KA probes generally provide more context and are easier to contemplate, which may help the interviewee recall a richer incident. Ideally, the first probe will bring about a complex incident that can be expanded upon, but this may not always be the case (see Table 1 for KA probes and descriptions). In some cases, multiple probes should be surveyed to stimulate an incident. For the purposes of this exercise, we think either strategy listed can be effective in generating a productive incident.

Step 2: Incident timeline

After identifying an incident for further exploration, the next step is to establish a timeline for the incident. The CDM emphasizes the importance of this step because it highlights the major events taking place in the story, and it brings to light the tasks and decisions the interviewee had to perform throughout the incident. In this hybrid approach, we recommend creating a preliminary timeline, which includes the starting event, major plot points, and the finishing event. In doing this, it is critical to identify critical events, decision points and shifts in situational awareness because these will guide which KA probes to use throughout the interview.

One of the goals of this approach is to provide adequate space for KA probes and the follow-up questions to guide the interviewee in articulating their expertise in the context of the incident. Thus, we suggest generating a brief outline that the interviewer can refer to and naturally flesh out over the course of the interview. The advantage of this approach is that it provides constraints on how judgments and decisions were made, enabling plausibility checks on the memory of the interviewee.

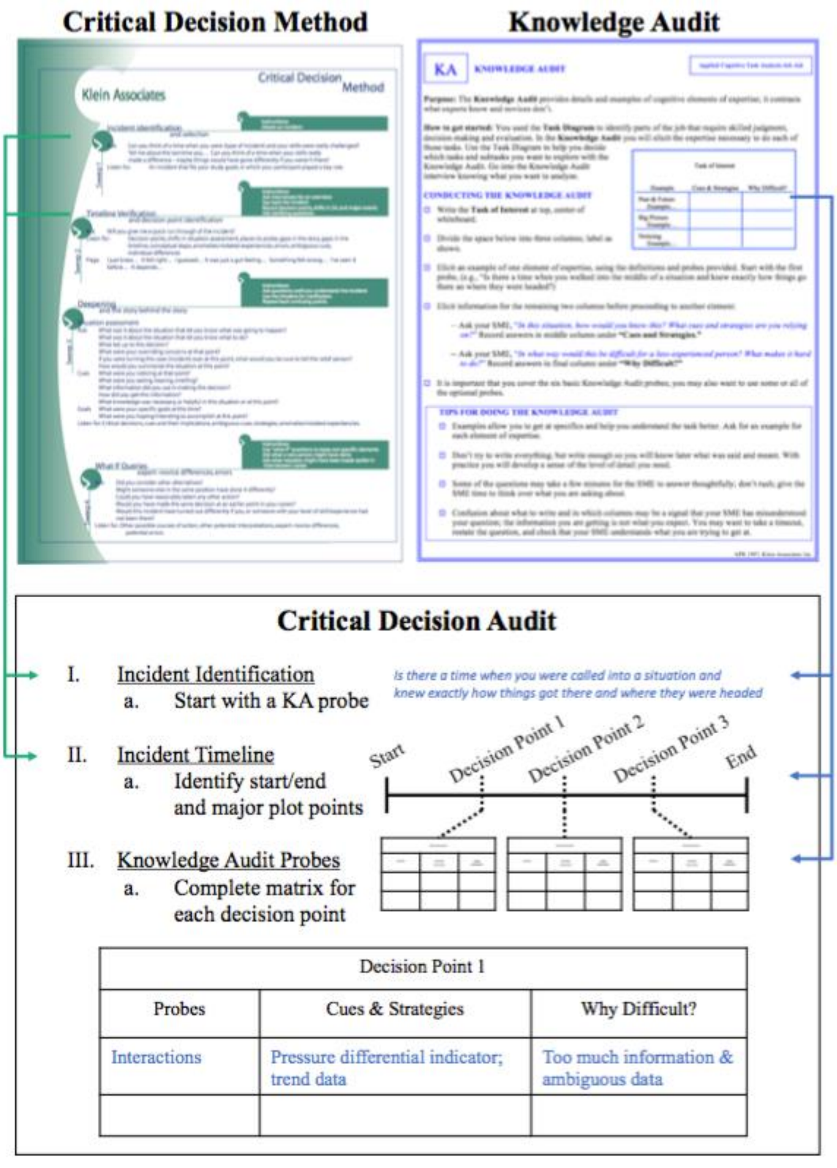


Figure 1. Critical Decision Audit

Step 3: Knowledge Audit probes

The majority of the interview should be centered around the KA probes. The goal is to use these probes to explore and analyze the cognitive elements of expertise within a critical incident. The probes are designed to be applicable to most complex domains and can be introduced in no particular order. However, if the interviewer has background knowledge in the domain and/or the incident of interest, s/he should use this information to select the most relevant probes. Also, as the interview progresses we become more knowledgeable about the incident and can inject more appropriate probes. For example, in our interview we learned that the expert was making sense of a complex network of information, which prompted us to introduce the *interactions* probe (See Table 1).

Table 1. List of Knowledge Audit probes

Knowledge Audit Probe	Probe Description
Past and future	Experts are able to recognize problems before they happen. They have a foundational understanding of the situation, including what has happened up to this point so they can accurately forecast what is to come.
Interactions	Experts are able to grasp various elements and piece them together quickly to understand the situation as a whole. Novices may only focus on small pieces of the situation and fail to see larger impacts/implications.
Paying attention	Experts are able to cut through the noise and attend to appropriate information (e.g., detecting cues, recognizing patterns). Novices may miss important cues and information altogether.
Opportunities & Improvising	Experts can improvise and adapt in complex situations. Their extensive domain knowledge allows them to understand what will work and what will not work in the given situation. Novices may have a hard time moving past guidelines.
Self-monitoring	Experts continually monitor their performance and make adjustments when they are necessary to achieve the desired goal. Novices may simply focus on task completion and not think about their performance.
Anomalies	Experts know what is normal vs. what is not normal. They can also pick up on the absence of something that is expected. Novices may not be able to distinguish what is atypical vs. typical.

For each probe, it is important to construct a matrix to elicit information about the cues and strategies they used and why the task/decision was difficult (see Figure 1). This approach integrates sweeps III (deepening) and IV (hypothetical) from the CDM into each KA probe. Thus, the probes should be followed by questions that explore the cues they were paying attention (or not paying attention to), the strategies they were using to make decisions, and why the task was so challenging. The KA user manual (Klein Associates, 1997), suggests asking the following questions: *In this situation, how would you know this? What cues and strategies are you relying on? In what way would this be difficult for a less-experienced person? What makes it hard to do?* We suggest coming up with one's own follow-up questions that pertain to the incident. The interviewee's responses should be recorded and examined in more detail when necessary. The *cues and strategies* prompts should uncover the interviewee's tacit knowledge, and bring forth the cognitive skills the interviewee relied on in the situation.

The latter questions are meant to elicit the novice perspective. That is, the elements of this situation that would cause a novice to struggle. In the past, we've found that many interviewees assumed most of their colleagues knew how to do the job, and they were uncomfortable engaging in hypothetical thinking. Instead, we framed the *why difficult* question as "If I were at the console making decisions (it is important to note here that we know very little about petrochemical processing), how could I mess this up?" This turned out to be a more fruitful strategy to elicit expert-novice differences, and bring to light the macrocognitive processes that our interviewee took for granted (e.g., checking trend data on the valves in the storage tank). It was easier for him to imagine the ramifications of having a completely inept person making decisions on the console, rather than one of his respected colleagues. Additionally, it was through this thought exercise that he became more cognizant of the number of decisions he makes and the amount of information that he has to attend to at any given moment, and especially during process upsets.

We recommend going through as many probes as time permits. As mentioned, the interviewer should develop an extensive understanding of the incident over the course of the interview, which should direct which probes and follow-up questions to use. The information that is captured from the *cues and strategies* and *why difficult* questions should guide the conversation, and inform the interviewer about which probe to introduce next. Ideally, the interview should be a fluid interchange that brings to light the cognitive elements underlying the critical incident.

DISCUSSION

Combining elements of the CDM and KA may be an effective way to streamline how we capture critical incidents and elicit cognitive skills. Previous literature has recommended using the KA at the front end of the CDM to become familiarized with the various elements of expertise involved in a particular task and/or when the interviewee has limited expertise (Klein & Armstrong, 2004). In this article, we have proposed a hybrid-interview strategy, the Critical Decision Audit, which incorporates components of both techniques and is intended to deeply investigate cognitive elements of expertise in the context of a single incident (as opposed to mini-cases). This approach leverages the KA probes and follow-up questions in correspondence with an incident timeline to elicit effective strategies for sensemaking, problem-detection, and decision making. Moreover, this strategy may be more suitable for less experienced CTA researchers and practitioners because it follows the pre-defined KA probes and requires less improvisation and adaptation from the interviewer.

The advantage of this approach is that it facilitates in-depth exploration of an incident as afforded by the CDM, while maintaining the accessibility and clarity of the KA. This approach emphasizes the KA probes, which provide a more straightforward script that can be adapted when necessary. We found these probes and the proceeding questions to be useful substitutes for the deepening and hypothetical (what if) sweeps of the CDM, which require more finesse and skill to elicit. Perhaps the most appealing aspect of using the KA probes to explore a single incident was that the resulting interview data linked directly to cognitive training requirements and provided a foundation for constructing training scenarios.

Scenario Development

One of the hallmarks of CTA is that it provides a means to represent and transfer tacit knowledge into effective training applications and job aids. Our team regularly applies CTA findings to construct scenario-based training. The Critical Decision Audit offers a way to streamline scenario development. We discovered that the KA probes and follow-up questions uncovered pertinant information that could be transformed into decision making exercises for training purposes (e.g., ShadowBox). For example, reflecting back on the interview with the panel operator we can see how his expertise can be transformed into training.

The first probe that we introduced asked him to describe an incident where he was called to a problem and quickly made sense of the incoming information and fixed the complication. His responses, particularly concerning the cues he attended to and the decision strategies he employed to notice the leaking valve provide two important pieces of information for scenario development. First, this information informs *where* we could introduce a decision point (i.e., question) within a training scenario. What cues and information was the expert using to detect the problem? Secondly, his responses and insights can help construct the expert model, which the trianee can compare their decisions against (Table 2). For ShadowBox training, we often include a variety of alternatives for the trainee to choose from at each decision point. An additional benefit of this approach is that each probe includes the follow-up question: Why is this such a difficult task for a novice? The expert's response to this question helps us better understand where novices have trouble, and allows us to represent inaccurate and limited mental models through foils and alternative options. These options become "traps" in the training scenario, which serve as learning moments for the trainees that fall into them.

Table 2. Elicitation strategies for scenario development

Question/probe	Scenario application
Why would this decision be more difficult for a novice to make?	The expert's response can inform what information is included (or not included) in the scenario, which impacts the difficulty of the impending decision.
What are the alternatives at this moment? What are the acceptable and unacceptable actions; and why?	The expert's action (assuming it was correct) and tenable alternatives are used to construct the expert model. The unacceptable actions can be used to develop foil options (i.e., traps).
How could I mess this situation up? What if I were to perform X or Y?	The expert's responses can be used to develop foils. If an expert does not prefer to engage in hypothetical thinking, the interviewer can take on the role of the novice and suggest inappropriate actions.

CONCLUSION

The flexibility of CTA is powerful because it allows researchers to adapt their data collection strategy to make new discoveries and create useful end products (e.g., decision aids, training exercises). We originally intended to conduct a KA, but adapted the interview to accommodate the interviewee's tendencies and satisfy our end goal, which was to build cognitive training scenarios. Part of this discovery was also finding and tweaking the KA probes to yield new insights. Table 1 lists the KA probes that we used in this case study. These probes in connection with an incident allowed us to ultimately develop an effective scenario from the incident. As with all CTA strategies, we encourage readers to further refine and adapt this approach around their own data collection objectives and project goals, which will hopefully yield new discoveries.

ACKNOWLEDGMENTS

We thank the Center for Operator Performance (Dayton, Ohio) and their partners for supporting this research. We also thank Emily Polander for her contributions in developing this CTA technique.

REFERENCES

Borders, J., Polander, N., Klein, G., & Wright, C. (2015). ShadowBox™: Flexible training to impart the expert mindset. *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.*

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: a technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science*, *16*(1), 42-51.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* MIT Press.

Hintze, N. R. (2008). First responder problem solving and decision making in today's asymmetrical environment. Master's thesis, Naval Postgraduate School Monterey, CA.

Hoffman, R., & Militello, L. G. (2008). *Perspectives on Cognitive Task Analysis: Historical Origins and Modern Communities of Practice*. New York: Taylor and Francis.

Klein Associates Inc. (1997). Critical Decision Method. Dayton, OH: APR.

Klein Associates Inc. (1997). Knowledge Audit. Dayton, OH: APR.

Klein, G., & Armstrong, A. A. (2004). Critical decision method. In N. A. Stanton, K. Brookhuis, E. Salas, & H. W. Hendrik (Eds.), *Handbook of human factors and ergonomics methods* (pp. 35-1). CRC Press.

Klein, G., & Borders, J. (2016). The ShadowBox approach to cognitive skills training: An empirical evaluation. *Journal of Cognitive Engineering and Decision Making, 10*(3), 268-280.

Klein, G., Borders, J., Wright, C., & Newsome, E. (2015). An empirical evaluation of the ShadowBox™ training method. Paper presented at the 12th International Conference on Naturalistic Decision Making, McLean, VA, 2015.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on systems, man, and cybernetics*, *19*(3), 462-472.

Klein, G., Hintze, N., & Saab, D. (2013, May 21-24). Thinking inside the box: The ShadowBox method for cognitive skill development. In H. Chaudet, L. Pellegrin & Bonnardel (Eds.), *Proceedings of the 11th International Conference on Naturalistic Decision Making, 2013*. Marseille, France. Paris France: Arpege Science.

Klein, G., & Militello, L. (2001). Some guidelines for conducting a cognitive task analysis. In *Advances in human performance and cognitive engineering research* (pp. 163-199). Emerald Group Publishing Limited.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618-1641.

Exploring the Decision Making of Police Officers Investigating Cases of Child Abuse within the Family

Laura BOULTON [a]

[a] *School of Forensic and Applied Sciences, University of Central Lancashire, Preston, PR1 2HE,*

ABSTRACT

Identifying the cognitive processes underlying investigative decision making in cases of child abuse is vital for reducing risk to safeguarding and justice through improved training. Despite this, very little research has been conducted into this specialised field. This study begins to address this gap by intitially exploring the decision making of four British Senior Investigating Offciers (SIOs) during challenging cases of child abuse using Cognitive Task Analysis methods. Whilst a range of cognitive, situational and organisational factors were identified as impacting on the decision making of the investigators, safeguarding was considered to be 'paramount' despite conflict with traditional investigative goals. This study provides some insight into the investigative decision making processes of specialist teams, but should be considered as a pilot study which will inform the design and provide a rationale for the proposal of a larger, more comprehensive study into SIO decision making in cases of child abuse.

KEYWORDS

Decision Making; Policing; Cognitive Task Analysis; Child Abuse Investigation.

INTRODUCTION

Child abuse has become and ever growing social problem. The NSPCC (2015) has estimated that, whilst the extent of child abuse in the UK may never be fully understood due to the hidden nature of such events, more than 57,000 children are currently at risk and in need of protection. This figure shows a significant increase from 2011 where only 50,000 children had been identified (NSPCC, 2015). Furthermore, the NSPCC identified that 47,000 of these cases were corded sexual offences against children (Bently, O'Hagan, Raff & Behatti, 2016).

The police service are one of the key stakeholders when it comes to protecting lives of children (NPIA, 2009), but they are the only agency that can investigate and take effective action against suspected offenders. When the police investigate allegations of child abuse there are specialist teams available to complete the investigations. In more current cases with relevance to child sexual abuse cases, there is the ability for the collection and use of forensics (i.e. ceasing of clothing, DNA examination and, where appropriate, medical examinations carried out by specialist doctors). Statements from victims are a critical part of the investigation and are usually taken by the first responder once the initial call comes in; further from here, video interviews are obtained at later stage of the investigation process. After all the evidence is gathered the decision to prosecute or not is in the hands of Crown Prosecution Service (CPS) who consider whether there is a possible conviction, and whether the prosecution is in the public interest (Citizens Advice, n.d.).

In order to fulfil the needs of investigating child abuse cases, it is essential that all police officers are confident in identifying child abuse and in their response to this identification in order to safeguard and promote the welfare of children, as well as protect their right to justice (NPIA 2009). High profile cases in the news in which the police, and other affiliated services, failed to protect children from abuse at the hands of parents and family members highlight the impact that police investigative failings can have on the victim, the family, the community and the wider general public. High profile cases in which this has occurred include the deaths of Poppi Worthington and Daniel Pelka. In both of these cases, a large number of potential opportunities for safeguarding of these children were missed by the Police, the NHS, social services and the school system. After the childrens' deaths, a further set of opportunities for justice were missed by the same agencies. In the case of Poppi Worthington, a High Court judge concluded that 12 basic Police failings had denied Poppi justice and that this was failure is un-recoverable due to lost opportunities to collect evidence at the time (i.e. missed opportunities to collect vital evidence, failure to secure the family home and delayed family interviews). The need for change within Police processes during child abuse and child death investigations have been strongly stated by the NSPCC who declared, "No child must ever be failed again in this way" (Rayner, 2016).

Naturalistic Decision Making (NDM) research examines how people make decisions in highly demanding and complex real-world situations, including how they can use personal experience to deal with challenging factors such as time pressures and restrictions, uncertainty, goals and objectives, internal organisational pressures and possible team conflicts (Klein, 2008). As a cognitively complex task, which is typically associated with situational challenges such as time pressure, uncertainty and high stakes, police investigation can be considered an NDM working environment.

In a qualitative comparison of British and Norwegian detectives, Fahsing and Ask (2013) identified 14 situational factors that impacted their decision making. These included; time pressure (the need to "wrap up the case"), availability of information and evidence, external pressure and community impacts ("reputation of entire Norwegian police force on your shoulders"), internal pressure (high workload) and organisational issues (Fahsing & Ask, 2013, p. 160). In support of these findings, research has shown that increased time pressure can negatively affect a decision maker's flexibility by reducing ability to generate alternative hypotheses and hypothesis testing strategies (Alison, Doran, Long, Power & Humphreys, 2013). Furthermore, when investigators are under time pressure, they have been found to seek hypothesis-consistent information to confirm initial beliefs regarding a crime, thereby avoiding hypothesis generation (Ask & Granhag, 2005). Whilst uncertainty and framing of information has been found to impact on evidence search strategies and interview question style, resulting in search strategies based on initial assessment of guilt or innocence (Hill, Memon & George, 2008; Rassin, Eerland & Kuijpers, 2010).

Fashing and Ask (2013) also identified 10 individual factors that impacted on investigative decision making. These included; experience ("best friend but also worst enemy"), personal characteristics ("investigative mind-set is vital") and training and education (Fahsing & Ask, 2013, p. 161). Such individual factors can be influenced by cognitive capacity overload which may reduce controlled processing ability and therefore impact decision making reliability (Kleider-Offutt, Clevinger & Bond, 2016), as well as vulnerability to forms of cognitive heuristics and bias (Croskerry, 2013). In their review of child abuse inquiry reports, Munro (1999) found evidence of both the availability heuristic (assessments of risk were based on a narrow range of evidence biased towards the information readily available and more memorable) and confirmation bias (a critical attitude towards evidence was found to correlate with whether or not new information supported existing views) in the risk assessment decisions of child protection professionals. Whilst this research is very dated now, it does suggest that errors in professional reasoning in child protection work are predictable on the basis of decision making and heuristics research (Kahneman & Klein, 2009; Tversky & Kahneman, 1975). If this is the case, errors may be reduced through training, i.e. implementing aids that recognise the central role of intuitive reasoning but offer systematic methods to reduce bias (Munro, 1999) or via high fidelity stress exposure training (Alison et al., 2013). However, currently there is little research examining this area of decision making, therefore in order to be able to recommend ways to improve decision making and avoid bias in these settings, first we must understand the processes involved in this domain through research.

The aim of this pilot study was to begin to identify the key decisional processes involved in the decision making of expert British SIOs during the investigation of child abuse cases in order to serve as guidance for further, more in depth, empirical evaluation. NDM research has promoted the study of expertise to highlight efficient decision-making techniques, which can be promoted and learnt from. The use of Cognitive Task Analysis (CTA) qualitative research methods can be used to generate meaningfully informed hypotheses suited to subsequent empirical testing (Wiltshire, Neville, Lauth, Rinkinen & Ramirez, 2014). For the present purposes, CTA methods are used as exploratory means to derive an integrated theoretical framework, which can then be tested empirically in future studies and contribute to the current evidence-based policing agenda. This approach is hoped to result in more informed policing and directed resources towards not only areas and issues that require it the most, but also in a way that has been found to be efficient.

METHOD

Design

This exploratory study employed CTA methods to qualitatively examine the decision making of British SIO's during the investigation of child abuse. Through consideration of the available CTA knowledge elicitation techniques it was decided that an interview protocol based on the Critical Decision Method (CDM) interview protocol would be most suitable to retrospectively examine the decision making processes of SIOs during a previously experienced investigation of child abuse which they considered to be non-routine. The CDM (Crandall, Klein, & Hoffman, 2006) is structured as an intensive incident based interview protocol which aims to identify the decision making processes involved in the judgments made during a 'challenging' incident that have been personally experienced. CDM has been used in a number of studies to identify the strategies, expertise, and knowledge requirements involved in other critical decision making situations which have led to important insights for designing better decision aids (see Wong & Blandford, 2002).

Sample

A total of four SIOs (all male) voluntarily participated in this study. Whilst this sample size is small, this is in line with many CTA studies which are typically based on a small sample size (<10) due to the large amounts of qualitative data that are generated by these approaches and limited access to experts of specific fields of interest (see Wiltshire et al., 2014). Furthermore, it is emphasised that this study is aimed to serve as a pilot upon which future research can be designed. All of the SIOs in this sample can be considered as experts as they work on a specialist team that deal with child abuse cases on a day-to-day basis, have led cases, have had specialist training and have national accreditation. Demographic data relating to the SIOs were not collected to protect anonymity. Prior to data collection, ethical approval was gained and each participant signed informed consent.

Materials

In order to conduct the analysis, permission was sought to record interviews in audio format using a Dictaphone (Olympus: WS-852). The participants were informed of this procedure before signing consent forms. The interviews were conducted guided by a script. Qualitative data analysis software NVivo 11 (QSR International) was used to assist the qualitative analysis of the current studies.

Procedure

Access was granted from the head of the Force Major Investigations Team at Lancashire Constabulary, but emphasis was placed on the importance of officer anonymity. To protect anonymity, no demographic information was collected and participant numbers were used throughout analysis. Data collection consisted of semi-structured interviews using a CDM based script. Each participant was asked to walk through a ‘challenging’ and non-routine case of child abuse investigation that they have personally experienced as a SIO. The interviews were conducted in the officers’ workplace. Each interview lasted between 45-60 minutes.

Data Analysis

All audio recordings were transcribed and the transcripts were reviewed for accuracy immediately after collection. The data analysis reflected a framework analysis methodology, which allowed for both a ‘top-down’ (theory-driven) approach and a ‘bottom-up’ (data-driven) identification of emergent patterns (Wiltshire et al., 2014). Firstly, the data set was read multiple times whilst considering issues which appeared to be relevant to the analysis. The interview transcripts were then inductively coded for repeated ideas, which were reviewed and grouped into themes and subthemes. This process was iterative and was conducted by the named author.

RESULTS

The data collected referred to the investigation of child abuse within a family environment, three of the cases were current cases and one referred to historic offences. All of the cases (n=4) had successful convictions at court, with all suspected offenders pleading guilty and concluding with custodial sentences. When looking at the work styles only one participant worked on the investigation as a single detective rather than as part of a team. In addition to this only one participant attended the investigation as an initial responder. In all cases discussed, the victim of the abuse was female and under the age of 13, whilst the suspect in all cases was male and 18 years or older. In two cases, the suspect was the victim’s brother, one suspect was the victim’s uncle and in one case the suspect was a non-blood relation at the same group home. In three cases, the abuse included rape; however, the offense type was not disclosed in the one other case. The analysis identified three themes, each with related subthemes. These themes included; (i) cognitive factors, (ii) situational factors, and (iii) organisational factors.

Cognitive Factors

All SIOs discussed cognitive factors which impacted on their decision making. These included; (i) prioritisation (safeguarding vs. investigation), (ii) information gathering, (iii) intuitive first judgments and, (iv) the considered need to remain impartial and adaptive to re-evaluation. The results highlighted the difference between the participants in the factors they feel should lead their investigations, specifically in relation to the decision to prioritise safeguarding or the need for an investigative result. All four participants referred to this. Whilst the investigative result was acknowledged to be a significant factor, this was often discussed in relation to the importance that other officers gave to it, whereas the three participants expressed the paramount need for safeguarding to be the forefront of the investigation at all times;

> P1: “The main priority at the time was safeguarding, (…) the paramount decision at that point is to safeguard those children in case he still posed to any of them (…) A lot of times in the office we have conversations about investigations coming over safeguarding (…) I am very safeguarding minded, whereas many officers want to get the investigation.”

> P2: “I would say we have to look at the best interest of the child, whereas sergeants or inspectors will look at the best interests of the investigation (…) I am quite victim centred as some of my colleague are, whereas higher ranking officers aren’t as concerned about what’s best for the victim but there is more concern about what is best for the case instead.”

P3: "Safeguarding is paramount and nowadays it tends to come before the investigation."

Information gathering to serve as evidence was, understandably, an action described by all participants. This occurred via multiple processes, including; initial response intelligence, conducting ABE (Achieving Best Evidence) interviews and forensic examination. Forensic examination and collection of evidence for forensic testing was highlighted as being extremely important to all participants, except the participant who discussed a historic case;

P4: "we needed to establish what was going on to progress anything else, sometimes the logs aren't always accurate, sometimes the comms operator misheard things or typing errors (…) From then on it was about getting as much evidence as we could for a trial (…) we would ask for fingernail cuttings, request his clothing, we would do the swabs (…) the key thing was the forensics in this case."

P2: "In order to move forward what we did was the forensics first."

When considering the decision making processes involved, three of the participants expressed that there was an initial intuitive judgement from the case briefing information based on recognition of key cues;

P3: "because of who she was, as bad as that sounds (…) I'm not saying you would stereotype but I suppose it would have an influence on some of the actions and how you would go about dealing with that person and the case in its early stages."

P4: "Sometimes you can judge it and see if it's going to be much of a job or not and that's more from experiences of what you had dealt with an the things that have been said."

However, it was also made clear by the same thee participants that the need to remain impartial and to re-evaluate judgments as the case continuously develops was at the forefront of their investigation. Here, the SIOs demonstrate knowledge of the need to be able to adapt cognitive processes and revisit and/or reject initial intuitive judgements in light of new information;

P1: "it was a consideration that it could have been a revenge report, it couldn't be investigated in that manner it has to be investigated impartially."

P3: "it never always goes to plan. There are certain things that change."

P4: "It could be helpful, it could be unhelpful, so we have to remain impartial (…) you've just got to take it step by step. Sometimes bits of information come in and send you off on completely different leads than you ever thought there would be, and you literally have the whole investigation changed before you. It's completely continuous you literally just don't know."

Situational Factors

All participants expressed their concerns with regards to NDM related situational factors that led or affected their ability to make decisions. These situational factors included; time pressure, conflicting options, 'luck', impact, as well as welfare and safeguarding considerations. Time pressure was reported by all four participants as being influential in their decision making. This manifested in the form of custody and bail timings, as well forensic examination and evidence collection time constraints. Furthermore, it was highlighted that when dealing with victims who are very young, time pressure also arises in terms of deciding when ABE interviews should be conducted considering the amount of time these, in addition of forensic examinations can take and how tired young children can get, especially after experiencing trauma;

P1: "we can't have people on bail for long anymore, and that decision has been taken out of our hands

by the government (…) we have to get authority from out superintendent to extend to extend that bail, if they don't agree they are released no charge (…) so there are definitely time pressures in terms of bail."

P2: "we did have to consider the time restraints because she is only five and should we really be interviewing a child after a certain time? (…) forensically there was time pressures definitely (…) then we have the added pressure of custody time restrains."

P4: "it's a really long process and that's why he was kept in custody because we couldn't charge him without her interview and full detailed account."

Long-term time pressure was also recounted. Two participants discussed feeling external pressure to conclude cases of child abuse to achieve justice as quickly as possible and highlighted how this conflicts with the need to conduct a thorough investigation, especially with the amount of intelligence which is generated online, i.e. social media;

> P1: "I think we are always under pressure to make decisions, especially in child abuse cases as the courts, family and the police want these sort of cases dealing with as quickly as possible. The reality of it is, that these cases are the ones that take the most length of time. So there is never a quick outcome to these cases, they could quite easily take two years to investigate."
>
> P3: "from time the investigation starts to the time he was convicted and sentenced it was probably eighteen months so as much as you do feel the pressure, it just takes time".

Uncertainty and conflicting options were highlighted by three of the participants as an additional and influencing pressure. However, interestingly, three officers also referred explicitly to luck playing a crucial part in the decisions made and outcomes of the investigation;

> P1: "we had to make a decision about how do we deal with her, treat her as a witness or a suspect? That was a really important decision, how we were best to do it (…) I mean luckily in this case this suspect pleaded guilty at the earliest opportunity (…) if I'm honest with you that is the key decision in the case that actually swayed him to plead guilty".
>
> P3: "luckily in this case because the social worker, or should I say two were already involved they played the role in ensuring that she was safeguarded".
>
> P4: "I was quite lucky because I was there right from the beginning".

Furthermore, consequential impact of any decision made was reported to be thoroughly considered. This impact manifested in terms of cost (i.e. of running certain forensics tests) and in terms of community impact and potential threat, either from community members towards a suspect if they are released on bail, or towards a community/children in that community from a suspect;

> P4: "should we send the condom off to see if her DNA was on the outside. But logically, its cost implicated and we believed he would plead guilty."
>
> P2: "But then we also have to look at the fact that offender lived round the corner from the alleged victim and we also had a lot of community issues as well (…) We had already been told by family members that if he went back to that area there would be consequences, so we had to look at the community impact if we allowed him back into that area with those bail conditions (…) No it's relatively quick, it's very costly but it's quick."

Overwhelmingly, welfare and safeguarding considerations were reported to be the biggest influential factor on decision making during the cases recalled. All four SIOs discussed this and in addition to safeguarding being referred to many times within the transcripts, it was emphasised strongly within the content of the description as being 'paramount' to the investigation;

> P1: "the main priority at that time was the actual safeguarding of any children in the case. For example our victim was alleging that she has been sexually abuse by her brother, I found out that her brother currently has his own children that are also female. So paramount decision at that point was to safeguard those children in case he still posed a risk."
>
> P2: "the uncle [suspect] also had children of his own who were a similar age to our victim which again is imperative in the case because we have got to look at safeguarding, which is paramount".
>
> P3: "safeguarding is paramount and nowadays it tends to come before the investigation".

Organisational Factors

Organisational factors that influenced SIO decision making broadly involved multi-agency conflict and factors that related to team decision making. Conflicting aims and differing needs with social service and the CPS in particular, was reported by three participants to have put additional pressure (both time and workload) on the officer and the investigation or created additional considerations to be made:

> P1: "There are pressures from CPS. When we put in advice files, or take a case to them, it is very often they will give us a further action plan and are given a time limit to do that. We then have to balance that action plan alongside our other cases."
>
> P2: "He was going to be bailed, so we had to fight with CPS, which was another issue"
>
> P3: "nowadays within what we do in the CSE [child sexual exploitation] office, they [social services] are a massive stakeholder and partner for us, I'd probably say we are in contact with them on a daily basis, more than other department in the police, we work with them very closely. Even now though it can be difficult and that communication isn't as fluid as it should be, sometimes might happen and it be three of four weeks before we get that information through".

Instances and descriptions of team conflict, processes of team decision-making and verbal development of shared mental models were discussed by three participants. The cases being discussed were all described as involving no team conflict, however this was caveated with descriptions how team conflict is typical in other cases. For the most part, this conflict was described positively as a way to generate and discuss ideas, reflect on options and develop shared mental models in order to facilitate team decision making;

> P1: "there wasn't any conflict in terms of decisions. We were all singing from the same hymn sheet. There is often in this sort of investigation a difference of opinions from supervisors and officers in the case, but a lot of the time we sit and trash the ideas out and come up with a plan we are all comfortable with."
>
> P2: "people do have differences of opinion and sometimes we are railroaded down a certain path which sometimes we don't think is the right one."
>
> P3: "I thrive off discussions and I personally would always value someone else's opinion on a job, two heads are better than one. Discussion is important so I would always much rather prefer that prior to making a decision."

DISCUSSION

The aim of the study was to explore the decision making process within an investigative setting in cases of child abuse within the family environment. The results support past research that emphasise the impact of complex situational demands such as time restraints, uncertainty and organisational pressures on the decision making that occurs in investigative settings (Fahsing & Ask, 2013).

Whilst early information framing was found to impact of SIOs initial intuitive hypotheses in relation to case legitimacy (Hill et al., 2008; Munro, 1999; Rassin et al., 2010), the officers in this sample were able to remain 'impartial' via their recognition of the continuous and changing nature of such cases. In this sense, the sample were able to retain cognitive flexibility (Diamond, 2013; Ward, Ericsson & Williams, 2012) to adapt existing hypotheses and generate new ones in response to new information. This reflect adaptive expertise (Klein & Jarosz, 2011; Kozlowski, 1998; Morrison, Wiggins, Bond & Tyler, 2013) and has been found in other policing context such as the expert decision making of Specialised Firearms Officers (Boulton & Cole, 2016).

A key finding expressed as being important to the decision making of all participants in this sample is the approach taken to child sex abuse cases in the modern day; decision-making should always be made whilst keeping safeguarding at the forefront of the mind. The participants explained that in cases with children, there needs to be additional care taken due to the age, intellectual ability and anxiety levels that can be found in children that are faced with the prospects of a criminal investigation after making a complaint. However, it was also acknowledge that the decision to prioritise safeguarding can often conflict with the investigative goals of higher-ranking officers and time pressure in both the short and long term. Whilst the need of the investigation to ensure justice for the victims was understood by the sample, the participants reflected that due to the face to face work that they do with those involved in child sex abuse cases, their mind-set had changed to prioritise the people involved rather than the investigative outcome. They believed that this does, and should, lead their decision making process.

Limitations

It should be noted that analysis is based on only four participants, all recruited from a department within the same British Police force. Although it is acknowledged that generalisability is a shortcoming, SIOs investigating child abuse are a very specific group of decision makers and as such, generalisation to a larger population is not a major consideration (McAndrew & Gore, 2013). However, it is possible that these results represent force-specific SIO decision-making themes and a larger sample generated across UK wide forces could help decipher the generalisation of these findings more accurately. Future research may seek to clarify these issues through replication with officers across different forces to examine these themes more thoroughly.

CONCLUSION

The results show that the paramount factors in investigative decision making in these cases of child abuse is the broader duty of safeguarding the children involved. However, time restrictions and organisational pressures to generate an investigative result can affect this. This study provides some insight into the decision making process on specialist teams to deal with such crimes, but should be considered as a pilot study which will inform the design and provide a rationale for the proposal of a larger, more comprehensive study into the decision making of SIO's in cases of child abuse and child death investigation.

ACKNOWLEDGMENTS

We thank Lancashire Constabulary for supporting this project and organising access to participants.We also thank Olivia Bloomfield for collecting and transcribing the data presented in this paper.

REFERENCES

Alison, L. Doran, B. Long, M. Power, N. & Humphrey, A. (2013). The effects of subjective time pressure and individual difference on hypotheses generation and action prioritization in police investigation. *Journal of Experimental Psychology: Applied. 19*(1), 83-93.

Ask, K. & Granhag, P. A. (2005). Motivational Sources of Confirmation Bias in Criminal Investigation. *Journal of Investigative Psychology and Offender Profiling*, *2*(1), 43-63.

Boulton, L., & Cole, J. (2016). Adaptive Flexibility Examining the Role of Expertise in the Decision Making of Authorized Firearms Officers During Armed Confrontation. *Journal of Cognitive Engineering and Decision Making*, *10*(3), 291-308.

Bentley, H. O'Hagan, O. Raff, A. & Bhatti, I. (2016). *NSPCC: How safe are our children: The most comprehensive overview of child protection in the UK.* Retrieved from https://www.nspcc.org.uk/globalassets/documents/research-reports/how-safe-children-2016-report.pdf

Citizens Advice (n.d). Child abuse – How the police investigate allegations. Retrieved from https://www.citizensadvice.org.uk/relationships/children-and-young-people/child-abuse/police-involvement/child-abuse-how-the-police-investigate-allegations/

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge: MIT Press.

Croskerry, P. (2013). From mindless to mindful practice – Cognitive bias and clinical decision making. *English Medical Journal*, 2445-2448.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168.

Fahsing, I. & Ask, K. (2013). Decision making and decisional tipping points in homicide investigations: An interview study of British and Norwegian detectives. *Journal of Investigative Psychology and Offender Profiling*, *10*(2), 155-165.

Hill, C. Memon, A. & McGeorge, P. (2008). Role of confirmation bias in suspect interviews. *Legal and Criminological Psychology*, *13*(2), 357-371.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American psychologist*, *64*(6), 515.

Kleider-Offutt, H. M., Clevinger, A. M., & Bond, A. D. (2016). Working Memory and Cognitive Load in the Legal System: Influences on Police Shooting Decisions, Interrogation and Jury Decisions. *Journal of Applied Research in Memory and Cognition*, *5*(4), 426-433.

Klein, G. (2008) Naturalistic Decision Making. *Human Factors: The Journal of Human Factors and Ergonomics. 50*(3), 456-460.

Klein, G., & Jarosz, A. (2011). A naturalistic study of insight. *Journal of Cognitive Engineering and Decision Making*, *5*(4), 335-351.

Kozlowski, S. W. J. (1998). Training and developing adaptive teams: Theory, principles, and research. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for training and simulation* (pp. 115-153). Washington: APA Books.

McAndrew, C., & Gore, J. (2013). Understanding Preferences in Experience-Based Choice A Study of Cognition in the "Wild". *Journal of Cognitive Engineering and Decision Making*, *7*(2), 179-197.

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 211-226.

Munro, E. (1999). Common errors of reasoning in child protection work. *Child abuse & neglect*, *23*(8), 745-758.

NPIA. (2009). *Guidance on investigating child abuse and safeguarding children* (2nd Ed). Retreived from https://www.ceop.police.uk/Documents/ACPOGuidance2009.pdf

NSPCC. (2015). Child protection register statistics UK: 2011-2015. Retrieved from https://www.nspcc.org.uk/globalassets/documents/statistics-and-information/child-protection-register-statistics-united-kingdom.pdf

Rassin, E. Eerland, A. & Kuijpers, I. (2010). Lets find the evidence: an analogue study of confirmation bias in criminal investigation. *Journal of Investigative Psychology and Offender Profiling*, *7*(3), 231-246.

Rayner, G. (2016). *Poppi Worthington: Call for public inquiry into sudden death of 13-month-old shortly after she was sexually assaulted by her father*. Retrieved from; http://www.telegraph.co.uk/news/uknews/crime/child-protection/12107318/Poppi-Worthington-Toddler-was-sexually-assaulted-by-father-before-sudden-death-judge-rules.html

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141-162). Springer: Netherlands.

Ward, P., Ericsson, K. A., & Williams, A. M. (2013). Complex perceptual-cognitive expertise in a simulated task environment. *Journal of Cognitive Engineering and Decision Making*, *7*(3), 231-254.

Wiltshire, T. J., Neville, K. J., Lauth, M. R., Rinkinen, C., & Ramirez, L. F. (2014). Applications of cognitive transformation theory examining the role of sensemaking in the instruction of air traffic control students. *Journal of Cognitive Engineering and Decision Making, 8*(3), 219-247.

Wong, B. W., & Blandford, A. E. (2002). *Analysing ambulance dispatcher decision making: Trialing emergent themes analysis.* Paper presented at the Human Factors 2002, the Joint Conference of the Computer Human Interaction Special Interest Group and The Ergonomics Society of Australia, HF2002, Melbourne.

Removing concussed players from the field: the factors influencing decision making around concussion identification and management in Australian Rules Football

Katelyn DAWSON[a], Paul M. SALMON[a], Gemma J. M. Read[a], Timothy NEVILLE[a], Natassia GOODE[a] and Amanda CLACY[a]

[a]*Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Queensland, Australia, 4558*

ABSTRACT

Concussion in sport represents a major issue worldwide. One component that has received little attention is the factors that influence decision making around concussion identification and management. This study examined the factors that influence player, coach, and medical team decisions regarding the removal of Australian Rules Football (AFL) players from the game following a suspected concussion. Using the critical decision method, participants from an Amateur AFL club were asked to reflect on their decision-making during situations where they had either suffered or managed a concussion. The findings demonstrate that the identification and management of concussion in amateur AFL is influenced by a myriad of factors from across the overall AFL system. This includes factors related to the game, players, and coaches and also factors related to assessment tools, concussion guidelines and social influences. The implications for the practical management of concussion in both AFL and other sports are discussed.

KEYWORDS

Concussion, Decision making, Australian Rules Football, Critical Decision Method, Systems thinking

INTRODUCTION

Increasing attention is being placed on the link between brain injuries caused by concussion sustained in sporting activities and the development of serious long-term health issues (Newton et al., 2014). Research has identified a range of serious consequences that can result from sports-related concussions, including neurological and cognitive disruptions (Edwards & Bodle, 2014), early cognitive decline and psychiatric disturbances (Partridge & Hall, 2014), long-term chronic effects such as Parkinson's disease or Chronic Traumatic Encephalopathy (CTE), and short-term acute consequences such as Second Impact Syndrome (SIS; Cobb & Battin, 2004).

Rapidly identifying, evaluating and managing players that have suffered a concussion is key to limiting adverse health impacts (McCrory et al, 2013). In particular, removal from the field of play in the event of a concussion is critical. Whilst there has been much research into the causes, symptoms, and consequences of concussion in sports such as NFL, football, ice hockey, and baseball, relatively less research has examined the factors influencing concussion identification and management. This is particularly so in the domain of Australian Rules Football (AFL). AFL is a full contact, fast ball sport played on large turfed ovals. Whilst the game rules attempt to ensure that players' heads and shoulders are protected (Neville, Salmon, Read, & Kalloniatis, 2016), players do not wear padding or headgear and there are numerous situations in which players are vulnerable to head trauma. These include head to head, body to head, stationary body, moving body, and head-to ground contacts. As a corollary, concussion is now recognised as a major issue. Indeed, the AFL Medical Officers Association (2015) report that, on average, six to seven concussions occur per team across a season at the professional level.

Research into contact sports such as NFL, rugby, ice hockey, and football have identified a range of factors that influence concussion identification and management (Clacy et al, 2015); however, to date there has little research examining the factors influencing concussion identification and management in AFL, either at the professional or amateur level. In other sports the importance of studying concussion at the amateur level has been emphasised, particularly where it is unclear whether professional level guidelines and systems are being implemented at the amateur level (e.g. Clacy et al, 2015; Hollis et al, 2012).

This article describes the findings from a study that aimed to identify the factors that influence decision making around concussion management, in particular decision making regarding the removal of a player from the game (referred to hereafter as concussion identification and management). The study was undertaken in AFL and involved an amateur level club from South-East Queensland, Australia. The overarching aim was to identify areas where interventions could be used to improve practice around concussion identification and management.

A SYSTEMS PERSPECTIVE ON SPORTS CONCUSSION

The study adopted a so-called systems thinking lens through which to examine the factors that influence concussion identification and management. Systems thinking models, such as Rasmussen's risk management framework (Rasmussen, 1997, see Figure 1) have been applied in to identify and examine the systemic factors that influence behaviour in different contexts, including sport (Clacy et al., 2015). Key tenets of the framework are that it views systems as comprising a hierarchy of different people and organisations, and that it seeks to identify the factors across the hierarchy that influence behaviour. It is argued that behaviour is influenced by the decisions and actions of people and organisations across all levels of a system (rather than just be the result of one bad decision or action) and that adverse events have multiple, interacting contributory factors. Any interventions designed to improve behaviour therefore need to consider influencing factors across the overall system rather than merely factors at the so-called sharp-end (Goode et al, 2016). In the present context, the AFL system can be viewed then as a hierarchy made up of different actors across multiple levels, ranging from players to coaches and medics who are overseen by regulatory and governing bodies. According to the model, decisions around concussion identification and management will be influenced by a range of factors from across this system. This study attempted to determine a. what these factors are, and b. where in the amateur AFL system they reside. The intention was to gather information to support the development of interventions designed to improve concussion identification and management in AFL.

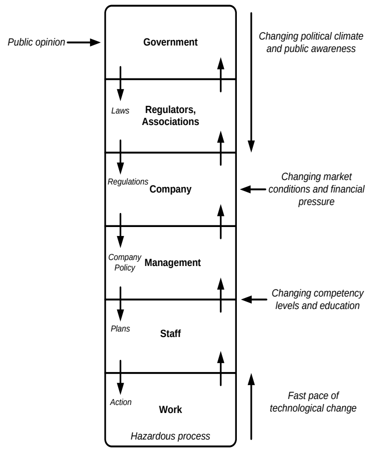


Figure 1. Rasmussen's risk management framework (Rasmussen, 1997).

METHOD

Participants

Participants were 15 members of an amateur AFL club from South-East Queensland, Australia, including 9 players, 2 coaches, and 4 members of the club's medical team. Inclusion criteria for the study required participants to have experienced, witnessed, or suspected a concussion during an amateur AFL game in the previous two years (2013-2015). Participants' ages ranged from 20 to 54 years old (M = 30.20, SD = 10.39) and on average they had been involved in AFL for 9.26 years (SD = 6.75). All participants were male, which reflects the typical demographic makeup of amateur AFL clubs in Australia.

Materials

To gather data on participants' decision making following a concussion incident, the Critical Decision Method (CDM) semi-structured interview technique was used (Klein, 1989). CDM has been used extensively to examine decision making in naturalistic environments in a range of areas, including sport (Salmon et al., 2010). A series of predefined questions from the CDM literature were used (e.g. Crandall et al., 2006), some of which were modified specifically for use in this study (see Table 1).

Procedure

Ethics approval was provided by the host institution's human research ethics committee. Approval for the study was also obtained from the president of the participating football club. Before data collection commenced two pilot interviews were conducted to ensure the adapted probes were appropriate.

The interviews were conducted at the football clubhouse and were recorded using a standard dictaphone. After completing demographic and consent forms, players were asked to describe an incident in which they had experienced a concussion or suspected concussion. The medical team and coaches were asked to describe an incident where they had witnessed a concussion or suspected concussion and whether they allowed the player to play on or instructed the player to come off the field. Following this, the adapted CDM probes were used to gain further information regarding decision making during the incident.

Data Analysis

The interview recordings were transcribed verbatim using Microsoft Word and then coded using the NVivo qualitative data analysis software (version 10; QSR Software) following an inductive coding approach (Hsieh & Shannon, 2005). Participants' responses to each probe were coded to that specific question and then coded into themes to create a coding template. For example, the statement "There was no time pressure because a stretcher was called" (Medical Team 1) was the answer to the question "How much time pressure was involved in making your decision". The statement was coded to the question 'Q8timepressure' and then allocated to the theme of 'No time pressure'. The data were then further analysed for factors influencing behaviour. An example of one of the statements was "I needed to go back on and I needed to win". This statement was coded to the theme of 'Winning'. The probes were also categorized to assist the data analysis process. To ensure responses were coded consistently, two researchers reviewed the resulting codes, one of whom had extensive experience in research on concussion in fast ball sports. Any disagreements were discussed further until consensus was achieved.

Table 1. Critical Decision Method probes used during study.

Probe categories	Probe
Scenario description, goals and symptoms	Can you please describe a time where you experienced a concussion or suspected concussion and continued to play on or hide you symptoms? Can you please draw on this piece of paper where you were situated on the field when the decision was made? Can you please tell me about your surroundings – who was near you? Where was the ball? What were your specific goals at the time of making your decision? What were you experiencing at the time, including hearing, feeling or seeing?
Time pressure	How much time pressure was involved in making the decision to play on?/ How much time pressure was involved in making the decision to let the player play on or to pull them off the field?
Knowledge of concussion and situation assessment	Can you tell me about your knowledge surrounding concussion? Did you use any of this knowledge when making the decision to play on /come off? Can you please tell me some of the consequences of concussion? What training of experience do you think was necessary in helping to make your decision? Did you follow a formal or informal rule when making your decision to play on? What mistakes are likely when assessing a player with a concussion or suspected concussion?
Situation awareness	What information did you use to make your decision? How would the situation have turned out differently? What was the most important piece of information used to make your decision?
Courses of action Analogy and generalisation Course of action	If you were presented with the same situation, would you make the same decision? Were you at any time reminded of a previous experience which a similar/different decision was made? Was this the best possible decision? Did you seek guidance at the time? What other courses of action were considered? Did you consider the possible consequences of your decision?
Rules and recommendations	Do you think you could develop a rule, based on our experience, which could assist another person to make the same decision successfully? What recommendations would you make to the AFL in regards to the rules and regulations surrounding concussions?

RESULTS

Factors influencing decision-making

The factors identified were mapped onto Rasmussen's Risk Management Framework to show the level of the system at which they reside (see Figure 2). In Figure 2, the number of players reporting each factor is represented by 'P', coaches by 'C' and the medical team by 'M'. Within Figure 2, two scenarios are included. The first includes those in which the players played on through the concussion or came off the field and then returned to play. Frequencies associated with decision making in this scenario are represented in the round brackets (f). The second scenario, for which frequencies are shown in the square brackets [f], includes those in which the player was taken off the field and remained off the field.

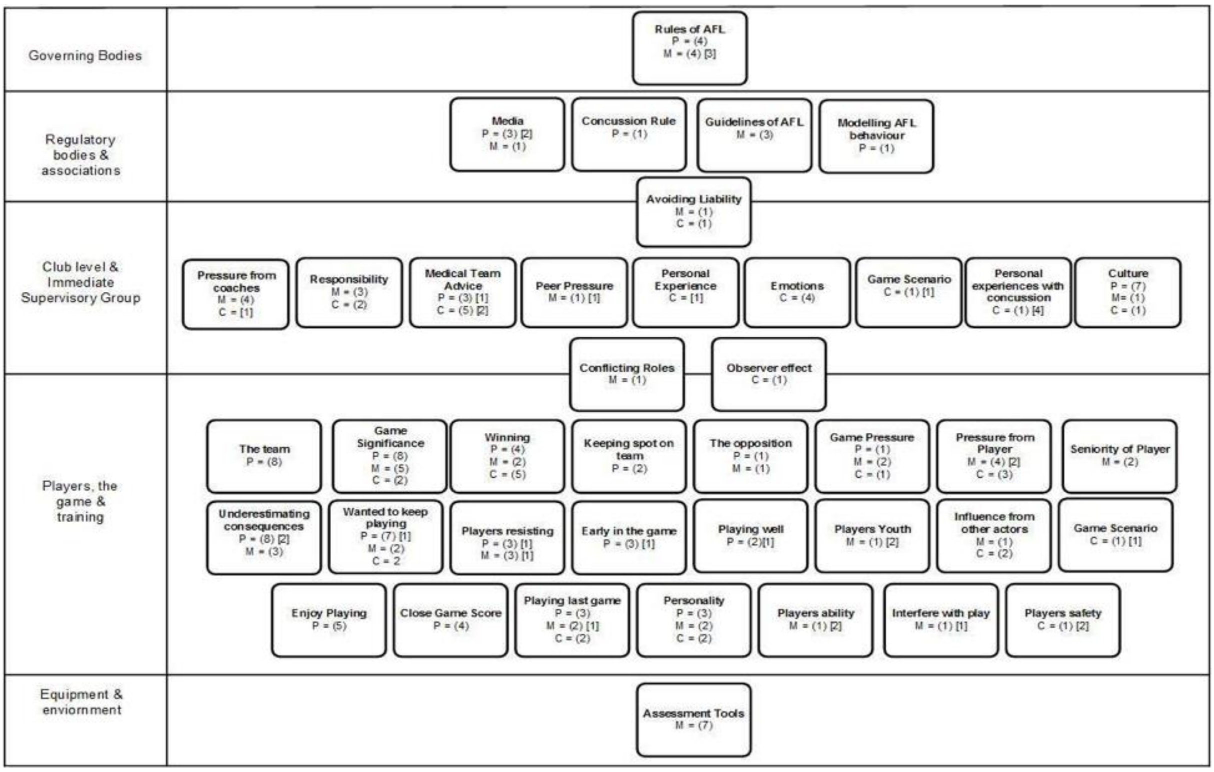


Figure 2. Frequencies of factors identified influencing the decisions of players, medical staff and coaches

Factors influencing player decision making

The most frequent factor reported to influence players' decisions to play on when concussed was an underestimation of the consequences of concussion. Following this, the most influential factors involved needing to continue playing their role for the team and 'game significance' such as a grand final. The majority of players (n = 7) reported that important games were a time when they would try to continue to play on after a concussion.

> "No matter what happens in a grand final I would have kept playing. No matter what happened in my last game I would of played" (Player 6)

> "This game was the first finals game and because it was finals I felt like I had to go back on and play" (Player 9)

The next most frequently reported factors included wanting to keep playing and game culture.

> "I think it was the way I was brought up you know when you get hit you get up and you keep going. I suppose that comes from the culture" (Player 6)

> "We try to be renowned for playing tough football; it's in the culture I suppose" (Player 3)

One player reported as a result of a concussion rule he did not want to allow himself to be in the position to be assessed by the medical team:

> "I really didn't want to go off because now with the concussion rule if you go off and they wanted to monitor you then you might not be allowed back on so you don't really want to go off in the first place" (Player 1).

Factors influencing medical team decision making

The most influential factor reported to impact the medical team's decision-making was the score provided by the concussion assessment tool, known as the Sports Concussion Assessment Tool (SCAT; McCrory et al., 2013).

This was followed by AFL rules, pressure from players, game significance, pressure from coaches, player age, and players resisting. Three out of four participants from the medical team (75%) stated that they are much more conservative when diagnosing and treating younger concussed football players.

One interesting finding was that, as a result of pressure from coaches, the medical team reported that they often find it difficult to communicate with the coach that a player is not returning to play.

> "Early in my career and certainly with a couple of coaches I would have had a tough time mentally going to the coach and saying he's not coming back. I have had coaches that were aggressive about it at every level" (Medical Team 3).

> "Coaches may push the boundaries a bit when they want a player to keep playing and even to the extent that the medical staff can be under pressure" (Medical Team 2)

The factors influencing decisions most frequently reported by the medical team included the AFL guidelines around concussion, player playing their last game, underestimating the consequences, formal responsibilities within their role and a player's ability. As AFL guidelines are reported frequently by the medical team this helped illustrate that some of the higher level factors are promulgating down and being reflected in the decisions and actions occurring at the lower levels of the system. Additional frequent themes identified included winning, game pressure, and that the player wanted to keep playing.

Factors influencing coach decision making

The coaches reported that they relied heavily upon the medical team's decisions and advice. Following this, winning and a personal experience with concussion were the next most frequent factors reported. One notable finding was that the coaches reported an awareness that their own emotions played a role in their decision-making and in one case this led the coach to remove themselves from the decision-making process.

> "I never got involved with the assessment because emotionally I wanted him to play" (Coach 1).

Other factors reported by coaches included pressure from players and concerns for the player's safety, game significance, playing the last game, players wanting to keep playing, personality of the player and formal responsibility. Lastly, game scenario, game pressure, an observer effect (i.e. whether somebody was watching them or not) and pressure from other coaches were also mentioned throughout the interviews with the coaches.

DISCUSSION

The findings from this study show that there are a diverse set of factors that influence decision making around concussion events, and that these factors reside at all levels of the AFL system, ranging from the game situation and club itself, to the higher governing body level (i.e., AFL game rules). A further important finding is that the factors influencing decision making differ across players, coaches, and medical staff. Discussion of the key findings in relation to each group is presented below.

Players

Underestimating the consequences of concussion was the most frequent factor influencing a player's decision. When asked if the players considered the consequences of playing on after suffering a concussion the majority responded that they did not. Over three quarters of the players interviewed reported that if they were in the same situation they would make the same decision to play on. In addition, over half of the players reported having minimal to no knowledge of concussion and its impacts. These findings suggest that there is a lack of awareness and consideration of the consequences of concussion and indeed of playing on following a concussion. Reinforcing this is the finding that most players were only able to identify one or two consequences of a concussion, while others could not name any. This is consistent with other findings in the literature, namely that players do not have a proper understanding of concussion symptoms and do not recognise the possible consequences (McAllister-Deitrick, Covassin & Gould, 2014; McCrea, et al., 2004). A key implication of this study is that improved player education on the symptoms and impacts of concussion is required.

Medical Team

Some of the social pressures reported to influence the medical team's decision making related directly to player and coaching staff behaviours. Although the AFL guidelines stipulate that medical trainers should not be swayed by the opinion of players or coaches, the findings suggest that medical teams experience such pressures in practice. This finding is similar to that of Anderson & Gerrard (2005), who found that physicians feel pressure from players and coaches to put players back on the field, and confirms Partridge's (2014) assumption that the same effect is

present within AFL. Situational game pressure was also described as being a strong influence on the medical team's decision making, indicating that game pressure may drive concussion misdiagnoses.
A related finding that has received relatively little attention in the literature is the medical team's assertion that aggression is a strong symptom of a concussion (McCrory et al., 2013). Whilst all other symptoms reported in this study have been identified in previous studies (Edwards & Bodle, 2014), aggression has not.

Central to many current debates in concussion management in many codes of contact football, 'final call authority' was also highlighted. This debate centres on the issue of who is the most objective person on the field to make the final call about concussion identification and management (e.g., Clacy et al., 2013; Partridge & Hall, 2015). This issue may be a result of the lack of guidance and rules set by the AFL, which state that first aid trainers may undertake concussion assessments. Further, funding constraints in amateur level football clubs could possibly restrict the availability of full trained medical personnel.

Coaches

The findings show that coaches have a heavy reliance upon the medical staff to manage players after a concussion and to make decisions regarding concussed players continued involvement in games. This is interesting, given the pressure mentioned by the medical team, as it indicates that the coach's behaviour towards the medical team may be counterproductive to getting information that is critical for their decision-making. One assumption as to why coaches may rely heavily upon the medical team identified was the shortcomings in their own concussion management knowledge. This is a concern to other amateur or community clubs throughout Australia that may not have the resources to acquire adequate medical support, and rely on the decision of coaching staff or others at the club. If coaches are in fact making final-call decisions, this is of further concern as emotion was identified as a significant influencing factor in the study.

Perhaps unsurprisingly, the coaches mentioned winning as a factor influencing decisions more frequently than the players and the medical staff. This is another factor that has not previously been identified in the literature. This finding suggests that the combination of a highly significant game and players and coaches with a strong desire to win will be problematic in terms of managing concussed players appropriately.

Implications for concussion identification and management in AFL

Taken together the findings indicate that there are a range of measures that can be taken to improve concussion identification and management knowledge. To address the lack of knowledge and education across all levels of the system the way in which messages around concussion are delivered to the lower levels of the system may require improvement. This would support a better understanding of the consequences of concussion at the lower levels (e.g. players, coaches). One approach could involve implementation of educational programs and resources across the systems to address this issue, such as workshops for players and coaches to improve their knowledge of concussion, its symptoms, and its consequences. Potentially social pressures could also be addressed through the use of educational programs, such as an 'in the other shoes campaign' which could show different stakeholders the perspectives and constraints of other stakeholders. This could also address issues with players and coaches attitudes towards medical staff. Other interventions could be used to promote culture change in clubs by emphasising the role that culture plays in the mismanagement of concussion. These should attempt to address the fact that players feel obliged to sacrifice their own health and safety for the team; for example, "mates don't let mates play concussed". By tapping into the mateship that is imbedded within the culture this could exploit an opportunity to redirect attitudes towards healthier decision making (Clacy et al., 2015).

At the higher level of the system, funding from governments or the league could be used to resource independent medical staff to attend games and conduct concussion identification and management activities, removing the pressures that can be felt when the assessor is linked to a club. With regard to funding at the amateur level another suggestion is to include a finical incentive for those clubs who are testing players' cognitive performance. This could be achieved by rewarding regular administration of the SCAT test. Finally, strengthening the AFL concussion guidelines to improve clarity around roles and responsibilities and providing clarification within those guidelines in relation to final call authority will introduce greater liability towards the management team. Developing and implementing a process for monitoring or auditing practice against the guidelines may assist in presenting an 'observer effect' to promote more conservative decisions around player safety.

The small sample size used in this study needs to be acknowledged as does the fact that only one club participated in the study. While the findings are interesting, caution should be urged when generalising them. A larger scale study incorporating more participants and clubs is recommended. Future research should also employ additional measures such as questionnaires or objective data from medical reports. Supplementary research could also involve other actors and organisations at different levels of the AFL system to identify the factors influencing their behaviour. Finally, it is proposed that the approach adopted may be applied to investigate other prominent forms of sports-related injury and other sports-related issues such as doping and corruption.

CONCLUSION

This study has taken a first step to identify and understand the factors influencing the decision-making of players, medical teams and coaches following a concussion in amateur AFL. The findings demonstrate that the identification and management of concussion in AFL is a complex issue that is influenced by a myriad of factors from across the overall AFL system. This includes factors that are related to the game, players, and coaches and also factors related to assessment tools, concussion guidelines and social influences. Now that some of these factors have been recognized, future research should aim to support removal of the negative factors to achieve better management of concussions in amateur AFL.

REFERENCES

AFL Guidelines. (2013). The management of concussion in Australian football. Melbourne. AFL, 2013. Retriever from http://www.aflcommunityclub.com.au/index.php?id=66

Anderson, L. C., & Gerrard, D. F. (2005). Ethical issues concerning New Zealand sport doctors. *Journal of Medical Ethics, 31*(2), 88-92. doi:10.1136/jme.2002.00836

Clacy, A., Sharman, R., & Lovell, G. (2013). Return-to-play confusion: Considerations for sport-related concussion. *Journal of Bioethical Enquiry, 10*(1), 127-8. doi:10.1007/s11673-012-9421-8

Clacy, A., Sharman, R., Goode, N., Salmon, P. M., & Lovell, G. P. (2015). Responsibilities in the prevention of concussion in community rugby union. *Procedia Manufacturing*, *3*, 1173-1180. doi:10.1016/j.promfg.2015.07.195

Cobb, S., & Battin, B. (2004). Second-impact syndrome. *Journal of School Nursing (Allen Press Publishing Services Inc.), 20*(5), 262-267 6262-267 6p. doi: 10.1177/10598405040200050401

Crandall, B, Klien, G, Robert, Hoffman, R, R. (2006) Working Minds – A practitioner's guide to cognitive task analysis. A Brafdord Book. London England.

Edwards, J. C., & Bodle, J. D. (2014). Causes and consequences of sports concussion.*Journal of Law, Medicine and Ethics, 42*(2), 128-132. doi:10.1111/jlme.12126

Goode, N., Read, G., van Mulken, M. R. H., Clacy, A. Salmon, P. M. (2016). Designing System Reforms: Using a Systems Approach to Translate Incident Analyses into Prevention Strategies. *Frontiers in Psychology*, 7, 1974

Hollis, S. J., Stevenson, M. R., McIntosh, A. S., Shores, E. A., Collins, M. W., & Taylor, C. B. (2009). Incidence, risk, and protection factors of mild traumatic brain injury in a cohort of Australian nonprofessional male rugby players. *The American Journal of Sports Medicine*, 37:12, 2328-2333.

Hsieh, H., & Shannon, S.E. (2005). Three Approaches to Qualitative Content Analysis.*Qualitative Health Research, 15*(9), 1277-1288. doi: 10.1177/1049732305276687

Klein, G. A. (1989). Recognition-primed decisions. *Advances in Man Machine Systems Research, 5*, 47-92. doi:10.1518/001872008X288385

McAllister-Deitrick, J., Covassin, T., & Gould, D. R. (2014). Sport-Related Concussion Knowledge Among Youth Football Players. *Athletic Training & Sports Health Care: The Journal For The Practicing Clinician*, *6*(6), 280-284.

McCrea, M., Hammeke, T., Olsen, G., Leo, P., & Guskiewicz, K. (2004). Unreported concussions in high school football players: implications for preventions. *Sports Medicine, 14*, 13-17. doi:10.1097/00042752-200401000-00003

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R. J., & Turner, M. (2013). Consensus statement on concussion in sport: the 4th international conference on concussion in sport held in Zurich, November 2012. *British Journal of Sports Medicine,* 47(5), 250 – 258. doi: 10.1136/bjsports-2013-092313

Neville, T. J., Salmon, P. M., Read, G. J. M., & Kalloniatis, A. C. (2016). Play on or call a foul: testing and extending distributed situation awareness theory through sports officiating. *Theoretical Issues in Ergonomics Science, 17*(1), 80-103. doi:10.1080/1463922X.2015.1106617

Newton, J. D., White, P. E., Ewing, M. T., Makdissi, M., Davis, G. A., Donaldson, A., & Finch, C. F. (2014). Original research: Intention to use sport concussion guidelines among community-level coaches and sports trainers. *Journal of Science and Medicine In Sport*, *17*, 469-473. doi:10.1016/j.jsams.2013.10.240

Partridge, B. & Hall, W. (2015) Repeated head injuries in Australia's collision sports highlight ethical and evidential gaps in concussion management policies. *Neuroethics*, *8*(1), 39-45. doi: 10.1007/s12152-014-9217-0

Partridge, B. (2014) Dazed and confused: sports medicine, conflicts of interest, and concussion management. *Journal of Bioethical Inquiry*, *11*(1), 65-74. doi:10.1007/s11673-013-9491-2

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, 27:2/3, 183-213.

Salmon, P. M., Stanton, N. A., & Gibbon, A. C. (2010). Human factors methods and sports science: a practical guide. Taylor & Francis, Boca Raton, USA.

How Analysts Think: Navigating Uncertainty – Aspirations, Considerations and Strategies

CELESTE GROENEWALD, B.L.WILLIAM WONG, SIMON ATTFIELD, PETER PASSMORE and NEESHA KODAGODA
Middlesex University, London, UK

ABSTRACT
While there has been considerable research in understanding the process of sense making in criminal intelligence analysis as well as the formulation of arguments in the domain of law, there remain gaps in our understanding of how to move seamlessly from the fluidity of the sense making activities to the rigour of argumentation construction. This matter is complicated further with high uncertainty which accompanies sense making and which propagates through to the rigour that accompanies argumentation. This paper attempts to understand how Criminal Intelligence Analysts navigate uncertainty from fluidity to rigour constructs and outlines some of the considerations and strategies deployed by the Criminal Intelligence Analyst to reach, or increase, certainty at a given point in time during the analysis process. This paper concludes by proposing preliminary suggestions with the aim to narrow the gaps in this journey from fluidity to rigour, at least, marginally.

KEYWORDS
Uncertainty; Sense making; Security; Government and Law

INTRODUCTION

"How Analysts Think" is a series of papers aiming to enhance our understand on how Criminal Intelligence Analyst's think, with the aim to be better informed on how to design the software they use as part of their daily activities. In Wong's (2014) paper, he introduced the concepts of sense-making fluidity and rigour. This work continued with Wong and Kodagoda's (2015) paper, which describes the inference making strategies in the analytic reasoning process and how uncertainty propagates through the analysis of criminal intelligence cases. Following on from this, Gerber et al.'s (2016) paper, proposed a framework on how intuition, leap of faith and insight occurs during Criminal Intelligence analysis. The research in this paper continues with this series and investigates how Criminal Intelligence Analysts (from now on referred to as Analysts) navigate through uncertainty. Navigating uncertainty during sense making is the process by which Analysts uses their expert skills to continually identify potential sense making problems, understand how it hinders them from reaching their aspirations and taking into account the considerations which affords the undertaking for the best possible course of action.

In this paper, we describe the Analysts' aspirations, considerations and objectives for each of the corresponding strategies employed during analysis to reach points of certainty. These points of certainty serve as possible anchoring points, which afford traction for further analysis or the launch of new lines of enquiry. Lines of enquiry can be described as facts or assertions, which are actively being pursued by Analysts or investigative officers to gain clarity on the level of truthfulness and are described as being, "specific to key facts on an investigation" (Cook *et al.*, 2013). A qualitative study using transcripts from five Criminal Intelligence Analysts reveals the problems that hinder their sense-making efforts during a variety of analytical tasks and the strategies used to resolve each. These problems relate to their mental efforts as well as the environment in which they work. Each problem hinders the Analysts' aspirations and the inverse of each problem could potentially reveal details on what those aspirations are. By studying the relationship between the strategy used and the attained aspiration, the potential considerations that have to take place before deploying the appropriate strategy are revealed. Similarly, the relationship between the aspiration, consideration and strategy reveals the potential objectives that serve as a tool to judge the level of certainty surrounding each outcome. This reveals the compound nature of certainty as a collection of multiple influential factors, which are expressed in this paper as eleven distinct aspirations. Due to space limitations, we are only able to cover three of the eleven aspirations and they are: certainty, believability and plausibility. The next section briefly outlines some of the relevant literature that inspired this research.

LITERATURE

At the onset of a case, Analysts seldom have adequate information they require to reach a conclusion. Instead, they rely on analytical, inference and sense making processes to guide them through the uncertainties they face. Pirolli and Card's (2005) Notional Model of Analytical Sense Making provided researchers with a model, clearly illustrating the complexities of the processes involved to move from raw data to presenting a conclusion to the decision makers. Wong (2014) describes similar complexities, but concentrates on the type of thinking required by Analysts as they move from the fluid to the rigour stages of sense making. According to Wong (2014), in the fluidity phase of the sense-making activities, the stories that the Analysts construct are based on a loose assembly of data with many uncertainties and the commitment to follow through on outcomes are low. In this phase, Analysts would rely on creative thinking methods, which in turn, could leverage them into being able to generate ideas or develop a better understanding of the situation or problem and thus gain traction to progress the investigation. As the Analysts gain better understanding of the investigation in the form of new ideas or evidence, the Analysts enter the rigour stage of the sense-making process, where certainty is high and the commitment to follow through with outcomes are also high. This process is not linear and can be chaotic and cyclical, depending on the available data, goals to satisfy (e.g. to gain traction or to prove), the desired claims to be made, and their experience and/or state of knowing at that time. Wong and Kodagoda (2015) further investigated how the analytical inferencing process could invite uncertainties to propagate from one conclusion to another, thus illustrating that the final outcome should not only be tested against weak evidential sources, but also of instances where inference propagation took place.

Klein et al. (2007) suggested that the sense-making process of gaining traction involves the use of anchors and defined anchors as, "key data elements that serve to create understandings that guide subsequent inquiry". Wong and Kodagoda (2015) added to this definition by stating that they believe that "data elements can also be non-data, such as suppositions where no data exists or is ambiguous as is often the case in intelligence analysis, and are used for the purpose of gaining traction." They further elaborated this concept by stating that anchoring is the, "process of using anchors to create and evolve understanding". From the perspective of the legal domain, Wagenaar et al. (1993, p39) refer to the concept of anchors in their anchored narrative model as; "ordering the evidence in such a way that it forms anchors between the story and the ground of accepted common-sense rules". They then continue by saying that, "all critical episodes in the story need to be anchored through evidence". Bex and Verheij (2013) built upon Waggenaar et al.'s concepts and subsequently created a Hybrid theory and explain that, "facts are organised into multiple hypothetical stories, coherent accounts of what might have happened in the case. Arguments based on evidence can then be used to justify these stories, as these arguments can be used to support elements in a story with evidence or, in other words, to anchor the story in evidence (cf. Wagenaar et al. 1993)". Therefore, the sense making domain concentrates on finding traction using data and non-data elements as anchors and the law domain concentrates on an argumentation approach, which requires the justification for the interpretations of derived facts. When these interpretations are grounded in evidence or common sense rules, then those interpretations are considered strong anchors. The distinction between the use of the term anchor in the domain of sense making and law are complementary and necessary as it may reduce confusion when the same term is used in different contexts.

The Analysts' conclusions should adhere to the highest degree of certainty. Any unresolved issues or doubtful conclusions could undermine the prosecutor's case during court proceedings. This is outlined in the Association of Chief Police officers (ACPO) Core Investigative Doctrine as, "Hypotheses that are formed from limited or uncertain information can, at best, only amount to an assumption of what may have occurred and this could be influenced by personal bias or stereotyping" (ACPO, 2005). It is therefore of interest to understand which sense making problems can hinder Analysts from reaching outcomes of the highest possible degree of certainty, especially if those outcomes are used as sense-making anchors to gain traction or to launch new lines of enquiry. The next section outlines the research conducted and what the results revealed for answering RQ1 – RQ5.

METHODOLOGY

In this paper, we report on the analysis of transcripts from Cognitive Task Analysis (CTA) interviews with five experienced Operational Criminal Intelligence Analysts. The interviewers, using the Critical Decision Method (Klein et al., 1986), investigated the inference and sense making processes of the Operational Criminal Intelligence Analyst participants from different police forces in the UK and Belgium. The interviewers wanted to understand how each Analyst resolved a particularly memorable case. Participants are more able to recall the details associated with a memorable case and the influencing factors it had, than just an ordinary case. Volume crime and serious crimes are two types of case classification that Analysts work with. The National Policing Improvement Agency (NPIA) within the Volume Crime Management Model (VCMM) defines volume crime as, "any crime which, through its sheer volume has a significant impact on the community and the ability of the local police to tackle it. Volume crime often includes priority crimes such as street robbery, burglary and vehicle-related criminality, but can also apply to criminal damages or assaults" (College of Policing, 2009). The Police Act of 1997 define serious crime as, "(Section 93.4a) [the crime] involves the use of violence, results in substantial financial gain or is conduct by a large number of persons in pursuit of a common purpose or (Section 93.4b) the offence or one of the offences is an offence for which a person who has attained the age of twenty-one and has no previous convictions could reasonably be expected to be sentenced to imprisonment for a term of three years or more ” (Legislation.gov.uk, 1997). The National Intelligence Model (NIM) defines the aim of crime analysis as, “to interpret a range of information to develop inferences, which are conclusions about what is known or what is believed to be happening” (Centrix, 2007). A third party anonymised, transcribed and reviewed the transcripts due to the sensitivity of the contents it contained. A third researcher performed the data analysis on five of the transcripts from the CTA interviews.

Crandall et al. (2006) outline the typical phases of data analysis of CTA interviews to be: preparation, data structuring, discovering meanings and representing findings. Their recommendation is to make multiple passes through the data in order to gain the most out of the richness and complexity of the data set. The first pass through our data set was to answer RQ1 and RQ2. By using the Open Coding technique as part of Grounded Theory (Corbin and Strauss, 1990), we used “problem” and “strategy” as high-level codes to identify each in the available transcripts. The results underwent re-analysis in order to sort them thematically into groups based on their characteristics. The members of each group were re-analysed and compared with other group members to insure correct assignment. When members were misplaced, we moved them into more appropriate groups. This process reiterated until saturation. This resulted in eleven distinct groups namely, Uncertainty, Skeptisism, Suspiciousness, Complexity, Obscurity, Disparity, Gaps, Misconceptions, Exhausted Options; Errors (Data Quality) and Mental Blocks. Table1 outlines each of the identified problems with their corresponding strategies and they relate to both the Analysts’ mental efforts as well as the environment in which they work. After determining the results for RQ1 and RQ2, we realised that further questions remained and we added RQ3 and RQ4 to determine why and how Analysts applied particular strategies. We used the results from RQ1 and RQ2 for the analysis of RQ3 and RQ4. Our final question (RQ5) was to determine if and how it would be possible for Analysts to verify that a sense-making problem has been successfully resolved, so we used the findings from RQ1-RQ4 to achieve this.

The next section outlines the eleven different sense-making problems we uncovered in our analysis. Due to space constraints, we chose to elaborate on what we consider as three significant problems for Analysts.

RESULTS

This research attempts to answer the following five research questions: **RQ1:** What are the problems that hinder sense making? **RQ2:** Which strategies do Analysts use to overcome each problem? **RQ3:** Why do Analysts apply specific strategies? **RQ4:** How do Analysts know which strategy to apply? **RQ5:** How could the Analyst judge if a strategy resolved a particular problem? Below is a summary of the findings for each question.

RQ1: What are the problems that hinder sense making? The research identified Uncertainty, Skeptisism, Suspiciousness, Complexity, Obscurity, Disparity, Gaps, Errors, Misconceptions, Exhausted Options and Mental Blocks, as possible problems that could hinder sense making.

RQ2: Which strategies do Analysts use to overcome each problem? The research identified that for problems in uncertainty, skeptisism and suspiciousness, could be overcome by using the following strategies : Resolving ambiguity/doubt; Establishing certainty points ; Strengthening the evidence ; Strengthening their reasoning process; Considering alternatives; Increasing the understanding of outcomes; Increasing the confidence in processes; Questioning the facts; Questioning system outputs; Using simplistic two-way tests to resolve anomalies; Merging multiple confirmations into one concept.

RQ3: Why do Analysts apply specific strategies? To answer this question, the researchers revised each problem and it became apparent that the inverse of each problem represents Analysts’ aspirations. This revealed that the selection process for using a particular strategy was influended by Analysts’ aspirations at various moments during the analysis. The identified aspirations are: Certainty, Believability, Plausibility, Simplicity, Clarity, Creativity, Connectivity, Identifying New Possibilities; Identifying Meaning/Information; Determining Correctness (Data Quality); and Increasing Understanding.

RQ4: How do Analysts know which strategy to apply? To answer this question, the researchers investigated the relationship between the aspiration (RQ3) and the deployed strategy (RQ2) and inferred the considerations the Analysts may have had which allowed them to narrow down their choices. The considerations for certainty, believability and plausibility are expressed as questions and they are summarised in Table 2.

RQ5: How could the Analyst judge if a strategy resolved a particular problem? To answer this question, the researchers investigated the relationship between the aspirations (RQ3), considerations (RQ4) and strategies (RQ2) and outlined potential objectives that could serve as a method to judge if a particular problem has been successfully resolved. The objectives for certainty, believability and plausibility are summarised in Table 2.

The next section describes the aspirations: certainty, believability and plausibility in detail.

Certainty, Believability and Plausibility

This section discusses the research results in more detail for the Analysts' aspirations of certainty, believability and plausibility. Table 2 offers a summary of these results. The explanations below outline the meaning of the Analysts' aspirations and the strategies employed to reach each of the aspirations.

Each of the strategies are then explained and start off with what Analysts could have considered at that moment in time in the analysis phase, given the data that was available. This is followed with the relevant interview extracts. Interview extracts are in italics and follow the format of the participant number, the relevant lines in the interview and finally the interview extract. Extracts have been desensitised to be gender and location neutral, due to the sensitivity of the data set.

Table 1 – Summary of the research results for RQ1, RQ2 and RQ3

#	Problem (RQ1)	Aspiration (RQ3)	Context of Example	Example	Example of a strategy used to reach the aspiration (RQ2)
1	Uncertainty	Certainty	Modus Operandi (MO) Details	(P2:267-275) "...I couldn't do anything because we had nothing to go on further, only a car and a figurative name... could be an anagram…The phone… they can see where they are living and then can go and look in the neighbourhood if a name occurs over there..."	Task officers to ask for the information within the neighbourhoods
2	Skeptisism	Believability	Findings during analysis	(P3:488-490) "...Whether [colleague] could see anything different to me or just what I can see, reassurance that checking that [colleague] agrees, seeing if there's anything... different that s/he would look at..."	Talk to other colleagues to confirm analysis/findings
3	Suspiciousness	Plausibility	Hotspots on a map	(P3:353-354) "...Shopping centres and things like that will show higher than the things that you actually want..."	Question how plausible it is for a hotspot to be significant on a specific region on a map
4	Complexity	Simplicity	Starting a case	(P12:2) "...What do we know about this one? We are checking everything…police reports.., everything…this we don't know… this is a probability… this is unknown… and now we know exactly what is what…at that time we set it all out…"	Structuring information into manageable pieces
5	Obscurity	Clarity	Too much data	(P1.353-355) "...I just ask my database how many times he has been there…Day minus one, day minus two, day minus three…"	Comparisons between data sections
6	Disparity	Connectivity	Unknown relationships	(P12:36) "...We need to know if it is a series... If it is an organisation behind it..."	Create relationships to other data points
7	Mental Blocks	Creativity	Unable to resolve the problem at hand	(P3:184-185) "...we do go out and see what the areas are like where the offending is happening..."	Use creative methods to see the situation with fresh eyes
8	Exhausted Options	Finding New Possibilities	Normal solutions does not fit the problem	(P3:157-158) "...Look at a few different places that weren't necessarily that close, but there might be arterial routes if the offenders where using cars..."	Increase the scope of the analysis
9	Gaps	Finding Meaning / Information	Too many gaps in the data to identify offenders	(P6:83-86) "...Where we're having smash and grabs, that's people smashing the windows, just highlighting to officers…stop and search people within the area, see what tools they've got on them that would be useful to break a window..."	Enrich the available data with new information
10	Errors (Data Quality)	Determining Correctness (Data Quality)	Data errors	(P6:442-443) "...The issue I had with offender B was there were too, there was files created for him/her so all of his/her information was split over different links..."	Verify and correct data errors
11	Misconception	Increasing Understanding	Information does not conform to the Analysts mental model of the crime	(P1:185-191) "...I need to know more about the [parent] and I ask the detective squad about the [parent]…"	Talk to other colleagues to obtain more information when an offender's behaviour is not as expected

Certainty

Certainty refers to the degree of correctness or truth. This is more easily described in terms of physical properties or scientific results. These are things that Analysts can easily determine as it is available in front of them – in black and white so to speak. An example is: It is certain that the time of death is 9pm, as it is the time that the officer reported shooting the offender in self-defence. Certainties most likely would require to be underpinned by data. This in turn requires good quality and reliable data. Certainty refers to both the data as well as the Analyst's mental state. The strategies employed by Analysts to increase the certainty were through the actions of: (1) Resolving ambiguity/doubt, (2) Establishing certainty points and (3) Strenghtening the evidence. The section below discusses each strategy in more detail.

(1) Resolve ambiguity / doubt: One of the strategies that Analysts can employ to increase certainty, is to resolve ambiguity in the data in order to ensure that Analysts knows exactly who or what they are working with. In this example the Analyst is unsure if the name in the crime report is an anagram or if the offender uses his/her real name. To increase certainty on this matter, the Analyst asked the officers to confirm the details in the offender's neighbourhood and thus lends to resolving the ambiguity.

> (P2:267-275)"...I couldn't do anything because we had nothing to go on further, only a car and a figurative name... could be an anagram…The phone… they can see where they are living and then can go and look in the neighbourhood if a name occurs over there..."

(2) Establishing certainty points: In some cases, there is either too little or too much data and the Analyst is required to focus on a particular entity, which offers the most certain information from which to gain traction. In this example the Analyst uses the phone found on the victim, who in this case is also a known offender, as a starting point to establish the identities of the other possible offenders by looking at the previous calls made.

> (P12:15)"...And you might see that there aren't any communications going out anymore… because s/he is dead… if there is incoming, okay that is good…but, I have to establish what happened here… if there is no more outgoing communications and the phone was found in his/her pocket then I say... the phone belongs to him/her...at this point in time..."

(3) Stengthening the evidence: The Analyst can go one step further (from the outlined point under number 2) and strengthen the certainty by resolving the remaining doubt on whether or not the phone does belong to the victim, by cross-checking the phone's location with the location in which the victim lives. In doing so, the Analyst has successfully created a certainty point from which to drive the investigation with enough evidence to support his/her conclusion.

> (P12:16)"...What I also do is find out where does s/he live... If s/he is not illegal, then s/he has a house...and every house (belongs) a telephone mast...and if I say that most of the communications from that mast is in the morning and evenings... Then I can say that s/he lives there..."

Table 2 - Summary of the research results for RQ1-5 for the Analysts' apirations of: certainty, believability and plausibility

Problem (RQ1)	**Aspiration (RQ3)**	**Consideration (RQ4)**	**Objective (RQ5)**	**Strategies Deployed (RQ2)**
Uncertainty	Certainty	How certain am I that the data is clear to me and that I understand what I see or read?	Demonstrate understanding and data clarity	Resolve ambiguity/doubt
		Which details are the most certain at this point in time, that I can use to progress my case?	Differentiate between what is certain and uncertain	Establish certainty points
		How can I increase the certainty of particular findings/details?	Demonstrate highest certainty level of findings/details	Strengthening the evidence
Skeptisism	Believability	How believable is it that the findings/analysis is correct or showing what I think it is showing?	Demonstrate a verified reasoning process	Strengthening the reasoning process
		How believable is it that the current approach is correct or alternatives could be playing a role?	Demonstrate a verified approach	Consider alternatives
		How believable is it that the current nominal is involved based on the data that points to this nominal?	Demonstrate verified outcomes	Increase understanding of outcomes
		How believable are the documents / analysis / findings that are being handed down from one person to another?	Demonstrate a verified process	Increase confidence in processes
Suspiciousness	Plausibility	How plausible is it that the current nominal is the offender and that the evidence is not super-imposed?	Demonstrate impartiality	Questioning the facts
		Is it plausible that the system is giving me false positives given this information?	Demonstrate objectivity	Questioning system outputs

	What is the most plausible option given the current anomaly?	Demonstrate decision making and choices	Use simplistic two-way tests to resolve anomalies
	How plausible is it that different people/objects/elements in different places are actually the same thing?	Demonstrate confirmations obtained	Merge multiple confirmations into one concept

Believability

Believability is the likelihood of something being true. It can be believable that offenders re-offend in Local Police Units (LPUs) after being released from prison. Is that a certainty? No, because it is not true for all offenders and other factors might also play a role, such as offenders migrating between LPUs. These instances are most likely underpinned by Analysts' general beliefs, domain and/or personal experiences. This in turn requires verification through further analysis or confirmation from external sources such as the police officers patrolling the streets. Believability refers to both the data as well as the Analyst's mental state. The strategies employed by Analysts to question the believability were through the actions of: (1) Seeking reassurance, (2) Considering alternatives, (3) Increasing their understanding on a particular topic and (4) Increasing their confidence level of the work they produced. The section below explains each strategy in more detail.

(1) Seeking Reassurance: The Analysts can reassure themselves that their reasoning is valid, by reducing any doubt they have in respect to their own capabilities as well as the data. This was achieved through increasing their confidence in the believability of their findings as well as confirming assumptions about data. Both routes involved the gathering of additional information, but were accomplished through different strategies. **(1a)** To reassure themselves of the believability of their findings, the Analysts brainstormed with other colleagues to get their point of view on the subject and thus strengthened their own reasoning and beliefs. This in turn increases the overall certainty of the analysis performed.

> (P3:488-490)"...whether [colleague] could see anything different to me or just what I can see, reassurance that checking that [colleague] agrees, seeing if there is anything... different that [colleague] would look at…"

(1b) To confirm assumptions made on the data, the Analysts used a probing activity by tasking officers to keep an eye out for particular details and to report back on their findings. Therefore, Analysts may believe that something is pointing to an offender, but they do not have the evidence to prove it. So although believable, it is not certain.

> (P6:361-367)"…Here's the intelligence we need [officer] to gather, here are offenders we specifically want [officer] to look at because they are who we think might be doing it – we've got the evidence pointing that way, intelligence as to say, pointing that way, but there's nothing concrete..."

(2) Considering Alternatives: The Analysts will try to increase the believability of their work through considering alternative pathways (possibilities or explanations) thus clarifying any assumptions they may have about the data. The question here is if it is believable that the offenders could travel between hotspots. The Analyst can achieve this by testing the probability of the assumption being possible. This is achieved through an internal questioning process which allows them to build a likely profile of an offender and/or the location.

> (P6:381-386)"…I looked at what crime they'd been arrested for and then looked at what, where that crime took place, if any took place within the location or – 'cause there was only one within, near my hotspot, I then think oh, maybe it's not too difficult for them to go here to there…or were any of these crimes a very similar MO for a specific MO that I was looking at..."

(3) Increasing Understanding: Nominal is a term used by UK Police to refer to a person who is a witness, suspect, offender or a victim. Analysts could try to increase the believability of a nominal as an offender, by increasing their understanding of the circumstances surrounding that nominal. This is accomplished by expanding their horizon by branching out and asking alternative questions. This could result in more information and thus a deeper understanding of the nominal and their actions within a particular scenario. This questioning process is performed as part of an internal dialogue and will not necessarily be externalised.

> (P1:396-397)"...Every move he made… Every communication… Why him?..."

(4) Increasing Confidence: The Analysts are not necessarily the first to analyse the data. They may have doubts about how exhaustively prior work has been completed, i.e. no stone was left unturned. The Analysts do this by questioning the validity of what they observe by double checking the documents / processes.

> (P6:159-168)"…I want to know what the officer's done already, if s/he's, you've looked at the pattern and if you've done exactly what it's found to do then you can do it again, or if I don't believe you've done it thoroughly enough I'll do it again. Erm, I need to know what you've done, so I went through, checking to see have they, [System Searched] him, have they found, erm, because I mean looking, it was looking through these notes that I found out that they hadn't got the property from him/her – so I didn't know that before, it didn't state that in the MO…"

Plausibility

Plausibility is the degree to which something makes sense. It is believable that offenders re-offend in a LPU, but is it plausible that offender A could travel to all three locations in under half an hour? These instances are most likely underpinned by the reasoning capabilities of Analysts based on the extent of their knowledge of the how the world works and especially their domain. Plausibility refers to how much the information or analysis conducted makes sense. The strategies employed by Analysts to judge the plausibility of a situation were: (1) Questioning the facts, (2) Questioning system outputs, (3) Using simplictic two-way tests to resolve anomalies and (4) Merging multiple confirmations into one concept. The sections below explain each strategy in more detail.

(1) Questioning the facts: Analysts ensures that the investigation is based on the available facts and that an offender is not wrongly acused of a crime based on previous convictions. Although the Analysts may already have knowledge of the offender's previous offences, which makes it plausible that the offender is involved again, Analysts must consider the information in front of them and not superimpose their beliefs on the offender.

> (P6:266-280)"...stereotyping ... I mean they are innocent until proven guilty so we can't just start looking – er, well, you always do it so it's your fault, so I'm just gonna put everything on you or find stuff to prove that it's you, I need to come from an unbiased point of view... previous offending can't count at all – it does in terms of my thought process but it doesn't in terms of my analysis, I'm still gonna take the exact same analysis regardless of – this offender I've never heard of, but even if I had I'd be doing exactly the same things in a way to make sure I've not superimposed him/her in that position as opposed to him/her actually being there..."

(2) Questioning System Output: The Analyst is unsure if the visualisation is actually producing the correct presentation. The visualisations are driven by the underlying data which can produce false positives. In these cases the Analyst is required to question the plausibility of outputs and consider factors which could be influencing the outputs.

> (P3:353-354)"...Shopping centres and things like that will show higher than the things that you actually want..."

(3) Using simplistic two-way tests to resolve anomalies: To more quickly progress the analysis, the Analyst creates a two-way test to judge which option is more plausible given the current anomaly as presented in the data. The path which is most plausible can be followed first in the anlaysis. In the example **(1a),** it is more plausible that one offender is being driven to the crime location, than not.

> (P12:25)"...I see him communicating with this number... driving to this phone mast...the communication stops...gap...so hypothesis... He is calling him to say that I am going to fetch you or not..."

Example (1b) is testing the involvement of an offender given the data. In this instance the Analyst judges the plausibility of the offender being involved based on his/her behaviour moments after the offence took place. Given the data that the offender called his/her spouse to possibly inform what had happened and that he/she is required to go silent (lay low) for a while, increases the plausibility that the offender had something to do with the event, than not.

> (P12:29)"...Hypothesis... Is this (P)...probably as the phone mast covers the whole region of (P)... The communication with the [spouse] and then the silence after the event..."

(4) Merging multiple confirmation into one concept: The Analyst can combine data based on the plausibility that it is the same object, appearing in different contexts.

> (P12:33)"...When I have two or three confirmations that the number is (P), then I'll merge them..."

Plausibility and Believability can work with each other or against each other. When working with each other it could move the Analyst towards clarity and understanding and eventually towards certainty. When working against each other it creates contradictions. The next section discusses the implications of the research findings and how it could influence design considerations for applications in sense making for Criminal Intelligence Analysis.

DISCUSSION AND CONCLUSION

This research identified eleven aspirations that can influence the collective level of certainty surrounding the outcomes of analytical sense-making activities. This paper outlined three of those aspirations in detail. The results section outlined various components, which could be present during a typical analytical sense-making activity. As sense-making problems surface during the analytical process, they present the Analyst with sense- making blockages, which increase the collective level of uncertainty. These blockages hinder Analysts from moving from where they currently are to where they want to be - which is to be on a more certain foothold. To work through a particular sense-making problem, the Analyst could use various *strategies*. During the sense- making activity, Analysts do not apply the strategies randomly, but uses their current *aspirations* to select the best one. Each aspiration reveals the current mind set of Analysts, indicating what they are striving for, but cannot reach until they have overcome the problem. When reached, each aspiration adds to the collective level of certainty. In order to apply the appropriate strategy, Analysts have various *considerations*, which they need to address. These considerations could be in the form of questions, which Analysts could ask internally when faced with a problem. This should subsequently point them in the right direction on which strategy to apply. A corresponding objective can serve Analysts with a method to judge if the problem has actually been successfully resolved, given the outcomes.

In navigating uncertainty, Analysts use their expert skills to continually identify potential sense-making problems, understand how they hinder them from reaching their aspirations and take into account the considerations which affords the undertaking for the best possible course of action.

By considering why each strategy was applied (as to only which ones exist), we are able to differentiate between the multitude of factors that influence certainty (in its collective form) during sense-making activities. This differentiation may be important when designing software for assisting sense making in Criminal Intelligence Analysis, as it may allow for additional sense-making affordances. If it is possible for an Analyst to be aware of the changing levels of certainty within current analytical activities, along with the influx of additional information, then it could serve as reflectors marking the way through a given task, as and when the Analyst is performing it. This affords similar functionality to cat's eye reflectors marking a road for a motorist during poor visibility, affording them the ability to navigate safely.

One possible way to achieve this is by borrowing the concepts of perspectives (Bex and Verheij, 2013) from the law domain and by introducing a new certainty perspective, which outlines the certainty levels for each different type of aspiration as per the considerations made at each stage. The objectives accompanying each consideration and matching strategy can serve as a method to judge the level of certainty. Rigid argumentation structures may force Analysts to use all of the outcomes if used from the onset, so a low-commitment option may be more desirable which is what a certainty perspective could offer. Analysts would then have a way to judge the most certain outcomes, at that point in time during the analysis, allowing them to effectively progress their lines of enquiry without over committing to the outcomes. Using the certainty levels to judge "proven" outcomes could allow for more effective argumentation using a factual perspective during the rigour stages of sense making.

Further research is required to establish if the certainty perspective can be reached from the onset or if addditional perspectives are required (illustrated by a question mark in Figure 1). Figure 1 depicts the roles of perspectives in their respective areas. Factual and Legal Consequences perspectives form part of the rigour part of sense making in the legal domain where the facts or consequences are known and the commitment to use in an argumentation structure or schema is high. A certainty perspective has varying levels of certainty and these certainty levels could evolve over time as the Analyst discovers new information. It therefore requires the Analysts to have low commitment to outcomes as they are still in the exploration stage of the analysis.

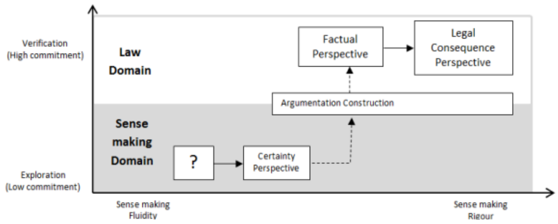


Figure 1 - Depiction of the different perspectives and how they could relate between domains

ACKNOWLEDGMENTS

The research results reported here has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) through Project VALCRI, European Commission Grant Agreement N° FP7-IP-608142, awarded to B.L. William Wong, Middlesex University and partners. We are also very grateful for the enthusiasm of the police analysts in sharing with us their experiences that made this report possible. Celeste Groenewald received a VALCRI PhD studentship to conduct her reasearch at Middlesex University.

REFERENCES

ACPO (2005). Investigative and evidential evaluation (extract from ACPO (2005) Core Investigative Doctrine). Available at: http://www.app.college.police.uk/app-content/investigations/linked-reference-material/

Bex, F. and Verheij, B. (2013). Legal stories and the process of proof, Artificial Intelligence and Law, Volume 21, Issue 3, Pages 253-278, ACM

Cook, T., Hibbitt, S. and Hill, M., (2013). Blackstone's Crime Investigator's Handbook. Oxford University Press.

Centrix (2007). Practice Advice: Introduction to Intellegence-led Policing. Centrix. Available at: http://www.fairplayforchildren.org/pdf/1291430265.pdf

College of Policing (2009). Practice Advice on The Management of Priority and Volume Crime: The Volume Crime Management Model. Second Edition. NPIA. Available at: http://library.college.police.uk/docs/acpo/VCMM-191109.pdf

Corbin, J.M. and Strauss, A., (1990). Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative sociology, 13(1), pp.3-21.

Crandall, B., Klein, G., Hoffman, R. (2006). Working Minds – A Practitioner's Guide to Cognitive Task Analysis, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England

Gerber, M., Wong, B.L.W., Kodagoda, N. (2016). How Analysts Think: Intuition, Leap of Faith and Insight. HFES

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting, 1, 576–580. Cited in Klein, G. (2008). “Naturalistic Decision Making”. HUMAN FACTORS, Vol. 50, No. 3, June 2008, pp. 456–460. DOI 10.1518/001872008X288385.

Klein G., Phillips J. K., Rall E. L., Peluso D. A. (2007). A Data-Frame Theory of Sensemaking. In Hoffman R. R. (Ed.), Expertise out of context Proceedings of the Sixth International Conference on NDM (pp. 113–155). Erlbaum.

Legislation.gov.uk (1997). Police Act 1997. Part III Authorisation of Action in Respect to Property. Section 93. Available at: http://www.legislation.gov.uk/ukpga/1997/50/section /93

Pirolli, P. and Card, K. (2005). The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, Conference: Proceedings of International Conference on Intelligence Analysis, Available at: https://www.researchgate.net/publication /215439203_The_sensemaking_process_and_leverage_points_for_Analyst_technology_as_identified_throug h_cognitive_task_analysis

Wagenaar, W., Koppen, P., Crombag, H. (1993). Anchored narratives : the psychology of criminal evidence

Wong, B. L. W. (2014). How Analysts Think (?): Early Observations. In 2014 IEEE Joint Intelligence and Security Informatics Conference (pp. 296–299). IEEE

Wong, B. L. W. and Kodagoda, N. (2015). How Analysts Think: Anchoring, Laddering and Associations. HFES

Making Sense of Magic

Simon HENDERSON[a]
[a]*Cranfield University*

ABSTRACT
This paper discusses the sensemaking processes a spectator experiences when they are fooled intensely by a strong magic effect. The paper focuses on the precise 'moment of magic'– the point at which a spectator's sensemaking process itself breaks down, and, for a moment, their understanding of reality fails. Factors that moderate the magnitude of this moment are identified, and implications for better utilizing such factors to create similar disruption of sensemaking in other domains are considered, including within the military, the arts and entertainment, and sport.

KEYWORDS
Sensemaking; Magic; Mental Models; Pattern-Matching; Tonic Immobility

INTRODUCTION

"*Alice laughed: "There's no use trying," she said; "one can't believe impossible things." "I daresay you haven't had much practice," said the Queen.*" (Carroll, 1866, p. 100)

"*Thus, at the peak moment, the ring rises and blows away their rapid intellectualizing and leaves them with an entirely non-cerebral event.*" (Brown, 2001, p. 36)

In his 1998 show 'David Blaine: Magic Man', Blaine approaches a boy on the street, and begins to shuffle a deck of cards. He hands the deck to the boy, and asks him to cut it anywhere, and to then remove the top five cards at the location he cut to. The remainder of the deck is taken back by Blaine. Whilst turning to face away from the boy, Blaine instructs him to spread the cards he has just removed, to look at their faces, and mentally to select and think of just one card. The boy is then instructed to place the removed cards back into the middle of the deck, which is then discarded to Blaine's side. The boy is told to place his hand on Blaine's chest, approximately over his heart, and to name aloud his thought of card. The boy states that he is thinking of the four of hearts. Blaine then lifts his t-shirt to reveal a tattoo on his chest of a playing card, the four of hearts, in the precise location where the boy has just touched. The boy stands motionless and expressionless for a few seconds, staring at the tattoo. He mutters the words "Oh, man", touches his chin, and continues to stare for several more seconds. He then turns away and stares downwards, slumping, shrugging and shaking his head in a repeated cycle that lasts about 15 seconds. The performance can be viewed at https://youtu.be/FqlfrbbFRls?t=22s (from 00:22 – 00:43). Similar reactions to Blaine's performances can be seen at https://youtu.be/N8Yhaz4xDRM?t=1s (from 00:01 – 00:43) and at https://youtu.be/masWR2VMWZE?t=10m31s (from 10:31 – 11:30).

Such reactions are typical when a spectator experiences a profound moment of strong magic. However, these reactions are perhaps the opposite of what one might expect to see when a spectator experiences something impossible (i.e. screaming, shouting, laughing, running away, or expressing disbelief or delight to nearby witnesses, etc). In his book 'Strong Magic' (Ortiz, 1994) the magician Darwin Ortiz differentiates 'strong magic' from other types of magic experience that are less impactful on the spectator. He suggests that most magic does not truly engage the spectator, leaves them bored and unentertained, makes them feel like they have been challenged or that they have a puzzle to solve, and potentially even leaves them feeling annoyed or upset (see also Fitzkee, 1943, p. 3; and Brown, 2001, p. 40). In contrast, 'strong magic' creates a profound, deep and meaningful experience for the spectator, leaving them to believe that they have genuinely taken part in something impossible for which they have no explanation. Such contrast is also reflect in the strong magic differentiation between *experiencing* a magic effect, versus merely *observing* a magic effect. Earl (2016, p. 4) suggests that in this moment "The spectator experiences a fracture in the grounded sense of consciousness or reality, and instead it is replaced by a disconnection from any sense of context or understanding. That is the magic moment: an empty space, a void of pure nothingness in which the spectator dissolves, losing any sense of who or what they are. This is the magic moment: no thoughts, feeling or action, complete nothingness." It is the sensemaking activities involved in this experience of strong magic that provide the focus for this paper.

THE SCIENTIFIC STUDY OF MAGIC

The formal scientific study of magic as a basis for generating psychological insight has a long and rich history. Early psychological investigations of magic from the 1880s-1930s focused largely on how magicians employ sleight of hand to manipulate spectator attention, the mechanisms through which belief is shaped in both individuals and groups, how magicians' patter (the words they speak to describe, explain and guide the spectator through an effect), self-presentation and conviction support and amplify spectator belief. After a relative hiatus of interest in magic, the past 10-15 years has seen an incredible resurgence of interest in the scientific study of magic (see Tompkins, 2016) to explore a diverse range of psychological phenomena, including: the physiology and psychology of misdirection, change blindness, lower- and higher-level cognitive interactions, causal reasoning, insight, mental models, expectations, priming, choice and agency, temporality, belief, memory, learning, language, culture, expertise and ideology. This surge of psychological research using magic may, in part, be due to the public popularity of recent 'neuro-magic' research led by Macknik & Martinez-Conde (e.g. Macknik, Martinez-Conde, & Blakeslee, 2011), enhanced by the researchers' association with a number of big-name magicians. For a sample set of references relating to the scientific study of magic see Table 3.

Despite this extensive and rapidly growing body of research, one area that appears to have received little or no scientific attention is the moment of magic itself - the moment when a spectator's sensemaking is disrupted by experiencing something impossible. Often this moment is described in magic literature as the spectator experiencing a moment of *enchantment* (e.g. Curry, 1999; During, 2002, pp. 43-73; Anthony, 2010; Rolfe, 2015), *astonishment* (e.g. Harris, 1992), *wonder* (e.g. Ortiz, 2006, pp. 32-33; Henning, quoted in Randall, 2009; Elimelech, 2015), *amazement* (e.g. Lavand, quoted in Ortiz, 1994, p. 17; Elimelech, 2015), or m*ystery* (e.g. Swiss, 2002). This paper, in agreement with Earl (2016, p. 4), argues that these descriptions more accurately refer to the emotions experienced in the moment *after* the moment of magic - that is, they describe the spectator's *reaction* to experiencing something impossible, and are not pertinent to the moment of magic itself. This paper presents an alternative view, that identifies the moment of magic as a profound disruption in sensemaking caused by an induced breakdown in pattern-matching, and an inhibition of the potential for recovery.

SENSEMAKING ACTIVITY WHEN EXPERIENCING A MAGIC EFFECT

When a spectator experiences a magic effect, a range of psychological processes are utilized. First, attentional and sensory systems (primarily the spectator's eyes, ears, and sense of touch - although some effects also make use of smell and taste) are used to collect and track information about the stimuli that are present in the performance environment. Microcognitive processes perceive these stimuli as cues that inform the spectator's understanding about the features and actions that are present in the environment, including the objects that the magician is using and the actions that that he or she is performing. In parallel, with this microcognitive activity, the spectator engages in a higher-level macrocognitive activity of dynamic 'Sensemaking' to generate meaning from their experience of these events (literally, to make sense of what is happening as they experience the effect). This higher-level process includes the activities of actively managing and directing attention, managing uncertainty, detecting anomalies, generating expectations, etc. (Klein, Ross, Moon, Klein, Hoffman, & Hollnagel, 2003).

Klein, Moon, and Hoffman (2006); and Klein, Philips, Rall, and Peluso (2007) suggest that people make sense of their experience through a process of pattern matching. Pattern-matching can be either a subconscious, or a conscious and deliberative activity that enables people to make sense of their world. Humans learn by mentally creating patterns (characteristic associations, or 'frames', which may comprise stories, maps, organizational diagrams, or scripts) among objects, properties, behaviors, and causes and effects (the 'data' perceived from the environment). Consequently, these patterns are stored as mental models, and are employed as templates against which to compare our perception of real world features. Our 'pattern library' is thus our experience. When we see characteristic collections of cues that together we *recognize*, this triggers activation of the relevant pattern, which creates meaning and tells us the set of expectations and actions that are appropriate for the situation at hand – a process known as 'Recognition Primed Decision Making' (Klein, Calderwood, & Clinton-Cirocco, 1986). The spectator's expectations about what will happen in the environment then in turn direct their attention and ongoing collection of additional environmental data. Frames therefore determine what counts as data (i.e. which data are noticed or searched for) whilst at the same time, the perceived data activate, shape or generate the frames themselves (we thus construct our frames based on previously experienced data relationships).

Table 3 - Psychological studies using magic as a basis

Research focus	Psychological studies using magic as a basis
Early research on human attention, perception and belief	Hodgson and Davy (1887); Jastrow (1888); Dessoir (1891, 1893); Binet (1894); Jastrow (1896); Triplett (1900); Carrington (1907); Rockwood (1919); Blodgett (1927); Hahne (1929); and Bernhard (1936)
Physiology and psychology of misdirection	Tatler and Kuhn (2007); Kuhn, Amlani, and Rensink (2008); Kuhn, Tatler, and Cole (2009); Kuhn and Findlay (2010); Hergovich, Gröbl, and Carbon (2011); Kuhn and Martinez (2011); Otero-Millan, Macknik, Robbins, and Martinez-Conde (2011); Demacheva, Ladouceur, Steinberg, Pogossova, and Raz (2012); Smith, Lamont, and Henderson (2012); Rieiro, Martinez-Conde, and Macknik (2013); Kuhn, Caffaratti, Teszka, and Rensink (2014); Rensink and Kuhn (2014); Tachibana and Kawabata (2014); Phillips, Natter, and Egan (2015); Tachibana and Gyoba (2015); Caffaratti, Navajas, Rey, and Quian Quiroga (2016); Hergovich and Oberfichtner (2016); Thomas and Didierjean (2016a, 2016b); Wiseman and Nakano (2016)
Neuro-psychology of attention and perception	Martinez-Conde and Macknik (2007); Macknik, King, Randi, Robbins, Teller, Thompson, and Martinez-Conde (2008); Martinez-Conde and Macknick (2008); Macknik and Martinez-Conde (2009); Macknik, Martinez-Conde, and Blakeslee (2011); Quiroga (2016)
Change blindness	Memmert (2010); Most (2010); Smith, Lamont, and Henderson (2012, 2013); Aardema, Johansson, Hall, Paradisis, Zidani, and Roberts (2014); Smith (2015); Lamont and Wiseman (n.d.)
Perception-cognition interactions	Thomas, Didierjean, Maquestiaux, and Gygax (2015); Ekroll and Wagemans (2016); Tompkins, Woods, and Aimola Davies (2016); Ekroll, Sayim, and Wagemans (2017)
Causal reasoning and insight	Subbotsky (1996); Subbotsky (1997, 2001); Parris, Kuhn, Mizon, Benattayallah, and Hodgson (2009); Faber (2012); Danek, Fraps, von Muller, Grothe, and Ollinger (2013, 2014); Hedne, Norman, and Metcalfe (2016); Smith, Dignum, and Sonenberg (2016); Williams and McOwan (2016); Shtulman and Morgan (2017)
Mental models	Landman (2013); Beth and Ekroll (2015); Griffiths (2015); Tompkins, Woods, and Aimola Davies (2016)
Expectations	Danek, Ollinger, Fraps, Grothe, and Flanagin (2015); Van de Cruys, Wagemans, and Ekroll (2015)
Priming	Trinkaus (1980); Mohr, Koutrakis, and Kuhn (2014)
Choice and agency	Shalom, de Sousa Serro, Giaconia, Martinez, Rieznik, and Sigman (2013); Olson, Amlani, Raz, and Rensink (2015)
Temporality	Thomas and Didierjean (2016c)
Belief	Williams and McOwan (2014); Leddington (2016); Olson, Landry, Appourchaux, and Raz (2016); Leddington (in press 2017)
Memory	Wilson and French (2014); Olson, Demacheva, and Raz (2015)
Learning	Curzon and McOwan (2008); Moss, Irons, and Boland (2016)
Language	Jones and Shweder (2003)
Culture	During (2002)
Expertise	Jones (2011); Rissanen, Palonen, Pitkänen, Kuhn, and Hakkarainen (2013); Rissanen, Pitkanen, Juvonen, Kuhn, and Hakkarainen (2014)
Ideology	Saville (2013)

The Data-Frame sensemaking model (Klein, Moon, & Hoffman, 2006) provides a description of how people generate an initial account to explain (i.e. make sense of) events they are experiencing; how they elaborate that explanation to account for new information they have acquired; how they question that explanation when they discover data that does not fit; how they may potentially defend their explanation in light of contradictory data; how they discover limitations or flaws in their understanding; and how they reframe (switch or adjust) their explanation to an account that better explains the circumstances. The core activity within the Data-Frame model is the parallel process of applying a frame to the data, and the data to the frame. The frame explains how the data relate to each other (in terms of how they form a meaningful pattern) and the frame guides the search for new data whilst also defining what actually counts as data. Therefore, sensemaking can be viewed as "not merely joining the dots or generating inferences, but also identifying what counts as a dot, and how to go about seeking new dots." (Hutton, Klein, & Wiggins, 2008). Sensemaking activities within the Data-Frame model are depicted in Figure 2, and Figure 3 presents the same model with a magician's goals for shaping a spectator's sensemaking overlaid. The model identifies seven different forms of sensemaking, which are summarized and considered in terms of their relevance to magic in Table 4.

Note that the spectator's sensemaking will often begin to be influenced by a magician before the spectator is aware that the effect has begun. For example, the way in which the magician approaches the spectator, introduces themselves, frames their capabilities, introduces and sets-up the effect, removes items from their person, and invites the spectator to begin to participate, all serve to suggest a frame of meaning, and influence the spectator's sensemaking before the effect *appears* to have begun; yet all such activities may in fact be intrinsic to both the method and to the effect itself. Note also that during this phase the magician may be performing a range of 'secret moves' or *sleights*, which are hidden from the spectator through both covert manipulation, and as a result of the spectator's attention not yet having been engaged and directed towards the effect. Later, when the spectator attempts to recall the sequence of events leading up to the moment of magic, their recall will begin at the point in time when they activated their attention - i.e. at a point *after* the secret moves had been performed. From that point on, the spectator's Data-Frame matching process will therefore be operating on the basis of incomplete data.

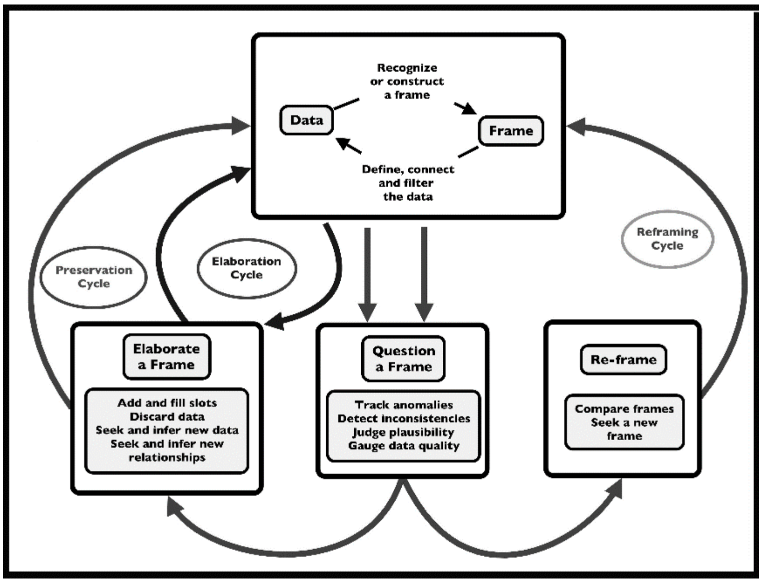


Figure 2 - The Data-Frame model of sensemaking (Klein, Moon, & Hoffman, 2006)

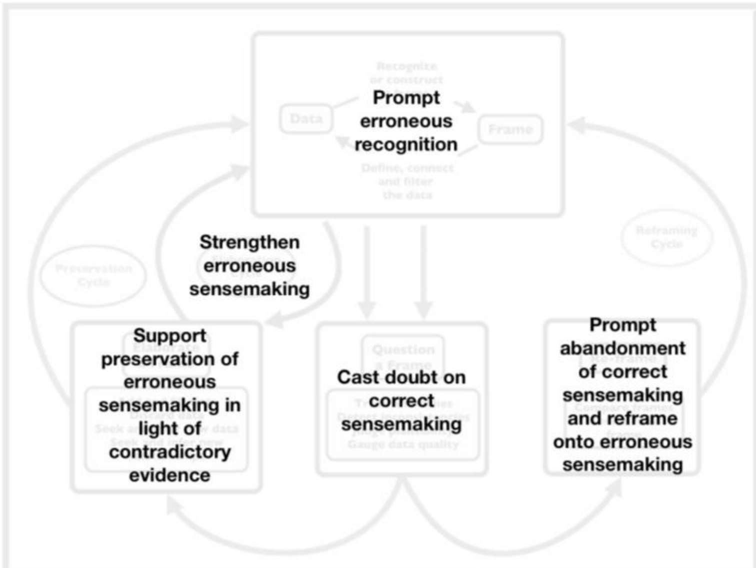


Figure 3 - The Data-Frame model of sensemaking overlaid with magician's deception goals (adapted from Klein, Moon, & Hoffman, 2006)

Within the primary matching loop of the Data-Frame model, the spectator applies frames about the physical world, objects, and actions, to make sense of what they are observing, directed and informed by the magician's patter, movements, and own attention. Such patter may draw the spectator's attention towards features that are salient to the sensemaking that the magician wishes to provoke, and draw attention away from (i.e. misdirect from) noticing features associated with the methods underpinning the effect. Actions are interpreted in terms of the spectator's general experience about how the world works, and how the objects employed by the magician work. However, this recognitional process (including associations, attributions and assumptions) is open to exploitation. For example, when a deck of cards is shuffled we know that the order is changed into an unknown sequence. However the magician may be performing a highly convincing false shuffle that retains the order of the deck, that visually is virtually indistinguishable (particularly by non-magicians) from the real thing. The magician may also be presenting false cues to the spectator to shape directly their erroneous pattern recognition and sensemaking process. For example, if magician and spectator are sat at a table, the magician might at one point appear to transfer a coin from the right hand to the left. Subsequently, after a little time has passed, he will slowly open his left hand to show that the coin has disappeared. The spectator now suspects that the coin must really still be retained in the magician's right hand, which he notices now appears to be a little stiff and unnatural (these are false cues presented by the magician). In fact, the coin was ditched from the magician's right hand into his lap whilst the spectator was focusing on the magician's left hand revealing itself to be empty. The fact that the spectator's attention is now directed intently onto the magician's empty (but apparently, coin-retaining) right hand creates the perfect opportunity for the magician to use his left hand to secretly steal the ditched coin back from his lap. Note that throughout this effect, the magician will be directing *his own* attention to the hand where he wants the spectator to look, as this form of social cueing is powerfully seductive.

Table 4 – Sensemaking activities when observing a magic effect

Sensemaking activity	**Explanation from Klein et al. (2007)**	**Sensemaking activity when experiencing a magic effect**
Sensemaking attempts to connect data and a frame	The specific frame a person uses depends on the data or information that are available; and on the person's goals, the repertoire of the person's frames, and the person's stance (e.g., current workload, fatigue level, and commitment to an activity).	In a magic performance, the spectator's data and frames will relate to: the performance *context* (e.g. a performance on stage; a performance in the street, etc); the *magician* (how they are dressed, their demeanor; how they speak, etc); the *effect introduction* and set-up (e.g. a demonstration of mind reading; a card trick, etc); the *objects* used (e.g. coins, cards, clipboards, etc); the *magician's actions* (e.g. shuffling a deck, writing a prediction, etc); and the *spectator's actions* (e.g. thinking of a friend's name; selecting a card, etc). Frames will also be informed by the spectator's pre-existing familiarity with magic.
Elaborating the frame	As more is learned about the environment, people will extend and elaborate the frame they are using, but will not seek to replace it as long as no surprises or anomalies emerge. They add more details, fill in slots, and so forth.	The spectator's frame will be elaborated as they observe the effect progressing. For example, the magician may shuffle a deck of cards, have the spectator select a card, loose the card back in the deck, and then produce an envelope from their pocket. Each action and manipulation of the objects serves to elaborate the spectator's sensemaking frame, advancing their understanding, and giving rise to new expectations.
Questioning the frame	Questioning begins when we are surprised - when we have to consider data that are inconsistent with the frame we are using. This is a different activity than elaborating the frame. (Lanir, 1991) has used the term "fundamental surprise" to describe situations in which we realize we may have to replace a frame on which we had been depending. He contrasted this with situational surprise, in which we merely need to adjust details in the frame we are using.	Throughout the magic performance, the spectator may question and adjust their frame. For example, when the magician produces an envelope from their pocket during a card trick, the spectator may expect their previously vanished card to be revealed within the envelope. However, when they rip open the envelope to find a small padlock key, their (situational) surprise and violation of expectations will result in a revision of their frame for making sense of what is happening. Note that whilst the moment of magic itself encompasses profound surprise, many smaller surprises within the effect may lead up to this moment.
Preserving the frame	We typically preserve a frame by explaining away the data that do not match the frame. Sometimes, we are well-advised to discard unreliable or transient data. But when the inconsistent data are indicators that the explanation may be faulty, it is a mistake to ignore and discard these data. Related to fixation.	When observing a magic performance, the spectator may have been led by the magician to suspect certain methods to be used to achieve the effect. This expectation can then be exploited by the magician to divert the spectator's attention away from the real method, and, (should the spectator catch any of the real method) encourage them to dismiss or explain away their observation in favor of the stronger frame.
Comparing multiple frames	We sometimes need to deliberately compare different frames to judge what is going on. This may entail holding onto one frame, whilst concurrently elaborating a second, opposing frame.	Throughout the observation of a magic effect, the spectator may hypothesize as to what they believe is 'really' happening (as opposed to what the magician suggests is happening). This will involve holding onto the magician's framing of events, whilst concurrently generating and evaluating alternative frames that might account better for the data in respect of suspected trickery (which may itself, in fact, be prompted by the magician!).
Reframing	In reframing, we are not simply accumulating inconsistencies and contrary evidence. We need the replacement frame to guide the way we search for and define cues, and we need these cues to suggest the replacement frame. Both processes happen simultaneously.	Should the spectator decide that they are not 'buying' the magician's suggested frame for making sense of their actions, the spectator may decide to reframe events based on their reasoning as to what is 'really' happening. The new frame will better account for their suspicions, postulated methods being used by the magician, and expectations regards the outcome of the effect. A new frame may, for example, be based on an apparently analogous magic effect the spectator has experienced before; or could potentially be anchored around a known or suspected property of the objects being used, etc.
Seeking a frame	We may deliberately try to find a frame when confronted with data that just do not make sense, or when a frame is questioned and is obviously inadequate. Sometimes, we can replace one frame with another, but at other times we have to try to find or construct a frame. We may look for analogies and also search for more data in order to find anchors that can be used to construct a new frame. People can assemble data elements as anchors in the process of building an explanation.	At various points whilst experiencing a magic effect, the spectator may seek to construct a new frame that better explains the collected data up to that point, that reflects hypotheses about what is 'really' happening, that incorporates suspicions about any secret methods employed, and that includes the spectator's expectations about how the effect may end. For example, having experienced (a) their selected card being mixed back into the deck; and (b) the envelop produced by the magician turning-out not to hold their card, but a key; the spectator may now construct a new frame based on an expectation that their card will be found in a locked container of some kind. When a safe-box is produced later in the effect, this supports and further strengthens the new frame.

Note that at no point does the spectator make sense of the secret methods by which the effect is achieved, as these will be invisible to the spectator, even though they may be occurring in front of their eyes. Such secret methods can be made invisible through a range of strategies, for example: the magician may block such actions from being viewed by turning their body so that their hand momentarily cannot be seen; the actions may be embedded within natural action that provides cover – for example, ditching a coin in a jacket pocket when reaching for a pen; the spectator may not attend to the secret action as their attention has been directed elsewhere (i.e. they have been misdirected), for example, the magician looks at their right hand whilst their left ditches the coin in their pocket; or the spectator sees but does not register the action, as the actions are not task-relevant (i.e. the spectator has been rendered inattentionally blind to noticing those actions), for example, the magician may ask the spectator to count the number of red-faced (hearts and diamonds) cards as a pile is dealt onto the table, leading them to not notice that several duplicates of a particular card are present in the pile being dealt.

Throughout the performance of the effect, the spectator may also seek actively to avoid being fooled. They might become hypervigilant, seeking to notice everything that is happening before them, to catch every detail and miss nothing. The may seek deliberately to not be misdirected, trying instead to fixate on observing the magician's hands at all times. And they may try to anticipate what will happen next, to inform their attentional and sensemaking strategy to keep one step ahead of the magician. At the same time, the magician will seek actively to control the spectator's observation and collection process in line with the frame is he seeking to promote, for example, using conspicuity to attract the spectator's attention; creating expectations through repetition and portrayal of causality, so that the spectator directs their own attention to where the magician wants; and he may deliberately portray both real and false patterns that shape the spectator's sensemaking, including the spectator's attempts to second-guess what is really happening. Patter will also play a key role in influencing the spectator's attention, pattern activation, and expectations – for example, asking the spectator a question will make them momentarily make eye contact with the magician, taking attention away from his hands (and enabling a secret move to go undetected). False cues may be generated or pre-planted to suggest incorrect methods by which the effect is achieved, or to set-up false assumptions or expectations. Secret moves may also be employed, with any detectable signatures being attenuated, hidden or obfuscated from the spectator. And in a performance of strong magic, the spectator's pattern recognition and well-founded expectations will, at some point, be profoundly violated to create an intense and impactful moment of magic.

THE MOMENT OF MAGIC

> "*Regarding each moment in a magic effect, the audience says, 'Yes, that was fair.' 'Yes, that was fair.' 'Yes, that was fair.' And then the moment of magic occurs and the audience says, 'Wow! What happened here?'*" (Burger, 2003, p. 196)
>
> "*Attention, if sudden and close, graduates into surprise; and this into astonishment; and this into stupefied amazement.*" (Darwin, 1873, p. 278)

At some point during the performance of a magic effect, usually at the end, a 'moment of magic' occurs. If the magician's performance has been weak and unengaging, this moment may elicit a neutral, barely perceptible response from the spectator, and potentially even a negative response. However, if the magic has been 'strong' and the moment curated effectively, the spectator's response may be profound.

The moment of magic is created through a combination of surprise, sudden violation of strongly founded expectations, disruption of pattern matching, and a lack of available sensemaking 'recovery cues' which serves to prolong the moment. So whilst the moment of magic relies upon surprise (defined by Willis (2002, p. 1660) as "a highly transient reaction to a sudden and unexpected event.") within the context of magic many other processes occur that intensify the experience beyond mere surprise. The phases a spectator passes through when experiencing a moment of magic correspond with the four stages of surprise described by Luna and Renninger (2015, pp. 5-6). This framework suggests that people who are surprised transition through the phases of Freeze, Find, Shift and Share, key features of which are identified in Figure 4. Note that whilst the moment of magic relates to Luna and Renninger's Freeze phase, subsequent phases serve to amplify impact of this moment on the spectator's overall experience of the effect. Each stage will now be considered in relation to the moment of magic and associated sensemaking activities.

A sensemaking account of the Freeze phase in magic

Tomkins (1963) describes surprise as a "resetting" state because, for a fraction of a second, the mind is cleared of thought. The momentary interruption of thought and action is related to the basic adaptive function of facilitating rapid evaluation and response to sudden changes. Luna and Renninger (2015) suggest that "Surprise is the neuropsychological equivalent of a pause button. It makes us stop what we're doing, hijacks our attention, and forces us to pay attention." Such physiological 'freezing', which is evident in both performance clips of Blaine, has been studied in the laboratory by Luna (2013) who asked study participants to observe a short video whilst concurrently making a circulating motion in the air with their finger. Unknown to the participants, the video contained a sudden and unexpected event. On witnessing this event, participants could be seen momentarily stopping circulating with their finger, starting again when they emotionally expressed surprise (a clip from this study can be viewed at https://youtu.be/6JrbL4gBg-c?t=45s, from 0:45 – 1:12). This phenomenon has also been labelled as "Attentional Blink" (e.g. Shapiro, Raymond, & Arnell, 1997; Biggs, Cain, Clark, Darling, & Mitroff, 2013; Cialdini, 2016, p. 29); as a "Pattern Interrupt" (e.g. Luna & Renninger, 2015, pp. 123-126; Cialdini, 2016, pp. 76-79) and as a type of "Instantaneous induction" (e.g. McGill, 1996, pp. 224-227)

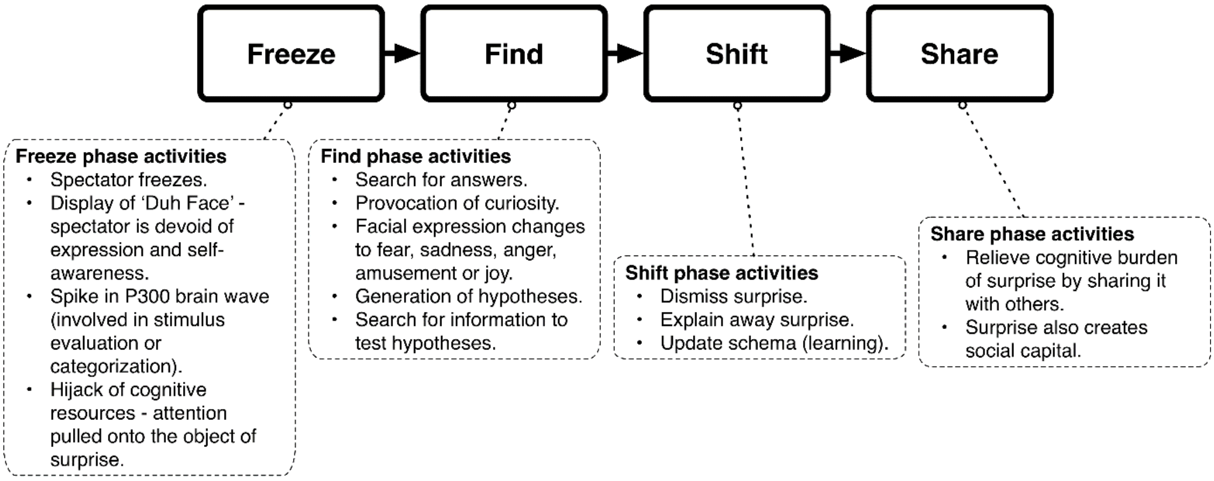


Figure 4 – The four phases of surprise (adapted from Luna & Renninger, 2015)

This physiological freezing phenomenon is consistent with the phenomenon of 'Tonic Immobility', which has been studied in animals since the 1600s (see Völgyesi & Klumbies, 1966) - later emerging as a field colloquially, (and perhaps unhelpfully) referred to as 'animal hypnosis' (Crozier, 1923; Gallup & Maser, 1977). Tonic immobility is a temporary behavioral state of motor inhibition thought to occur in response to situations involving intense shock or fear (literally, being 'scared stiff'). Nash, Gallup, and Czech (1976) identified that Tonic Immobility presents as a stuporous catatonic-like immobility with muscular hyper- or hypo-tonicity, odd postures, suppression of vocalization, intermittent eye closure, Parkinsonian-like tremors, and reduced responsiveness to external stimulation. Physiological correlates have been found to include decreases in heart rate and body temperature, increased respiration rate, and altered electro-encephalograph patterns. Because of the potential for tremor, eye closure, occasional head turns and vocalizations, the behavior cannot be described as absolute immobility, however, there is clearly a profound state of response inhibition. Studies have indicated that the state may last from fractions of a second to over several hours, and the phenomenon has been identified in scores of different species including insects, crustaceans, fish, amphibians, reptiles, birds, mammals and primates (Gallup & Maser, 1977). However, Abrams, Nicholas Carleton, Taylor, and Asmundson (2009) suggest that most research effort to date has been directed towards studying the phenomenon in animals, with relatively little attention having been paid to the phenomenon in humans.

Ratner's Defensive Distance Hypothesis (Ratner, 1967) suggests that tonic immobility has evolved to provide a last line of survival under conditions of predation, such immobility reducing the likelihood of further attacks, minimizing visual cues to the predator, and increasing the probability of predator boredom or distraction (in animals, the process of playing dead is known as 'thanatosis'). As the distance between predator and prey decreases, Ratner suggests that the prey moves through the stages of freezing, attempted flight, struggle with the predator, and lastly tonic immobility. Gallup and Maser (1977) have tested this hypothesis experimentally, showing that the duration of freezing is indeed affected by predation distance; specifically, the proximity the predator's eyes alone (when the eyes are hidden or removed from a stuffed predator, its proximity no longer moderates freeze duration). It is interesting to note that during the moment of magic, various magicians have developed an approach of staring intently at the spectator whilst themselves remaining totally motionless and silent (for example, Blaine can be seen doing this in all the examples referenced earlier, and the effect is especially pronounced in his performance at: https://youtu.be/rCskKEhaKhY?t=1h4m21s, from 1:04:21 – 1:07:58). It is suggested that this behavior on the part of the magician also serves to minimize the cues available to the spectator that might help them recover their sensemaking – thereby serving to further prolong the moment of magic.

Luna and Renninger (2015, pp. 5-6) state that during the Freeze phase, the spectator will usually for a few moments exhibit (what the authors refer to as) the 'Duh Face', wherein the face itself becomes frozen, blank, and devoid of expression. This account is at odds with traditional views on the facial expressions associated with surprise (e.g. Darwin, 1873, pp. 278-309) which describe features such as widened eyes, raised eyebrows, and open mouth. However, such expressions only arise later whilst searching for meaning, which Luna & Renninger define as the

'Find' phase. As a corollary, work by Schützwohl and Reisenzein (2012) found that common stereotypes about facial reactions to surprise were evident in just 5% of the participants that experienced intense surprise and novelty in their experiments.

The types of sensemaking described in the Data-Frame model help link together Luna & Renninger's Freeze and Find phases, in terms of an ongoing data-frame matching activity that becomes radically disrupted by the sudden presence of data that fundamentally does not fit the current operating frame and associated expectations – a form of 'fundamental surprise' (Lanir, 1991) that requires the existing frame to be replaced. The intensity of the Freeze phase depends on the degree of surprise, and the degree of surprise itself depends on the degree of frame discrepancy. Another factor that may effect the duration of this phase is culture. Choi and Nisbett (2000) tested the hypothesis that East Asians, because of their purported 'holistic reasoning', take contradiction and inconsistency for granted and consequently are less likely than Americans to experience surprise. Their studies suggest that Koreans display less surprise than Americans when a target's behavior contradicts their expectations, even when contradiction is created in highly explicit ways.

Following a momentary period of paralysis, the spectator's sensemaking then moves into a process of seeking a new frame, and a new explanation for their experience, corresponding with the 'Find' phase in Luna & Renninger's model. The Freeze phase may be transitioned into the Find phase in response to a reactivation of sensemaking (potentially based on the identification of a 'recovery cue' obtained from recall or from the external environment), dynamic internal physiological events, or from environmental events and associated new data.

A sensemaking account of the Find phase in magic

Following the initial phase of freezing, the spectator will likely attempt to recall and reconstruct the sequence of activities leading to the moment of magic, seeking an alternative frame to make sense of what they have experienced. This activity may involve formulating questions or hypotheses, necessitating the comparison of multiple frames. Early theories of surprise, including Darwin's (Darwin, 1873), argued that surprise is predominantly a basic emotion, a viewpoint supported by others such as Ekman and Friesen (1971). More recent theories have used the potential for surprise to be positive or negative to develop a more cognitive view of surprise that casts it as a process of ''making sense of surprising events". Foster and Keane (2015) advance the view that the essence of sensemaking following surprise is explanation; specifically, that people's perception of surprise is a metacognitive estimate of the cognitive work involved in explaining an abnormal event; and therefore, the degree of surprise is determined by how difficult the event is to explain. Experimentally, Foster and Keane have shown that the difficulty of explaining a surprising outcome is the best predictor for people's perceptions of the surprisingness of an event.

To the spectator, all the data necessary for understanding what is happening appears to exist in the present moment, in front of them within the context of the effect performance (note that they are unaware of the other data that has been deliberately hidden from them, relating to the secret moves or gimmicked objects employed). Thus, when the moment of magic occurs and their understanding of the current data fails, the *only* place to begin looking for an explanation is in their recall of the past sequence of events (i.e. memory). However, memory involves 'attention in the past', and the magician has already manipulated the attention of the spectator - therefore memory is a compromised resource for supporting the spectator's recovery of sensemaking. The spectator has been carefully guided through the process of the effect, therefore their data and frames have been subtly created and managed by the magician. Memory is the only major resource available to potentially unravel the effect, and humans generally have misplaced faith in the validity of their own memory. Memory is by nature a distortion of reality, an approximation of real events. Magicians know this and intentionally manipulate this resource, counting on spectators to over-rely on their ability to correctly recall events.

The Data-Frame model is also predicated upon the utilization of expert knowledge, familiarity (based on our stored experiential pattern library), and feature recognition to enable sense to be made of a given situation. In the context of a magic effect, most spectators are novices in that they do not have experience of, familiarity with, or expertise in, the performance of magic effects. They do not know about the secret methods by which magicians achieve their effects (which are deliberately kept outside of public awareness); and they have little or no experience even of having previously observed the impossible feats they have just witnessed. To have the expertise necessary for recognizing the patterns underlying the achievement of a magic effect, the spectator needs to have experience (that goes beyond knowledge alone) of effect design, sleights, apparatus, performance, and spectator impact. The interaction between a magician and spectator is thus founded upon a fundamental mismatch in levels of domain experience. Thus, any expertise in magic will reduce significantly the intensity of any surprise, and thus the intensity of the moment of magic (Alberdi, Sleeman, & Korpi, 2000; Foster & Keane, 2015). Even a partial explanation that a spectator can generate will serve to reduce the intensity of the moment (Foster & Keane, 2015).

Foster and Keane (2015) propose that any factor that acts to increase the cognitive work in explaining a surprising event results in higher levels of surprise. Additional factors that increase a spectator's cognitive load when experiencing a magic effect include: the spectator's lack of scientific and psychological knowledge, their lack of conjuring knowledge, their inability to assess the plausibility of potential methods employed by the magician (e.g. the actual method may be dismissed from consideration as a result of it being perceived as too simple or too complex), the spectator misremembering, and finally, the magician using deliberate strategies to misdirect the spectator's recall and reconstruction, such as having laid-out false cues, and recapping events in a deliberately misleading or suggestive way (see Lamont & Wiseman, 2005, pp. 82-101). This means that not only does the spectator not know *where* to look, or *what* to look for to construct meaning regards how a magic effect has been achieved, but also that if they do detect something meaningful, they are unlikely to recognize *why* it is important.

In addition, when the magician stares at the spectator following the moment of magic, this may serve not only to reduce any potential 'recovery cues' transmitted by the magician, but also to seduce and fixate the spectator's attention away from the environment and any associated objects in-play, thereby reducing and constraining their search for other meaningful cues. The magician may also take deliberate action to help frame the spectator's search and sensemaking process following the reveal - for example, Fitzkee suggests that it is the magician's job to use "words, actions and implications" to 'interpret' and add meaning to the spectator's experience of the magic effect (Fitzkee, 1945, pp. 35-37). Similarly, Brown suggests that the purest experience of magic is simply 'confusion', and that the job of the magician is not just to invoke confusion, but subsequently to ascribe meaning to it (Brown, 2001, pp. 39-44).

It is also worth noting that some research suggests that individuals are likely to experience greater surprise when participating in a magic effect as an individual, than when experiencing magic in a group. In a group setting, the collective will have wider experience from which to draw, have more cognitive resources available, can gather and analyze in parallel more data, and can reach a collaborative consensus of post-effect explanation. In addition, if a spectator is selected by a magician from a group to be the focus of the effect, the other group members may serve to provide distractions, and act as points of reference for the spectator to 'check-in' with as the effect progresses. These are all factors that are likely reduce the spectator's surprise (see Tachibana & Kawabata, 2014; Foster & Keane, 2015).

Finally, the Find phase involves the spectator experiencing an emotional reaction to their experience of surprise, which may be fear, anger, amusement or joy. Such emotions are usually telegraphed externally via facial expressions and verbal utterances, and are usually (and as suggested in this paper, erroneously) associated with occurring *at* the moment of magic, as opposed to being a later reaction *to* the moment of magic.

A sensemaking account of the Shift phase in magic

Luna and Renninger (2015) suggest that, having acquired new data via a search process, the spectator may dismiss or ignore the cause of the surprise as irrelevant ("I know it's impossible to bite coins in half and then restore them by blowing, so I figure that was just silly and not worth any further consideration."). They may develop an explanation for the effect, which may be correct or incorrect ("I know that teeth aren't strong enough to bite coins, so that guy must be wearing a special gum-shield that he removed when I wasn't looking"). Or they may update their schema to account for their experience ("I know most people can't bite and restore coins, but based on this experience I figure people with strange powers maybe can.") In the Data-Frame model, these activities follow-on from questioning the frame and seeking a frame, and correspond broadly with the sensemaking activities of preserving the frame, reframing, and elaborating the frame. And as cited earlier, any factor that serves to increase the cognitive work involved in these reasoning processes will likely increase the intensity of the moment of magic (Foster & Keane, 2015).

A sensemaking account of the Share phase in magic

Surprise creates a cognitive burden (Söderlund, 1998) as it is difficult to keep an emotionally and cognitively intense experience to ourselves. Surprise also arms us with social capital, enabling us to share an interesting story about our experience with others. Research demonstrates that the more significant the surprise, the sooner and more frequently we share it (Rimé, Philippot, Boca, & Mesquita, 1992). Repetition of a story also increases retention, and in the case of a magical experience that story is already atypical (Munnich, Ranney, & Song, 2007; Adler, 2008; Foster & Keane, 2015; Schomaker & Meeter, 2015). Repeated telling also increases the likelihood of distortion and confabulation, exaggerating and amplifying the power of a recalled effect. After performing a strong magic effect to an individual selected from a group (for example, in a street magic setting) it is common to see the selected spectator socially unloading and sharing their experience, seeking validation and experiential normalization, checking-in with other group members to compare their personal experiences of the effect.

A magician can readily capitalize on these short-term and longer-term effects, using spectators' exaggerated personal testimony to increase public perceptions about their powers, and to feed the rumor and publicity mill to seed expectations about the power of their act relative to other magicians, and thereby gain greater bookings.

IMPLICATIONS OF THE DATA-FRAME MODEL FOR AMPLIFYING THE MOMENT OF MAGIC

A sensemaking data-frame analysis of the moment of magic helps identify a range of strategies that serve to moderate the intensity of the spectator's experience of a magic effect. Table 5 summarizes these, classified into strategies employed ahead of effect execution, those employed during effect execution, and those employed after effect execution. To maximize the intensity and duration of the moment of magic, it is suggested that multiple strategies should be exploited across all phases.

These strategies are potentially exploitable systematically by magicians to help shape the design of both their effects and their performance to maximize spectator impact – in other words, to make their magic stronger. It is suggested that such strategies are also potentially available to planners, designers and practitioners in other domains that wish to create similar experiences for their audience. For example, fiction writers could supplement narrative development using these strategies to amplify reader immersion, impact and entertainment (e.g. Bae & Young, 2008). Military deception planners could utilize the strategies to increase the degree of surprise that their operations achieve (e.g. Whaley, 2007). Sports team could exploit such strategies to support the design of deceptive plays that create surprise against their competitors on the sports-field (e.g. Pfleegor & Roesenberg, 2013). Advertisers could develop adverts based on these strategies to support enhanced brand recognition and retention within their target audience (e.g. Alden, Mukherjee, & Hoyer, 2000). And industrial designers might employ these strategies to design objects that surprise, confound and delight users in relation to their outward appearance; that operate in unexpected or magical ways; or that support the user in unexpectedly helpful ways they never envisaged (e.g. Ramírez, 2014).

CONCLUSIONS

'Strong magic' provokes a profound, deep and meaningful reaction in a spectator, leaving them to believe that they have genuinely experienced something impossible for which they have no explanation. Whilst the moment of magic is often described as a moment of enchantment, astonishment, wonder or amazement, in agreement with Earl (2016, p. 4) it is suggested that these are in fact emotional reactions that occur *after* the moment of magic itself. The moment of magic is characterized by a profound disruption in the spectator's sensemaking resulting from a breakdown in pattern matching, coupled with an induced inability to recover. This results in the spectator experiencing a moment of temporary physiological and psychological paralysis, before they then try to recover and generate meaning from mental replays of their apparently impossible experience. Fundamental to enabling the moment of magic is a mismatch between the magician's and the spectator's levels of expertise. A range of strategies for amplifying the moment of magic have been identified, that could have utility for creating and amplifying this moment within both a magic performance, and within a range of other domains and applications.

ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Benjamin Earl: (a) for highlighting that a spectator's primary data source for the 'Find' phase within a magic effect is their memory; and (b) for his essay on the moment of magic ('The empty space', Earl, 2016, pp. 4-5) that precipitated the thinking underlying this paper.

Table 5 – Key strategies for amplifying the moment of magic

Factors before the effect	**Factors during the effect**	**Factors after the effect**
• Perform for individuals in preference to groups. • Select individuals that have no knowledge of magic. • Perform in an environment that does not create expectations of magic - for example, street magic contexts may serve to amplify surprise, as there is no pre-existing performance context that serves to inform expectations of encountering the impossible. • Use objects and actions that are familiar to the spectator, to reduce within and post-effect attention, examination, searching and critical analysis. • Consider carefully how you frame and signal your own abilities through demeanor, dress, voice, introduction, patter, actions, and objects used. Bear in mind that these will also inform the spectator's post-effect sensemaking and attempt to explain their experience. • Convey confidence, self-belief, and absolute conviction in the effect, so that contextual credibility and sincerity are telegraphed to the spectator. • Employ secret moves ahead of the spectator being aware that the effect has started, before their attention towards your actions has been activated. • Note that cultural studies of surprise suggest that more intense surprise may be experienced by Western spectators than by East Asian spectators.	• Clearly frame the (apparent) start of the effect to the spectator, so that their subsequent recall of events commences at the wrong point. Wherever possible, secret moves should occur prior to this moment, to minimize within-effect anomalies. • Make the flow of the effect easy to follow. Use plain language, and clear, easily discernable, and natural movements (which may differ from popular beliefs about what is 'natural' - see Whaley (2013)). • Use the simplest means possible to achieve the effect. Minimise the use of props, phases, gimmicks, and sleights, and any anomalies they create. • Create a strong, grounded and reinforced base frame by portraying easy to recognize cues that support clear, unambiguous, and easily discernable pattern recognition. • Strengthen the base frame through apparently independent or coincidental confirmation and reinforcement of the desired pattern. • Minimise distractions and other irrelevent sources of conspicuity or interest. • Perform only natural manipulations of objects. If you must perform an unnatural manipulation, it should be physically, and preferably psychologically invisible. • Do not use objects, object properties or object functions in unusual ways, unless this is intended solely to be the moment of magic. • Telegraph purpose, motivation and justification behind your actions. • Reduce discrepancies in the patterns you are portraying to enable easier recognition. Make cue sequences and cue cluster easily discernable (although see also below). • Do not give the spectator the entire pattern 'on a plate'. Present pattern fragments, and make them invest effort to build a complete, absolutely certain, and yet absolutely wrong understanding about what is happening. • Present convincers in a natural and justified manner, with appropriate time delay. • Divorce method from effect. Eliminate potential cues that may anchor post-effect explanatory sensemaking. The fewer cues that are available to the spectator for subsequent analysis, the better. • Attenuate, mask or obfuscate anomalies created through the use secret methods or gimmicks employed. • Where necessary, present false cues about potential methods to misdirect the spectator away from real methods, but ensure that these are subsequently cancelled-out to eliminate potential sources of explanation later. • Create, reinforce and amplify expectations through repetition, conditioning, and the portrayal of obvious causality. • Invest effort and genuine belief in your actions and their effects (i.e. Method Act). • Seek to maximize and amplify schema discrepancy by violating strong expectations, breaking or defying causality, defying physics, jarring with well-established pattern frames, and portraying actions and states that have no natural explanation. • Make the actions and states associated with the moment of magic as far removed from proceeding actions and states as possible.	• Stand still and stare intently and silently into the spectator's eyes, to prolong the duration of tonic immobility, and to minimize any unintentional recovery cues that may trigger a search for meaning. • When the spectator's tonic immobility is broken, use language and movement that is congruent with the effect they have experienced. • Do not break character, change demeanor, or make references to anything outside the frame of the effect, as this may serve to accelerate the spectator's initiation of explanatory sensemaking. • Do not correct the spectator misremembering, exaggerating or confabulating details about the effect. Your specificity about any details will reduce the opportunity for future exaggeration by the spectator. • Do not correct the spectator's attribution of any impossible powers to you, or seek to amplify this yourself. Leave them to do the work for you. • Encourage the spectator to share their experience with others, to (post-hoc) amplify the intensity of the moment of magic, to increase possible confabulation and exaggeration, and to act as publicity.

REFERENCES

Aardema, F., Johansson, P., Hall, L., Paradisis, S.-M., Zidani, M., & Roberts, S. (2014). Choice Blindness, Confabulatory Introspection, and Obsessive-Compulsive Symptoms:
A New Area of Investigation. *International Journal of Cognitive Therapy, 7*(2), 83-102.
Abrams, M. P., Nicholas Carleton, R., Taylor, S., & Asmundson, G. J. G. (2009). Human tonic immobility: measurement and correlates. *Depress Anxiety, 26*(6), 550-556. doi: 10.1002/da.20462
Adler, J. E. (2008). Surprise. *Educational Theory, 58*(2), 149-173.
Alberdi, E., Sleeman, D. H., & Korpi, M. (2000). Accommodating Surprise in Taxonomic Tasks: The Role of Expertise. *Cognitive Science, 24*, 53-91.
Alden, D. L., Mukherjee, A., & Hoyer, W. D. (2000). The Effects of Incongruity, Surprise and Positive Moderators on Perceived Humor in Television Advertising. *Journal of Advertising, 29*(2), 1-15. doi:10.1080/00913367.2000.10673605
Anthony, W. (2010). *Tales of Enchantment: The Art of Magic*. Orlado, Florida: Leaping Lizards Publishing.
Bae, B.-C., & Young, R. M. (2008). A Use of Flashback and Foreshadowing for Surprise Arousal in Narrative Using a Plan-Based Approach. In U. Spierling & N. Szilas (Eds.), *Interactive Storytelling: First JointInternational Conference on Interactive Digital Storytelling, ICIDS 2008 Erfurt, Germany, November 26 29, 2008 Proceedings* (pp. 156-167). Berlin, Heidelberg: Springer Berlin Heidelberg.
Bernhard, R. E. (1936). *The Psychology of Conjuring*. Stanford University: Robert Bernhard.
Beth, T., & Ekroll, V. (2015). The curious influence of timing on the magical experience evoked by conjuring tricks involving false transfer: decay of amodal object permanence? *Psychological Research*(79), 513–522. doi: 0.1007/s00426-014-0584-2
Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F., & Mitroff, S. R. (2013). Assessing visual search performan differences between Transportation Security Administration Officers and nonprofessional visual searchers *Visual Cognition, 21*(3), 330-352.
Binet, A. (1894). Psychology of prestidigitation. *Annual report of the board of regents of the Smithsonian Institution*, 555-571.
Blodgett, S. D. (1927). The Psychology of Legerdemain. *The Sphinx, 26*(4), 148-149.
Brown, D. (2001). *Absolute Magic*. Bristol: Derren Brown.
Burger, E. (2003). On the Structure of Magic Effects. In J. Jay (Ed.), Magic in Mind: Essential Essays for Magicians (pp. 193-198). Retrieved 17/01/2017, from: https://www.vanishingincmagic.com/magic downloads/ebooks/magic-in-mind/.
Caffaratti, H., Navajas, J., Rey, H. G., & Quian Quiroga, R. (2016). Where is the ball? Behavioral and neural responses elicited by a magic trick. *Psychophysiology, 53*(9), 1441-1448. doi: 10.1111/psyp.12691
Carrington, H. (1907). The Psychology of Deception. In H. Carrington (Ed.), *The Physical Phenomena of Spiritualism, Fraudulent and Genuine: Being a Brief Account of the Most Important Historical Phenomena, a Criticism of Their Evidential Value, and a Complete Exposition of the Methods Employed in Fraudulently Reproducing the Same* (pp. 48-63). Michigan: Turner.
Carroll, L. (1866). *Alice's Adventures in Wonderland*. New York: D. Appleton & Co.
Choi, I., & Nisbett, R. E. (2000). Cultural psychology of surprise: Holistic theories and recognition of contradiction. *J Pers Soc Psychol, 79*(6), 890-905.
Cialdini, R. B. (2016). *Pre-suasion: a revolutionary way to influence and persuade*. New York: Simon & Schuster.
Crozier, W. (1923). Reflex immobility and the central nervous system. *Experimental Biology and Medicine, 21*(2), 55-56.
Curry, P. (1999). Magic Vs. Enchantment. *Journal of Contemporary Religion, 14*(3), 401-412.
Curzon, P., & McOwan, P. W. (2008). *Engaging with computer science through magic shows.* Paper presented at the 13th annual conference on Innovation and technology in computer science education, June 30 - July 02, 2008, Madrid, Spain
Danek, A. H., Fraps, T., von Muller, A., Grothe, B., & Ollinger, M. (2013). Aha! experiences leave a mark: facilitated recall of insight solutions. *Psychol Res, 77*(5), 659-669. doi: 10.1007/s00426-012-0454-8
Danek, A. H., Fraps, T., von Muller, A., Grothe, B., & Ollinger, M. (2014). It's a kind of magic-what self-reports can reveal about the phenomenology of insight problem solving. *Front Psychol, 5*, 1408. doi: 10.3389/fpsyg.2014.01408
Danek, A. H., Ollinger, M., Fraps, T., Grothe, B., & Flanagin, V. L. (2015). An fMRI investigation of expectation violation in magic tricks. *Front Psychol, 6*, 84. doi: 10.3389/fpsyg.2015.00084
Darwin, C. (1873). *The expression of the emotions in man and animals*. London: John Murray.
Demacheva, I., Ladouceur, M., Steinberg, E., Pogossova, G., & Raz, A. (2012). The Applied Cognitive Psychology of Attention: A Step Closer to Understanding Magic Tricks. *Applied Cognitive Psychology*, n/a n/a. doi: 10.1002/acp.2825
Dessoir, M. (1891). The Psychology of the Art of Conjuring. In H. J. Burlingame (Ed.), *Around the world with a magician and a juggler : unique experiences in many lands : from the papers of the late Baron Hartwig Seeman, "The Emperor of Magicians" and William D'Alvini, juggler "Jap of Japs"* (pp. 137-165). Chicago: Clyde Publishing Co.

Dessoir, M. (1893). The Psychology of Legerdemain. *The Open Court, 7(291), 3599–3601; 7(292), 3608–3611; 7(293), 3616–3619; 7(294), 3626–3627; 7(295), 3633–3634*.
During, S. (2002). *Modern Enchantments: The cultural power of secular magic*. Cambridge, Massachusetts: Harvard University Press.
Earl, B. (2016). This is not a box. Self published: Benjamin Earl.
Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J Pers Soc Psychol, 17*(2), 124.
Ekroll, V., Sayim, B., & Wagemans, J. (2017). The Other Side of Magic: The Psychology of Perceiving Hidden Things. *Perspect Psychol Sci, 12*(1), 91-106. doi: 10.1177/1745691616654676
Ekroll, V., & Wagemans, J. (2016). Conjuring Deceptions: Fooling the Eye or Fooling the Mind? *Trends Cogn Sci, 20*(7), 486-489. doi: 10.1016/j.tics.2016.04.006
Elimelech, C. (2015). Magic… Evokes Amazement and Wonder. from: http://www.magiciantoronto.com/magic-evokes-amazement-and-wonder/
Faber, A. (2012). *Investigation of Insight with Magic Tricks: Introducing a Novel Paradigm.* (PhD in Neurosciences), Ludwig-Maximilians-Universität, München, Munich.
Fitzkee, D. (1945). *Magic By Misdirection*. San Rafael, CA: San Rafael House.
Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cogn Psychol, 81*, 74-116. doi: 10.1016/j.cogpsych.2015.08.004
Gallup, G. G., & Maser, J. D. (1977). Catatonia - Tonic immobility: Evolutionary underpinnings of human catalepsy and catatonia. In J. D. M. M. E. P. Seligman (Ed.), *Psychopathology: Experimental models* (pp. 334-357). New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co.
Griffith T. L. (2015). Revealing ontological commitments by magic. *Cognition, 136*, 43-48. doi: http://dx.doi.org/10.1016/j.cognition.2014.10.019
Hahne, N. C. (1929). Deceptive Psychology *The Sphinx, 28*(5), 174.
Harris, P. (1992). Art of Astonishment. 3 volumes.:, 1992. *A1-Multimedia, Rancho Cordova CA*.
Hedne, M. R., Norman, E., & Metcalfe, J. (2016). Intuitive Feelings of Warmth and Confidence in Insight and Noninsight Problem Solving of Magic Tricks. *Front Psychol, 7*, 1314. doi: 10.3389/fpsyg.2016.01314
Hergovich, A., Gröbl, K., & Carbon, C. C. (2011). The paddle move commonly used in magic tricks as a means for analysing the perceptual limits of combined motion trajectories. *Perception, 40*(3), 358-366. doi: 10.1068/p6866
Hergovich, A., & Oberfichtner, B. (2016). Magic and Misdirection: The Influence of Social Cues on the Allocation of Visual Attention While Watching a Cups-and-Balls Routine. *Front Psychol, 7*, 761. doi: 10.3389/fpsyg.2016.00761
Hodgson, R., & Davy, S. J. (1887). The possibilities of mal-observation and lapse of memory from a practical point of view. *Proceedings of the Society for Psychical Research*.
Hutton, R., Klein, G., & Wiggins, S. (2008). *Designing for sensemaking: A macrocognitive approach*. Paper presented at the Computer Human Interactions 2008, Florence, Italy:.
Jastrow, J. (1896). Psychological Notes Upon Sleight-of-Hand Experts. *Science, 3*(71), 685-689. doi: 10.1126/science.3.71.685
Jastrow, J. J. (1888). The Psychology of Deception. *Popular Science Monthly, 34*, 145-157.
Jones, G. (2011). *Trade of the tricks: Inside the magician's craft*. Bekeley and Los Angeles: University of California Press.
Jones, G., & Shweder, L. (2003). The Performance of Illusion and Illusionary Performatives: Learning the Language of Theatrical Magic. *Journal oflinguistic Anthropolog, 13*(1), 51-70.
Klein, G., Moon, B., & Hoffman, R. R. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems, 21*(5), 88-92.
Klein, G., Philips, J. K., Rall, E. L., & Peluso, D. A. (2007). A Data-Frame Theory of Sensemaking. In R. Hoffman (Ed.), *Expertise Out of Context: Proceedings of the Sixth Conference on Naturalistic Decision Making* (pp. 113-155). New York, NY: Lawrence Erlbaum Associates.
Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Human Centered Computing, May/June 2003*, 81-85.
Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). *Rapid Decision Making on the Fire Ground.* Paper presented at the Human Factors and Ergonomics Society 30th Annual Meeting, Dayton, Ohio.
Kuhn, G., Amlani, A. A., & Rensink, R. A. (2008). Towards a science of magic. *Trends Cogn Sci, 12*(9), 349 354. doi: 10.1016/j.tics.2008.05.008
Kuhn, G., Caffaratti, H. A., Teszka, R., & Rensink, R. A. (2014). A psychologically-based taxonomy of misdirection. *Front Psychol, 5*, 1392. doi: 10.3389/fpsyg.2014.01392
Kuhn, G., & Findlay, J. M. (2010). Misdirection, attention and awareness: inattentional blindness reveals temporal relationship between eye movements and visual awareness. *Q J Exp Psychol (Hove), 63*(1), 136 146. doi: 10.1080/17470210902846757
Kuhn, G., & Martinez, L. M. (2011). Misdirection - past, present, and the future. *Front Hum Neurosci, 5*, 172. doi: 10.3389/fnhum.2011.00172

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition, 17*(6), 925-944. doi: 10.1080/13506280902826775
Lamont, P., & Wiseman, R. (2005). *Magic in Theory*. Hertfordshire: University of Hertfordshire Press.
Lamont, P., & Wiseman, R. (n.d.). Seeing and Believing. Retrieved 15/06/2012, from: http://www.richardwiseman.com/resources/MAGIC article.pdf
Landman, T. (2013). Framing Performance Magic: The Role of Contract, Discourse and Effect. *Journal of Performance Magic, 1*(1), 47-68.
Lanir, Z. (1991). The reasonable choice of disaster. In J. Rasmussen, B. Brehmer, J. Leplat, & Werner-Reimers-Stiftung (Eds.), *Distributed Decision Making: Cognitive Models for Cooperative Work* (pp. 215-230). Oxford: Wiley.
Leddington, J. (2016). The Experience of Magic. *The Journal of Aesthetics and Art Criticism, 74*(3), 253-264.
Leddington, J. (in press 2017). Magic: Art of the Impossible. In D. Goldblatt, L. Brown, & S. Patridge (Eds.), *A Reader in the Philosophy of the Arts, 4th Edition*. Oxford: Routledge.
Luna, T. (Producer). (2013, 19/01/2017). Taken by Surprise. Retrieved' Date Accessed 'from: https://www.youtube.com/watch?v=6JrbL4gBg-c
Luna, T., & Renninger, L. (2015). *Surprise: Embrace the Unpredictable and Engineer the Unexpected*. New York, New York: Perigee Trade.
Macknik, S., Martinez-Conde, S., & Blakeslee, S. (2011). *Sleights of Mind: What the neuroscience of magic reveals about our everyday deceptions*. London: Profile Books.
Macknik, S. L., King, M., Randi, J., Robbins, A., Teller, Thompson, J., & Martinez-Conde, S. (2008). Attention and awareness in stage magic: turning tricks into research. *Nat Rev Neurosci, 9*(11), 871-879. doi: 10.1038/nrn2473
Macknik, S. L., & Martinez-Conde, S. (2009). Correspondence in response to article: Real magic: future studies of magic should be grounded in neuroscience, by Lamont, Henderson & More. Nature Rev. Neurosci. 9, 871–879 (2008). *Nature Reviews: Neuroscience*. doi: 10.1038/
Martinez-Conde, S., & Macknick, S. (2008). Magic and the Brain. *Scientific American Mind, December 2008*, 72-79.
Martinez-Conde, S., & Macknik, S. L. (2007). Mind Tricks: Cognitive scientists take a lesson from magicians. *Nature, 448*(26), 414.
McGill, O. (1996). *The New Encyclopedia of Stage Hypnotism*. Carmarthen: Crown House Publishing Limited.
Memmert, D. (2010). The gap between inattentional blindness and attentional misdirection. *Conscious Cogn, 19*(4), 1097-1101. doi: 10.1016/j.concog.2010.01.001
Mohr, C., Koutrakis, N., & Kuhn, G. (2014). Priming psychic and conjuring abilities of a magic demonstration influences event interpretation and random number generation biases. *Front Psychol, 5*, 1542. doi: 10.3389/fpsyg.2014.01542
Moss, S. A., Irons, M., & Boland, M. (2016). The magic of magic: The effect of magic tricks on subsequent engagement with lecture material. *Br J Educ Psychol*. doi: 10.1111/bjep.12133
Most, S. B. (2010). What's "inattentional" about inattentional blindness? *Conscious Cogn, 19*(4), 1102-1104; discussion 1107-1109. doi: 10.1016/j.concog.2010.01.011
Munnich, E. L., Ranney, M. A., & Song, M. (2007). *Surprise, surprise: The role of surprising numerical feedback in belief change.* Paper presented at the Proceedings of the twenty-ninth annual conference of the cognitive science society. Erlbaum, Mahwah, NJ.
Nash, R. F., Gallup, G. G., & Czech, D. A. (1976). Psychophysiological correlates of tonic immobility in the domestic chicken (Gallus gallus). *Physiology & Behavior, 17*(3), 413-418. doi: http://dx.doi.org/10.1016/0031-9384(76)90100-1
Olson, J. A., Amlani, A. A., Raz, A., & Rensink, R. A. (2015). Influencing choice without awareness. *Conscious Cogn, 37*, 225-236. doi: 10.1016/j.concog.2015.01.004
Olson, J. A., Demacheva, I., & Raz, A. (2015). Explanations of a magic trick across the life span. *Front Psychol, 6*, 219. doi: 10.3389/fpsyg.2015.00219
Olson, J. A., Landry, M., Appourchaux, K., & Raz, A. (2016). Simulated thought insertion: Influencing the sense of agency using deception and magic. *Conscious Cogn, 43*, 11-26. doi: 10.1016/j.concog.2016.04.010
Ortiz, D. (1994). *Strong Magic: Creative Showmanship for the Close-Up Magician*. Washington, WA: Ortiz Publications.
Ortiz, D. (2006). *Designing Miracles*. El Dorado Hills, CA: A1 MagicalMedia.
Otero-Millan, J., Macknik, S. L., Robbins, A., & Martinez-Conde, S. (2011). Stronger misdirection in curved than in straight motion. *Front Hum Neurosci, 5*, 133. doi: 10.3389/fnhum.2011.00133
Parris, B. A., Kuhn, G., Mizon, G. A., Benattayallah, A., & Hodgson, T. L. (2009). Imaging the impossible: an fMRI study of impossible causal relationships in magic tricks. *Neuroimage, 45*(3), 1033-1039. doi: 10.1016/j.neuroimage.2008.12.036
Pfleegor, A. G., & Roesenberg, D. (2013). Deception in Sport: A New Taxonomy of Intra-Lusory Guiles. *Journal of the Philosophy of Sport, 41*(2), 209-231. doi: 10.1080/00948705.2013.785424
Phillips, F., Natter, M. B., & Egan, E. J. (2015). Magically deceptive biological motion-the French Drop Sleight. *Front Psychol, 6*, 371. doi: 10.3389/fpsyg.2015.00371

Quiroga, R. Q. (2016). Magic and cognitive neuroscience. *Curr Biol, 26*(10), R390-394. doi: 10.1016/j.cub.2016.03.061

Ramírez, E. R. R. (2014). Industrial design strategies for eliciting surprise. *Design Studies, 35*(3), 273-297. doi: 10.1016/j.destud.2013.12.001

Randall, J. (2009). The Curious Case of Doug Henning. Retrieved 17/01/2017, from: http://magicgizmo.com/home/Magic-History/the-curious-case-of-doug-henning.html

Ratner, S. C. (1967). Comparative aspects of hypnosis. In J. E. Gordon (Ed.), *Handbook of clinical and experimental hypnosis* (pp. 335-371): New York: Macmillan.

Rensink, R. A., & Kuhn, G. (2014). A framework for using magic to study the mind. *Front Psychol, 5*, 1508 doi: 10.3389/fpsyg.2014.01508

Rieiro, H., Martinez-Conde, S., & Macknik, S. L. (2013). Perceptual elements in Penn & Teller's "Cups and Balls" magic trick. *PeerJ, 1*, e19. doi: 10.7717/peerj.19

Rimé, B., Philippot, P., Boca, S., & Mesquita, B. (1992). Long-lasting Cognitive and Social Consequences of Emotion: Social Sharing and Rumination. *European Review of Social Psychology, 3*(1), 225-258. doi: 10.1080/14792779243000078

Rissanen, O., Palonen, T., Pitkänen, P., Kuhn, G., & Hakkarainen, K. (2013). Personal Social Networks and the Cultivation of Expertise in Magic: An Interview Study. *Vocations and Learning, 6*(3), 347-365. doi: 10.1007/s12186-013-9099-z

Rissanen, O., Pitkanen, P., Juvonen, A., Kuhn, G., & Hakkarainen, K. (2014). Expertise among professional magicians: an interview study. *Front Psychol, 5*, 1484. doi: 10.3389/fpsyg.2014.01484

Rockwood, P. R. (1919). The Psychology Of Magic. *The Sphinx, 18*(10), 256-260.

Rolfe, C. (2015). Theatrical magic and the agenda to enchant the world. *Social & Cultural Geography*, 1-23. doi: 10.1080/14649365.2015.1112025

Saville, I. (2013). The Development of Socialist Magic: Reflections on the Place of Power and Ideology in Magic Performance. *Journal of Performance Magic, 1*(1), 2-18.

Schomaker, J., & Meeter, M. (2015). Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neurosci Biobehav Rev, 55*, 268-279. doi: 10.1016/j.neubiorev.2015.05.002

Schützwohl, A., & Reisenzein, R. (2012). Facial expressions in response to a highly surprising event exceeding the field of vision: a test of Darwin's theory of surprise. *Evolution and Human Behavior, 33*(6), 657-664. doi: 10.1016/j.evolhumbehav.2012.04.003

Shalom, D. E., de Sousa Serro, M. G., Giaconia, M., Martinez, L. M., Rieznik, A., & Sigman, M. (2013). Choosing in freedom or forced to choose? Introspective blindness to psychological forcing in stage-magic. *PLoS ONE, 8*(3), e58254. doi: 10.1371/journal.pone.0058254

Shapiro, K. L., Raymond, J., & Arnell, K. (1997). The attentional blink. *Trends Cogn Sci, 1*(8), 291-296.

Shtulman, A., & Morgan, C. (2017). The Explanatory Structure of Unexplainable Events: Causal Constraints on Magical Reasoning. *In Press, Psychonomic Bulletin & Review*. doi: 10.3758/s13423-016-1206-3

Smith, T. J. (2015). The role of audience participation and task relevance on change detection during a card trick. *Front Psychol, 6*, 13. doi: 10.3389/fpsyg.2015.00013

Smith, T. J., Lamont, P., & Henderson, J. M. (2012). The penny drops: Change blindness at fixation. *Perception, 41*(4), 489-492. doi: 10.1068/p7092

Smith, T. J., Lamont, P., & Henderson, J. M. (2013). Change Blindness in a Dynamic Scene Due to Endogenous Override of Exogenous Attentional Cues. *Perception, 42*(8), 884-886. doi: 10.1068/p7377

Smith, W., Dignum, F., & Sonenberg, L. (2016). The Construction of Impossibility: A Logic-Based Analysis of Conjuring Tricks. *Front Psychol, 7*, 748. doi: 10.3389/fpsyg.2016.00748

Söderlund, M. (1998). Customer satisfaction and its consequences on customer behaviour revisited: The impact of different levels of satisfaction on word-of-mouth, feedback to the supplier and loyalty. *International Journal of Service Industry Management, 9*(2), 169-188.

Subbotsky, E. (1997). Explanations of unusual events: Phenomenalistic causal judgements in children and adults. *British Journal of Developmental Psychology, 15*(13-36).

Subbotsky, E. (2001). Causal explanations of events by children and adults: Can alternative causal modes coexist in one mind? *British Journal of Developmental Psychology, 19*(1), 23-45. doi: 10.1348/026151001165949

Subbotsky, E. V. (1996). Explaining impossible phenomena: Object Permanence Beliefs andMemory Failures in Adults. *Memory, 4*(2), 199-223.

Swiss, J. I. (2002). The Search for Mystery. In J. Jay (Ed.), Magic in Mind: Essential Essays for Magicians. Retrieved 17/01/2017, from: https://www.vanishingincmagic.com/magic-downloads/ebooks/magic-in-mind/.

Tachibana, R., & Gyoba, J. (2015). Effects of different types of misdirection on attention and detection performance. *Tohoku Psychologica Folia, 74*, 42-56.

Tachibana, R., & Kawabata, H. (2014). The effects of social misdirection on magic tricks: How deceived and undeceived groups differ. *Iperception, 5*(3), 143-146. doi: 10.1068/i0640sas

Tatler, B. W., & Kuhn, G. (2007). Don't look now: The magic of misdirection. In R. v. Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye Movement Research: Insights into Mind and Brain* (pp. 696-713): Elsevier.

Thomas, C., & Didierjean, A. (2016a). The ball vanishes in the air: can we blame representational momentum? *Psychon Bull Rev, 23*(6), 1810-1817. doi: 10.3758/s13423-016-1037-2
Thomas, C., & Didierjean, A. (2016b). No need for a social cue! A masked magician can also trick the audience in the vanishing ball illusion. *Atten Percept Psychophys, 78*(1), 21-29. doi: 10.3758/s13414-015-1036-9
Thomas, C., & Didierjean, A. (2016c). Scientific Study of Magic: Binet's Pioneering Approach Based on Observations and Chronophotography. *American Journal of Psychology, 129*(3), 315–328.
Thomas, C., Didierjean, A., Maquestiaux, F., & Gygax, P. (2015). Does magic offer a cryptozoology ground for psychology? *Review of General Psychology, 19*(2), 117-128. doi: 10.1037/gpr0000041
Tomkins, S. S. (1963). *Affect imagery consciousness: Volume II: The negative affects*. New York: Springer Publishing Company.
Tompkins, M. L. (2016). A Science of Magic Bibliography. Retrieved 12/12/2016, from: http://www.matt-tompkins.co/blog/2016/9/23/a-science-of-magic-bibliography
Tompkins, M. L., Woods, A. T., & Aimola Davies, A. M. (2016). The Phantom Vanish Magic Trick: Investigating the Disappearance of a Non-existent Object in a Dynamic Scene. *Front Psychol, 7*, 950. doi: 10.3389/fpsyg.2016.00950
Trinkaus, J. (1980). Preconditioning An Audience For Mental Magic: An Informal Look. *Percept Mot Skills, 51*, 256.
Triplett, N. (1900). The Psychology of Conjuring Deceptions. *The American Journal of Psychology, 11*(4), 439-510. doi: 10.2307/1412365
Van de Cruys, S., Wagemans, J., & Ekroll, V. r. (2015). The put-and-fetch ambiguity: How magicians exploit the principle of exclusive allocation of movements to intentions. *i-Perception, 6*(2), 86-90. doi: 10.1068/i0719sas
Völgyesi, F. A., & Klumbies, G. (1966). *Hypnosis of man and animals*: Baillière.
Whaley, B. (2007). *Stratagem: Deception And Surprise In War*. Boston: Artech House.
Whaley, B. (2013). The principle of naturalness. In H. Rothstein & B. Whaley (Eds.), *The Art and Science of Military Deception* (pp. 41-43). Boston: Artech House.
Williams, H., & McOwan, P. W. (2014). Magic in the machine: a computational magician's assistant. *Front Psychol, 5*, 1283. doi: 10.3389/fpsyg.2014.01283
Williams, H., & McOwan, P. W. (2016). Magic in Pieces: An Analysis of Magic Trick Construction Using Artificial Intelligence as a Design Aid. *Applied Artificial Intelligence, 30*(1), 16-28. doi: 10.1080/08839514.2015.1121068
Willis, R. H. (2002). Surprise. In W. E. Craighead & C. B. Nemeroff (Eds.), *The Corsini Encyclopedia of Psychology and Behavioral Science* (Vol. 4, pp. 1660-1663).
Wilson, K., & French, C. C. (2014). Magic and memory: using conjuring to explore the effects of suggestion, social influence, and paranormal belief on eyewitness testimony for an ostensibly paranormal event. *Front Psychol, 5*, 1289. doi: 10.3389/fpsyg.2014.01289
Wiseman, R. J., & Nakano, T. (2016). Blink and you'll miss it: the role of blinking in the perception of magic tricks. *PeerJ, 4*, e1873. doi: 10.7717/peerj.1873

Developing Adaptive Expertise: A Synthesis of Literature and Implications for Training

Rob HUTTON[a], Paul WARD[b], Julie GORE[c], Patrick TURNER[d], Robert HOFFMAN[e], Andrew LEGGATT[a] & Gareth CONWAY[f]

[a]*Trimetis Ltd, UK*

[b]*University of Huddersfield, UK,*

[c]*University of Bath, UK*

[d]*QinetiQ UK*

[e]*Institute of Human Machine Cognition, USA*

[f]*Defence Science & Techology Laboratory, Ministry of Defence, UK*

ABSTRACT

This paper reports an on-going project to explore the challenge of developing adaptive performance skills. The paper describes an initial critical synthesis of literature which highlights the conceptual emphasis of the existing literature and a lack of empirical evidence for the efficacy of training interventions. We provide an operational definition of adaptive performance which highlights critical cognitive skills. Useful training principles from the existing literature are presented, and initial concepts for concrete training interventions from those principles are suggested. Initial experiences implementing the training interventions will also be presented if available.

KEYWORDS

Practical application; Macrocognition; Learning and training; Adaptive performance; adaptive thinking.

INTRODUCTION

The concept of adaptive expertise has been popular in both the applied practitioner and academic literature since the mid-1980s. As the work environment becomes more volatile and uncertain there is a need for 'experts who possess the required organisational domain expertise and can quickly overcome changes' (Bohle Carbonell, et al., 2014; 2015). It is asserted that such employees possess 'adaptive expertise'.

However, there are concerns that the concept of 'adaptive expertise' is not an empirically robust one and that an examination of its application and theoretical evolution may be worthwhile in order to inform training policy recommendations and development, in particular in high-reliability organizations. A scoping study, described in Ward et al. (2016) presents the findings from a critical interpretive synthesis of the extant literature. The aim of this paper is to very briefly review that synthesis, identify training principles that suggest benefits with respect to the development of adaptive performance and adaptive thinking skills, and present some initial concrete instantiations of those training principles for a military operational headquarters context.

Background

In contrast to their description of a routine expert, Hatano and Inagaki (1984/1986) described an adaptive expert as someone "who not only perform[s] procedural skills efficiently but also understand[s] the meaning of the skills and nature of their object" (p.28). Later, they simplified their description of adaptive expertise as a combination of procedural and conceptual knowledge (or understanding)—rather than of procedural skill/knowledge alone. Subsequently, they suggested that conceptual understanding is the primary basis for being flexible and adaptive (see 1984, p.30). They qualified the notion of 'performing procedural skills with understanding' (i.e., adaptive expertise) as being "when the performer can explain why it works, i.e., verbalize the principle involved; or at the least, when he/she can judge, in addition to the conventional version of the [procedural] skill, its variations as appropriate or inappropriate, and/or can modify the skill according to changes in constraints" (see 1984, p.28).
In addition to affording the kind of in-event adaptive thinking process and outcome described above, Hatano & Inagaki noted that a well-developed, conceptual mental model provided the opportunity to mentally simulate instances where adaptation may be necessary. In addition to deviating from existing procedures in situ, they speculated that mental simulation provided offline opportunities to explain (either explicitly and verbally, or implicitly as reference to a 'vague' mental image) unfamiliar situations, make new predictions and invent new procedures (see 1984, p.28).

Contact Information of First Author: rob.hutton@trimetis.co.uk

To summarise, Hatano and Inagaki (1984 / 1986,) described two of the core characteristics of adaptive expertise: efficiency and innovation (see also Schwartz, et al., 2005) and contrasted those with the procedural efficiency of routine experts.

Key Components of Adaptive Expertise

Conceptualisations of adaptive expertise and its development have often reduced the problem to just two of the core expertise characteristics identified above. While efficiency is, perhaps, more characteristic of routine expertise, the assumption is that it is a characteristic of all experts, and a starting point from which adaptive expertise can be developed. Innovation, on the other hand, relates to an adaptive expert's capability at handling novelty or change, i.e., modifying old and/or creating new methods to deal with changes in situational complexity, familiarity and frequency (see Bohle Carbonell, et al. 2014; Hatano and Inagaki, 1984; Schwartz, et al., 2005; see also Hoffman, 1998). Whether this component of adaptive expertise is an underlying ability or an acquired skill is open to debate. Although efficiency and innovation are often viewed as competing constructs that are mutually exclusive, in reality they should be viewed as orthogonal dimensions both of which can be developed (see Schwartz, et al., 2005).

Consistent with Hatano and Inagaki (1984/1986), a central feature of the adaptive expertise literature is consensus on the notion that adaptive expertise emanates from a well-developed conceptual understanding and associated knowledge structures that are contextually sensitive and malleable

In addition to the work of Hatano and Inagaki, current perspectives on adaptive expertise (e.g., Bohle Carbonell, et al., 2014; Schwartz et al., 2005) are often based, at least in part, on the empirical findings from Rand Spiro and Paul Feltovich's work on cognitive flexibility (e.g., Spiro, et al., 1988, 1995; Feltovich, et al., 1997). A central feature of the notion of cognitive flexibility is that for knowledge to be useful (i.e., highly accessible whenever needed in any relevant context and in a myriad of different circumstances), it has to be experienced, acquired, taught, organized, and mentally represented in different ways. When knowledge is not acquired flexibly, its use is limited to situations that resemble the initial learning context alone. The initial context, however, constitutes only a fraction of the situations where that knowledge may be applicable. When knowledge structures are acquired flexibly, knowledge assemblies can be built "*to fit the diverse future cases of knowledge application in that domain"* (Spiro et al., 1995, p.67). The ideas of flexible knowledge acquisition and constructed knowledge assemblies are consistent with the current theories of accelerated expertise in complex domains (e.g., Cognitive Transformation and Cognitive Flexibility [CT-CF] Theory merger proposed by Hoffman, Ward, et al., 2014). They are also consistent with the mental modelling processes that support adaptation (as described by Hatano and Inagaki) and high levels of proficiency (e.g., Klein, et al., 2003), and with the underlying data-frame mechanisms proposed in recent theories of expert sensemaking (e.g., Klein et al., 2006).

Recent literature on adaptive expertise has drawn on related constructs (i.e., cognitive flexibility), but has largely focused on identifying individual difference characteristics of expertise (e.g., Bohle, et al., 2014, 2015). While frequently acknowledged, with few exceptions, there has been less emphasis on adaptation being based on one's conceptual understanding, or on the cognitive mechanisms responsible for successful adaptation (cf. Hoffman, Ward, et al., 2014; Feltovich, et al., 1997; Klein and Baxter, 2006; Hoffman, Best, and Klein, 2014).

METHOD

Based on the assertions identified above, we conducted a synthesis of existing literature that has attempted to explore ways to develop adaptive performance and adaptive thinking. We conducted searches using 94 search-term combinations, and conducted an initial review of 1995 abstracts. From these abstracts, we subjectively identified those sources most relevant to our goals (n = 72). Based on subsequent reading of these and related articles, we collated a database of approximately 140 publications.

In addition we consulted with over 30 stakeholders working in high reliability contexts via two workshops and eight interviews in order to obtain practitioner perspectives on adaptive expertise.

Two key findings derive from this review. First, the literature base on adaptive expertise is largely conceptual. Empirical data are sparse indeed. Hence, our attempts at employing a systematic empirical evaluation framework met with limited success. Where empirical data existed, which was not often, primarily it pertained to measures of transfer of knowledge (which we have reviewed extensively elsewhere, see Hoffman, Ward, et al., 2014). Even when authors explicitly stated that an "adaptive performance test" or "assessment of adaptive expertise" was implemented, typically the measures used were of transfer rather than adaptivity or adaptive expertise *per se* (e.g., Shadrick, 2006). Second, there is a confusing array of terms used, often interchangeably, when referring to adaptivity or skilled adaptation of some kind.

Research synthesis

Over 140 resources were reviewed, from both the academic (journals and peer-reviewed journals and book/chapters) and the applied (e.g. government technical reports) literature bases. A wide range of perspectives, disciplines and objectives were captured in the review, including work from learning, experimental, occupational (industrial-organisational) and human factors psychology, education, and training primarily but not exclusively.

The concept of adaptive expertise spans at least three key areas of literature: (i) adaptivity, (ii) skilled performance, and (iii) skill development. Expertise is partly defined by the ability of the expert to skilfully adapt their performance to the changing task demands. The literature on the development of expertise and skilled performance, transfer of learning from one context to another (e.g., a closely related task or novel task), and on learning, training and development across different levels of proficiency, also provided rich sources of related evidence.

ANALYSIS

Defining Adaptive Expertise

Following the review of very diverse definitions of adaptive expertise in the literature, and of the inputs of stakeholders with respect to how they viewed adaptive expertise in high-reliability contexts, an operational definition for this context was proposed: '*Timely changes in understanding, plans, goals, and methods in response to either an altered situation or updated assessment about the ability to meet new demands, that permit successful efforts to achieve intent*.' This definition encompasses three key elements of adaptive performance for individuals and teams: (i) Understanding of a situation, (ii) actions required to achieve intent, and (iii) self-awareness to balance the situational and task demands with the ability of the individual (and the resources at his or her disposal) to achieve the intent (represented in Figure 1).

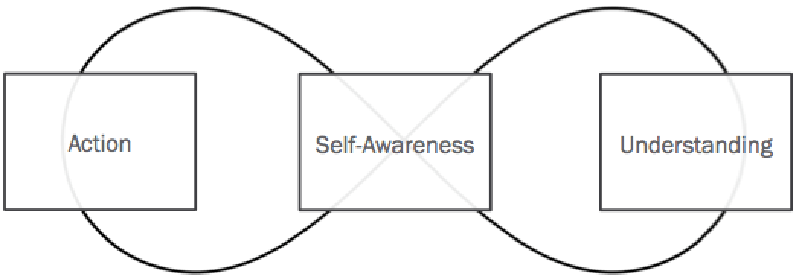


Figure 5. Key elements of adaptive performance for individuals and teams.

Training Adaptive Performance and Adaptability

A brief review of existing efforts to develop adaptive expertise revealed a number of interesting aspects, challenges and concerns. In summary, the UK view is that adaptable performance is valued, and it is implicit in current selection and training efforts but with no way of knowing how well it is being trained or how effective the training is. There is a recognition that thinking skills/cognitive skills are important to leadership and adaptive thinking/performance and there are some efforts, being led by Communications and Behavioural Science (CABS) at Royal Military Academy Sandhurst (British Army initial Officer training), to explicitly teach and train thinking skills. These have been adopted by the Royal Naval College, Dartmouth. The Royal Air Force are also revisiting their thinking skill requirements. But there is no clear direction or guidance about how to handle cognitive skills, particularly with respect to improving adaptability. There is a question as to whether adaptability should be "the" focus, or whether it is still acceptable to have it embedded implicitly. Why should adaptability be favoured over critical thinking? Or analytical rigour? Or other forms of cognitive skills? How should the UK Ministry of Defence (MOD) address adaptability? Our scoping study identified an increasing number of efforts in this direction and a desire to address adaptive performance head on. The future character of conflict (an increasingly dynamic technological environment that presents challenges and opportunities for the way we work and the way the enemy works) appears to demand adaptivity.

It was recommended that the development of adaptive performance be addressed explicitly within the MOD and that it issues a mandate for exploring improvements to the way that it currently develops adaptive performance (through education, training and development opportunities/continuing professional development/self-learning, institutional learning and operational learning). The requirement for the current study has indicated an initial effort to tackle the problem systematically. The scoping study began the process to provide an initial conceptual framework and language to support this effort, as well as training/development guidance.

The US DOD has already identified adaptive performance as critical and has provided a mandate (of sorts) through the Army Learning Concept 2015, based on anecdotal evidence (because there is no empirical evidence based on our own analysis of the research literature). However, they are still exploring what adaptive performance is, how to train it and how to assess it. None of these is straightforward. They have been working towards it for over 20 years with the advent of the various decision making and critical thinking development efforts in the 90s (Ward et al., 2016). Across the US and UK there appears to be a lack of concrete training requirements, interventions or training solutions, performance measurement techniques, evidence for efficacy, and support to the training community to evaluate the compliance and efficacy of any new programmes. Our recommendations are intended to address some of these shortfalls.

Training Principles

Based on a review of the current literature with respect to empirically based suggestions for immediate application to training, several training principles were described in order to begin to address the shortfall identified above. Fundamental to the achievement of adaptive performance is the opportunity for practice at problems that stretch current competency, and that permit acquisition of knowledge and reasoning skills exercised in differing situations or contexts (i.e., varieties of tough, rare tasks sampled from across the ecosystem) in safe-to-fail environments. We identified six candidate training principles that we expect will facilitate the development and acceleration of adaptive performance skill.

Flexibility-Focused Feedback. Instructors need to be able to provide feedback about performance that overcomes the human tendency to rigidly stick with 'what works', especially when the situation changes or when there is a misunderstanding about 'why it works'. Feedback given during learning opportunities should permit learners (a) to learn when their current strategies work effectively and when they do/will not, (b) to promote the development of new strategies and conceptions that permit them to demonstrate effective flexibility and adaptation to new situations or unexpected changes in the training exercise or learning experience, and (c) to quickly re-assess and re-appraise their current interpretation of a situation so that they can 'reassemble' current knowledge of a situation in flexible ways (including adding to it, subtracting from it, or even abandoning it entirely in favour of a new conception).

*Concept-Case Coupling***.** (Combining context and cognition) When aspects of adaptive performance and the required mental skills are presented in training they must be presented in conjunction with the situational factors that demand adaptations. Not only that, but the full range of situational contexts must be presented in order to illustrate the structural patterns in the situations that trainees must understand and pay attention to, and to minimize the 'transfer distance' between what they know and the situation to which they have to apply that knowledge. In sum, the more cases, and the more varied the number of cases, the more likely that concept will be learned flexibly, and adaptive performance developed.

*Tough Case Time Compression***.** Trainers should attempt to present a number of particularly difficult/low-frequency scenarios in order to increase the rate of challenging learning opportunities that are experienced.

*Case-Proficiency Scaling***.** The test bank of tough cases needs to be scaled to the trainee's proficiency level to permit current skills to be stretched. The intent is to maintain training at the edge of the 'zone of proximal development' (i.e. the edge of the current levels of proficiency, knowledge and experience), which is a moving target as individuals learn and improve.

*Complexity Preservation***.** This principle requires learners to learn with complexity to facilitate adaptation to novelty and changing demands, not to be presented with overly simplified scenarios that merely train procedures or Standard Operating Procedures. Trainers should shift learning away from conceptions of 'knowledge storage' to conceptions of 'thinking dynamically'. This requires constant reconfiguration of trainee understanding through experience of context variations / knowledge boundaries. It also requires regeneration and updating of mental models on-the-fly via dynamic and complex scenarios and requirements for anticipatory thinking.

*Active Reflection***.** Training for adaptive performance should facilitate critical thinking and reflective practice. Four key exercises are proposed to promote metacognitive awareness:
(i) Prediction or estimation of one's competency on a specific task, success on a given case, or learning (e.g., how long it will take, the level of performance attained, etc.) relative to actual (future or past) performance level and/or learning rate.
(ii) Experimentation with new cases, contexts, situations, methods, strategies and response actions.
(iii) Extrapolation of experience with prior incidents to new incidents.

(iv) Explanation (to self, to others, and by others) of what happened, why, and if/how success was achieved (through facilitated, inclusive After Action Review, for example).

DISCUSSION

The training principles identified above are quite abstract, in some cases are already considered best practice in non-cognitive areas of training, and could be applied in a training context on a variety of ways. One realisation is that adaptive performance is less likely to be supported by training the target audience with specific techniques or providing them with ‘tools’ to enhance adaptive thinking. The application of the training principles are more likely to be made concrete in the context of the training techniques, tools for trainers, facilitators, observer/mentors and performance evaluators. Rather than providing thinking techniques or processes per se, the general approach is akin to developing “habits of mind” and improved mental models which capture situation-action patterns, and the sense of typicality required to support anomaly detection and the “need to adapt.”

Figure 2 presents a mapping of training principles to potential concrete training interventions.

Figure 6 Mapping of training principles for adaptive performance to potential training system interventions.

Notice that the training interventions on the right side of Figure 2 are less about providing thinking techniques for becoming a (more) adaptive thinker/decision maker, and are less about the target training audience per se. Most of them are targeting changes to the “training system” in which the training is delivered: the scenario development and sequencing, the assessment/measurement process, facilitation of debriefs, after action reviews, and reflection, observation and mentoring guidance.

The interventions identified above are in the process of being implemented and tested in two ways, through an initial “quick win” opportunity at an imminent operational exercise, as well as in a more controlled, quasi-experimental way as part of a longer term project.

CONCLUSION

The intended impact of this work is to support policy and guidance decisions for training and education in the area of developing the complex cognitive skills that support building adaptive expertise and skilled adaptive performance.

Notwithstanding a number of gaps in the empirical evidence to support the effectiveness of existing training approaches aimed specifically at adaptive performance, there is an operational imperative (in high reliability related organizations in particular) to continue to improve the development of adaptive individuals, teams and agile work in the future. Crucially, it is apparent from this review that the evidence of the effectiveness of the training principles for improvements in a number of the supporting complex cognitive skills required for adaptive performance is sufficient to warrant further development and testing in a range of contexts.

REFERENCES

Bohle Carbonell, K., Stalmeijer, R. E., Könings, K. D., Segers, M., and Van Merriënboer, J. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational Research Review* 12, 14-29.

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), *Expertise in context* (pp. 125–146). Menlo Park, CA: AAAI/MIT Press.

Gore, J., Flin, R., Stanton, N., and Wong, W. (2015). Applications *for* Naturalistic Decision Making. *Journal of Occupational and Organizational Psychology*. 88, 223-230.

Hatano, G., and Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, and K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York: W. H. Freeman.

Hoffman, R. R. (2015). *Macrocognitive Measures and Metrics.* Institute of Human Machine Cognition Internal Technical Report. Pensacola, FL: IHMC

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., and Andrews, D. (2014). *Accelerated expertise: Training for high proficiency in a complex world.* New York, NY: Psychology Press. http://www.psypress.com/books/details/9781848726529/

Klein, G. A. (1997). Developing expertise and decision making. *Thinking and Reasoning, 3,* 337–352.

Klein, G., Phillips, J. K., Rall, E. L., and Peluso, D. A. (2006). A Data/Frame Theory of Sensemaking. In Robert R. Hoffman (ed.), *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (Boca Raton, FL: Taylor and Francis), 117-118 (115-155).

Shadrick, S. B., Lussier, J. W., and Fultz, C. (2007). *Accelerating the development of adaptive performance: Validating the Think Like a Commander training* (No. ARI-RR-1868). Arlington, VA: Army Research Institute for the Behavioral and Social Sciences

Spiro, R. J., Coulson, R. L., Feltovich, P. J., and Anderson, D. K. (1988). *Cognitive Flexibility Theory: Advanced Knowledge Acquisition in Ill-Structured Domains* (Technical Report No. 441, October 1988). Centre for the Study of Reading, University of Illinois-UC.

Knowledge Elicitation in Naturalistic Decision Making: Collegial Verbalisation with "Conspective Protocols"

Anders JANSSON[a] and Anton AXELSSON[a]
[a]*Division of Visual Information and Interaction, Uppsala University, Sweden*

ABSTRACT
For knowledge elicitation in work environments where participants are highly experienced, there exist two established verbalisation protocols – concurrent and retrospective – both of which are associated with methodological and practical issues. A third protocol – "conspective" verbalisation – and its theoretical background are presented together with Collegial Verbalisation (CV), a method that synthesises the use of the protocols into a cohesive methodological framework. Results from use of the CV method in three domains are presented. The method contributes to the unravelling of mental strategies and an enhanced understanding of naturalistic decision making tasks. Independent observers comment in the form of conspective protocols on the behaviour of target participants. It solves some of the problems with the established verbalisation protocols. Analyses of the protocols show the importance of regularities and environmental constraints in the organisation of decision making activities in as diverse domains as train traffic control, high-speed ferry operation, and train driving.

KEYWORDS
Decision making; Expertise; Transportation; Knowledge elicitation; Verbalisation

INTRODUCTION

Research conducted in the field of naturalistic decision making aims for careful understanding of how professional decision makers think and act in their specific work environments. One common goal is to account for the decision makers' acquired experiences while they perform different decision tasks. From an analysis point of view, this means that it is not enough to understand *why* people behave the way they do, or *what* they do in each situation, but also *how* they accomplish the activities associated with a certain decision task. Analyses of how activities are accomplished can be carried out through a strategies analysis, with a focus on mental activities, either in the form of categorised cognitive processes (Rasmussen, Pejtersen, & Goodstein, 1994), task procedures (Vicente, 1999), as sequences of mental and effector operations (Payne, Bettman, & Johnson, 1993), as task performance approaches distinguished by costs and benefits (Hassall & Sanderson, 2014) or adaptive heuristics for dealing with dynamics (Brehmer, 1990; Jansson, 1995). In all cases, some form of knowledge elicitation is necessary.

For the purpose of knowledge elicitation in naturalistic work situations, there are several methods available. Among them are different types of verbalisation methods. Traditionally, verbal reporting is carried out through concurrent or retrospective verbalisation protocols. Both these established protocols are associated with a specific methodological challenge: there is no necessary correlation between the mental behaviour responsible for the actions taken in a certain decision task and the mental behaviour behind the verbal reports about the same actions (Bainbridge, 1979/1999), this means that there is no guarantee that what is verbalised is an actual account of the mental processes involved in the performance of the decision maker. Historically, there were strong doubts about verbalisations as data because of this vagueness about the validity of the verbal protocols. These doubts came to an end however with the seminal work by Ericsson and Simon (1980, 1984) when they were able to support their claim for verbal reports as data with a strong theoretical model and concurrent verbalisations in the form of think-aloud protocols. There is however another issue with concurrent verbalisation procedures: they may jeopardise the representativeness of the decision task due to the fact that the work task is disrupted with the additional task of verbalising (e.g. Ericsson & Simon, 1980, 1984; Ericsson & Crutcher, 1991; Bartl & Dörner, 1998; Dickson, McLennan, & Omodei, 2000). This can have severe consequences in naturalistic decision investigations because the participants cannot prioritise verbalisations without changing the way they think and act (Dickson et al., 2000). Furthermore, it is often difficult to verbalise skill-related knowledge during task completion because much of the knowledge is tacit (Polanyi, 1967). Regarding retrospective verbalisations, the challenges are even bigger. Firstly, one can expect that time delays will affect the remembering of the control actions negatively since the mental behaviours corresponding to these measures will decay from working memory (Gibbons, 1983; Ericsson & Simon, 1984). Secondly, verbalisers often focus on problems closer at hand and thus infrequent problems might be overlooked (Wright & Ayton, 1987). Thirdly, since there is no way to separate the mental behaviour responsible for the non-observable actions taken and the mental behaviours responsible for the verbal reporting, we cannot learn from empirical data if a decision maker carrying out a verbalisation retrospectively is rationalising his or her behaviour (van Someren, Barnard, & Sandberg, 1994; Bainbridge, 1979/1999).

As a reaction to these challenges, a number of studies have explored different procedures of having others verbalise the actions of target users by, for example, letting colleagues or domain experts verbalise rather than users themselves (e.g. Dominguez, Flach, McDermott, McKellar, & Dunn, 2004; Miller, Patterson, & Woods, 2006; McIlroy & Stanton, 2011; Jansson, Erlandsson, & Axelsson, 2015). The rationale behind the aim of having an independent observer verbalise instead of the target user is to avoid the privacy problem (Bainbridge, 1979/1999) since this is a critical source to the issues discussed above. So far, it has however been little or no progress on theoretical explanations motivating the use of independent observers verbalising the actions of target participants. However, recently Jansson et al. (2015) provided a theoretical model motivating the use of the Collegial Verbalisation (CV) method, including a “conspective” verbal protocol. We have termed this form of verbalisation “conspective” due to the fact that the verbaliser is *observing* whilst thinking aloud. This distinction is important because this verbalisation can be performed both in real-time or with recorded material. The new protocol fits neither under concurrent verbal protocol since the verbaliser is not performing the work task, only observing it, nor under the retrospective verbal protocol since the verbaliser is seeing the events unravel for the first time.

The rationale behind concurrent, retrospective, and conspective verbal protocols is roughly similar: to extract data about mental behaviour associated with domain specific decision making behaviour and performance. The CV method suggests investigators to video record target operators performing work. The critical part of the method is the conspective verbalisation where colleagues of the target users verbalise on the recorded material. In conjunction with the conspective protocol, an investigator can choose to use (1) a concurrent protocol during recording of the target operator, (2) a retrospective protocol with the target operator, or (3) may choose to use both. Our objective here is to argue for the usefulness and value of this method for the purpose of knowledge elicitation in naturalistic decisions, emphasizing the role of long-term memory knowledge structures.

The Model Behind the Method – Theoretical Motivation

The established verbal protocol methods suffer from methodological issues which originate from the problem of verification of the validity and reliability of verbal reports. Introducing an independent narrator as a verbaliser on user actions will solve some of these problems. This has been recognised by previous authors where domain experts have been employed to verbalise on students or practitioners (Miller et al., 2006), or on other experts (Dominguez et al., 2004; McIlroy & Stanton, 2011). The CV method has similarities in procedure to the stimulated recall interview (Calderhead, 1981), however, CV is thought of as a verbalisation procedure utilising two or three data generation points. In using the CV method, it is important to notice that both the verbalising target operators and their colleagues are instructed to think-aloud without interrupting them with remarks for interpretations or clarifications. They should be exposed as closely as possible to the same control task procedure. Secondly, one assumption behind the CV method is that environmental constraints will affect the behaviour of experienced target operators and make it possible for likewise experienced colleagues to utilise the effect of these constraints when

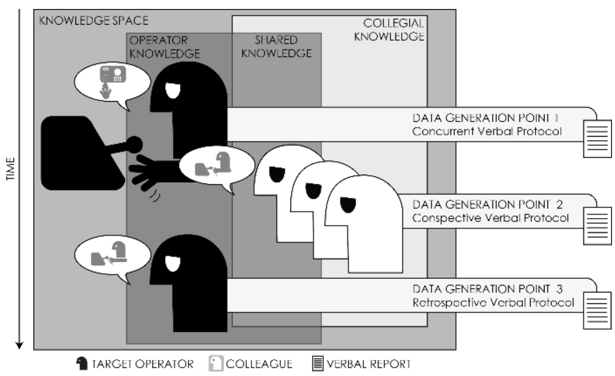


Figure 1. The Collegial Verbalisation method (repinted with permission from Jansson et al. 2015)

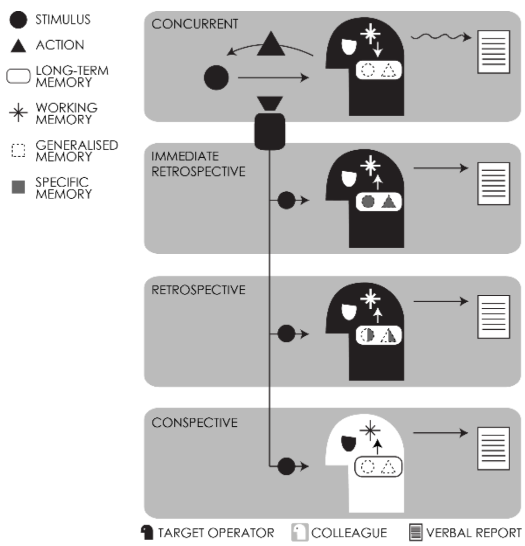


Figure 2. Verbal protocols in relation to time and familiarity (repinted with permission from Jansson et al. 2015)

they verbalise on the behaviour of their fellow operators. The critical part of the CV method is thus the conspective protocol, not its retrospective counterpart.

Collegial Verbalisation and Conspective Protocols

The CV model consists of three data generation points, one first point for data generation by a target operator (Data Generation Point 1), a second point for data generation by a colleague (Data Generation Point 2), and a third point for data generation retrospectively by the target operator (Data Generation Point 3). The three data generation points are referred to as Concurrent, Conspective, and Retrospective Verbalisation respectively (Figure 1). The first and second data generation points are independent of each other in the sense that at least two different narrators are involved. The same goes for the second and the third data generation points. The separation of data generation points is seen as the unique contribution of the CV method. However, a verbal protocol from any of the data generation points is not independent of the domain-specific task knowledge with which it is concerned. On the contrary, the content of the verbal reports is of central concern. Without reference to content, conspective protocols would be useless and the CV method meaningless. The consequence of this is that the method is restricted to research settings where the researcher has access to domain-specific knowledge in terms of expertise in the form of skill developed in relation to a specific task. Data generation is also, of course, limited by the number of operators or colleagues that can participate.

Even though the narrators participating are independent of each other at the data-generation points, they share experiences from the same environment, which means that their verbal reports will reflect these joint and common experiences. Here, it is interesting to note that Nisbett and Wilson (1977, p. 257) in their often cited review of verbalisation methods concluded that “[i]t is frightening to believe that one has no more certain knowledge of the working of one’s own mind than would an outsider with intimate knowledge of one’s history and of the stimuli present at the time the cognitive process occurred”. We argue that Nisbett and Wilson (1977), perhaps accidently, pointed to two important aspects with their remark: (1) in everyday situations, verbalisations often reflect the use of both working memory and long-term memory in conjunction; and (2) even though it can be hard to accept that an outsider who knows oneself well can predict one’s behaviour, this points to the possibility of having other people verbalising the actions of oneself. In environments where domain knowledge is shared between close colleagues, we might find that they also share cognitive strategies.

Figure 2 shows the relation between the three protocols when the retrospective protocol is divided into two phases, immediate retrospective and retrospective protocols, respectively. This organisation of the three protocols into four phases emphasise memory decay over time as the factor that ties the protocols together in a synthesised framework. As can be realised from the figure, a protocol based on immediate retrospective verbalisation is closer to protocols based on concurrent verbalisations, whilst protocols based on retrospective verbalisation distant in time from the target activity is closer to conspective protocols. The rationale for organising verbalisation protocols into these four phases is to show the importance and the role of long-term memory structures in domain-specific knowledge, and that knowledge elicitation using highly experienced decision-makers cannot ignore these long-term memory structures if the researchers’ goal is to account for the decision makers’ experiences when trying to understand their decision making behaviours.

EMPIRICAL FINDINGS

Below, the application of the CV method and the associated conspective protocol are described through three field studies, all of which had the same purpose: to understand in detail the behaviour and actions, including the mental behaviour, of the participating professional decision makers as a basis for suggestions for improved systems design. The procedures for implementing and using the conspective protocols and the CV method in the three different domains are briefly described. Details on the design of the individual studies can be found in Jansson, Olsson and Erlandsson (2006), Erlandsson and Jansson (2007, 2013), and the successive development of the method has been described in Jansson et al. (2015). The method was used differently in the three domains. Practical circumstances determined the number of participants available, both as target participants and as colleagues, as well as the design of the studies. The first two studies focused on the content of the verbal reports, in particular the conspective protocols in comparison with concurrent protocols, whereas the last study focused on the comparison between conspective and retrospective protocols from a methodological perspective. In all three studies, the colleagues were asked to describe what they believed the target operator in the recorded incident was paying attention to and taking meaning from, but not to imagine themselves being in the situation.

A Field-Study of Train Drivers

Method

Video recording sessions were conducted with six different professional train drivers while driving along four different types of real-schedule routes, such as long-distance routes, commuter traffic, and so forth. Three different video cameras were used to capture the driver, the instrumentation, and the signals along the tracks. They were asked to think aloud while they were driving. Seven other professional train drivers then individually performed conspective verbalisation while watching these video recordings. The recordings were muted so that they could not hear the target driver's comments. This conspective procedure was also recorded. The protocols allowed for comparisons in-between the colleagues as well as between colleagues and target drivers.

Table 1. Examples of mental actions in train driving identified with the help of 'conspective' protocols

Non-observable actions		
Leaving a station	**Out on the route**	**Approaching a station**
Judging time available and preparing to get away quickly to save time	Judging speed ahead in order to avoid second level warnings and/or automatic braking	Calculating braking power and braking distance to end up at the right place at the platform
Calculating power needed to leave station smoothly	Judging time available to manage to be in time	Preparing the entering of the station, attention directed towards platform and signals for switches

Results

Analyses of the conspective protocols showed that the drivers use information from the signal system and the instrumentation in the cab, to a large extent they also use information from the surroundings near the track. For example, the colleagues noted that the target drivers were checking for particular signs along the track, for reference points in the surroundings on when to apply the brakes, focusing attention on people on platforms, preparing for and expecting certain braking capacity and so forth. All these behaviours are difficult for a lay person to detect and understand. With conspective protocols it was possible to understand that these non-observable behaviours are important in the train-drivers' organisation of the decision making activities. Furthermore, the analysis of the train-drivers' behaviours showed that the task of driving a train can be divided into three phases. Out on the route, between two stations, the drivers focus their attention on the speed-limit so that they do not exceed the critical limits where the train brakes automatically. They also adjust the speed of the train, constantly weighting goals such as efficiency, safety, and comfort against each other, that is, the drivers use the difference between actual speed limit and braking speed limit to manage to keep up with the time table. When approaching a station, their focus shifts towards the surrounding environment and the braking conditions of the train at this particular station, for example they monitor the slope of the track, weather conditions and other aspects. Furthermore, they also prepare for things they cannot control themselves, such as people on the platform, trains coming the other way, or expected clear-signals through the switches. When leaving the station, the drivers focus on the possibilities to leave as quickly as possible since this is the part of the journey that is most time critical from a time-table perspective. If they, for example, lose time here it is often difficult to catch up later on, but if they get away quickly, they can have a smoother ride further down the trackway. Other important things noted during the conspective verbalisations were the calculation of how much power is needed to get away smoothly and being extra cautious with respect to passengers arriving late. The colleagues also noted that domain-specific knowledge (route-knowledge) is essential if one wants to reach the goals of driving smoothly and at the same time keep ahead with the time-table. Table 1 shows the mental behaviours identified with the help of conspective protocols (Jansson et al., 2006).

A Field Study of High-Speed Ferry Operation

Method

Four different video cameras were used to capture the crew, instrumentation, and the surroundings. Two officers, one captain, and one navigator participated as target officers on the bridge during this recording. Four colleague officers individually watched and verbalised on the actions and decisions made by the target officers in the video. This conspective procedure was also recorded, and the protocols from these sessions were then compared to examine to what extent the four colleagues agreed on observed behaviours. The protocols from the colleagues allowed for comparisons in-between the colleagues only, not between colleagues and target officers since concurrent verbalisation was not utilised due to risks of interference with procedures on the bridge.

Results

A detailed examination of the protocols revealed that there was a high degree of agreement between the colleagues on the main series of events. Many of the comments concerned non-observable actions and behaviours impossible for a lay person to understand completely or correctly. A comparison between the colleagues' conspective protocols showed that there is conformity among the officers in many situations. Some specific statements conflicted however between the protocols, indicating the possibility of maladaptive mental models within at least one of the colleagues since both colleagues' conceptions cannot be reconciled with reality at the same time. From the following statements, it is clear that the verbalising colleagues think and reason differently:

- [The action of verbally] handing over [between the bridge wing and the centre control] is very important. Everybody knows the procedure, but as long as I haven't said anything, I'm still responsible.
- Now, the control is transferred back [to the centre control]. If anything would fail, [the officer] would bring it up, but otherwise there is no need for any verbal hand over [procedure].

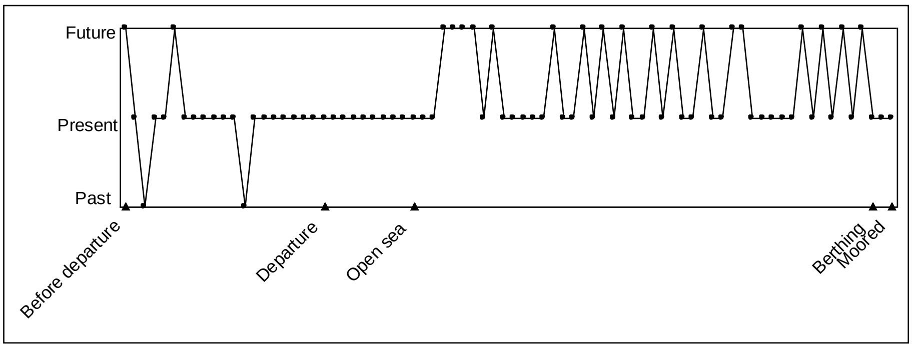


Figure 3. Past/present/future categories in a sequential order identified with conspective protocols. Each dot represents a verbal statement.

Moreover, the conspective protocols made it possible to categorise the sequential order of past/present/future events continuously discussed by officers on the bridge (Figure 3) revealing the importance of proactive decisions and historical events for present actions (Erlandsson & Jansson, 2007).

A Quasi-Experimental Field Study of Train Dispatchers

Method

A systematic comparison between conspective and retrospective verbalisation was made. Four train dispatchers were video recorded individually while working. These dispatchers then performed both a conspective and a retrospective verbalisation from these recordings in a quasi-experimental setup, that is, they verbalised on both their own actions and the actions of a colleague. It made it possible to compare conspective and retrospective protocols for the same events. In order to minimise the effect of remembering situation-specific details, there was a delay of a few weeks between the target situation and the verbalisations. By this procedure, the emphasis was on the comparison between long-term memory structures, assuming effects of recency to be under control.

Results

A comparison between the conspective and retrospective protocols was carried out on three different levels: (1) Protocol level – this is quantitative measure, showing the number of characters in the verbal protocols, measuring the amount of verbalisation activity going on during that sequence of verbalisation; (2) Statement level – this measure consists in number of assertions and utterances, measuring the amount of sentences and statements going on during the session; and (3) Topic level – this is a qualitative measure consisting in the amount of topics the participating dispatchers deal with during the session. As expected, there was a high degree of agreement on the protocol level between retrospective and conspective protocols, which means that the amount of text generated in the protocols varied as a consequence of the current activity. Also as expected, there was a very low degree of agreement on the statement level in that when the comparison is based on similarity in utterances and sentences, the conspective and retrospective protocols are very different. Finally, and most importantly, there was a high degree of agreement on the topic level between the retrospective and conspective protocols, that is, when statements were categorised into topics, the protocols covariate with the topics to a large degree. A Krippendorff's alpha reliability estimate showed a reliability coefficient of .8847, with a 95% confidence interval of .7912 – .9627 (Erlandsson & Jansson, 2013, p. 247).

Thus, the participating dispatchers seem to verbalise the same content to a large degree. This does not however necessarily mean that they interpret all specific actions in the same way. On the contrary, on some topics they have different explanations of whether the actions exhibited by the target operator are relevant behaviours in the particular situation or not. Table 2 shows an example where they agree on the content but disagree on the relevance of the actions. This information may be as important as any information showing the similarity between colleagues and target operators.

Table 2. Examples of statements where the colleague questions the actions of the practitioner (Erlandsson & Jansson, 2013)

Example	Retrospective	Colleague
#1	‘‘Now I'm replanning some trains here. 537 are 7 min late’’	‘‘2135 were late, and 537. But now he is moving the wrong line for 537, but, yes also for 2135 perhaps. He is moving the departure, but they are also late on arrival to Katrineholm’’
#2	‘‘Yes, it says here, detected but not locked path for 2148. Since 2148 has a departure here, he cannot complete that, until there is a path through Åby’’	‘‘But he believes that he clicks on arrival, so now he does not understand why it says that 2148 are there. … As he clicks on departure, and that is a bit tricky, but he has to check this box in order to see the arrival condition’’
#3	‘‘I'm trying to move that rotation point down. I realise that they will be late’’	‘‘I'm not really getting why he moves this rotation point down here? That gave a speed of 50 km per h, having a 50-train from here to here. That is a bit unrealistic, unless he have gotten some information that they are driving without ATC, so that it becomes a 70-train’’

DISCUSSION

Conspective protocols serve as a valuable complement to other information acquisition protocols in four different ways. First, it gives a lot more pieces of information than concurrent protocols do. Second, it specifically allows researchers to scrutinise the hypotheses on mental behaviours in naturalistic decision making tasks. Third, independent observers' verbalisations open up the possibility of having more participants verbalising on the same set of data. Finally, the introduction of independent observers also makes it possible to detect differences in understanding between target participants and colleagues.

It is only with the introduction of conspective protocols and the use of the synthesised CV method that it is possible to discriminate between different forms of understandings, something that can be critical in many domains, and an important input into future studies in the field of human factors. Since the colleagues in the high-speed ferry study were not present at the bridge when the target officers were running the craft, they, of course, cannot remember details from the particular situations. With the conspective verbalisation procedure, they rather recall similar situations in which they have been involved since the environmental constraints imposed on them in similar situations consist in regularities that are abiding. Erlandsson and Jansson (2007) concluded that the most controversial issue with the CV method and the conspective protocols is the idea of having other subjects than the target operators performing the verbalisations. With this approach, the colleagues have not been part of the target actions, and are therefore left with some form of recall when they verbalise. It is important to bear in mind, however, that the operators participating are highly familiar with the tasks, and that they all have long experiences with the same tasks and systems. Conspective verbalisation means a shift away from analysing working memory structures to analysing long-term memories. This also means different theoretical assumptions compared to more traditional forms of verbalisation tasks. Assuming that common experiences result in similar strategies for dealing with environmental constraints and regularities, multiple verbalisers constitute a possibility for field studies in the area naturalistic decision making. When data elicited through this protocol are in congruence with other protocols in the synthesised methodological framework, it is a valuable complement.

Knowledge elicitation in naturalistic decision making (Klein, Calderwood, & Clinton-Cirocco, 2010) as well as in dynamic decision making (Brehmer, 1992) demands new forms of verbalisation methods than existing models and methods (e.g., Ericsson, 2006). Concurrent verbalisation and think-aloud protocols may work well when it comes to chess players and math problems, or decision problems that are static or stationary, but expert performance in decision domains characterised by complexity, ill-structured problems, non-transparency and dynamics is based on recognition-primed decisions (Klein, 1992) and strategies for dealing with dynamics. In such study contexts, conspective protocols and the CV method have a role to play, we believe.

CONCLUSION

A model for verbalisation by colleagues is presented as the rationale for the Collegial Verbalisation (CV) method. It is based on the idea that in situations where domain knowledge is shared between colleagues one might find that they also share cognitive strategies that they can verbalise. Independent observers (colleagues) comment in the form of conspective verbal reports on the behaviour of a target operator. It solves some of the challenges associated with established verbalisation protocols like concurrent and retrospective verbalisation. The method is however sensitive to how close to the practitioner's experience the narrator is. Data generated by the CV method can, for example, be useful for practical purposes since correlations between colleagues' statements can be developed into team-learning and discovery of differences between team-members understanding of situations and contexts. It makes the method particularly interesting for research in naturalistic decision making. The most controversial issue, and at the same time the unique contribution, is the fact that it is not the practitioners themselves that provide the verbalisations. The narrator is left with doing some form of interpretation of the practitioner's actions based on their knowledge and experience. It is concluded that CV and conspective protocols are separate from existing verbalisation methods but that it is intended to be used in conjunction with these, not in isolation. The major implication is the contribution of an independent source of data to be used in applied research.

REFERENCES

Bainbridge, L. (1979/1999). Verbal reports as evidence of the process operator's knowledge. *International Journal of Human-Computer Studies*, 51, p. 213-238.

Bartl, C. & Dörner, D. (1998). Speechless when thinking: About the role of speech for problem solving. *Sprache und Kognition*, 17, p. 224-238.

Brehmer, B. (1990). Strategies in real-time, dynamic decision making. In R. Hogarth (Ed.), *Insights in Decision Making* (pp. 262-279). Chicago: University of Chicago Press.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211-241.

Calderhead, J. (1981). Stimulated recall: A method for research on teaching. *British Journal of Educational Psychology*, 51, (2), p. 211-217.

Dickson, J., McLennan, J., & Omodei, M. M. (2000). Effects of concurrent verbalisation on a time-critical, dynamic decision making task. *Journal of General Psychology, 127*, p. 217-228.

Dominguez, C. O., Flach, J. M., McDermott, P. L., McKellar, D. M., & Dunn, M. (2004). The Conversion Decision in Laparoscopic Surgery: Knowing Your Limits and Limiting Your Risks. In K. Smith, J. Shanteau, & P. Johson (Eds.), *Psychological Investigations of Competence in Decision Making* (pp. 7-39). New York: Cambridge University Press.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 223-242). New York: Cambridge University Press.

Ericsson, K. A. & Crutcher, R. J. (1991). Introspection and verbal reports on cognitive processes – two approaches to the study of thinking: A response to Howe. *New Ideas in Psychology, 9, (1)*, 57-71.

Ericsson, K. A. & Simon. H. A. (1980). *Verbal reports as data*. *Psychological Review, 87*, p. 215-251.

Ericsson, K. A. & Simon. H. A. (1984). Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press.

Erlandsson, M. & Jansson, A. (2007). *Collegial verbalisation – a case study on a new method on information acquisition. Behaviour & Information Technology, 26(6)*, p. 535-543.

Erlandsson, M. & Jansson, A. (2013). Verbal reports and domain-specific knowledge: A comparison between collegial and retrospective verbalisation. *International Journal of Cognition, Technology and Work*, 15, (3), p. 239-254.

Gibbons, F. X. (1983). Self-attention and self-report: the "veridicality" hypothesis. *Journal of Personality Assessment, 51*, p. 517-554.

Hassall, M. & Sanderson, P. (2014). A formative approach to the strategies analysis phase of cognitive work analysis. *Theoretical Issues in Ergonomics Science*, 15, (3), 215-261.

Jansson, A. (1995). Strategies in dynamic decision making: Does teaching heuristic strategies by instructions affect performance? In J.P. Caverni, M. Bar-Hillel, F.H. Barron & H. Jungermann (Eds.), *Contributions to Decision Making*, pp. 213-232. Amesterdam: Elsevier Science Publishers.

Jansson, A., Erlandsson, M., & Axelsson, A. (2015). Collegial verbalization – the value of an independent observer: An ecological approach. *Theoretical Issues in Ergonomics Science, 16, (5)*, 474-494.

Jansson, A., Olsson, E., & Erlandsson, M. (2006). Bridging the gap between analysis and design: Improving existing driver interfaces with tools from the framework of cognitive work analysis. *International Journal of Cognition,Technology and Work, 8, (1)*, p. 41-49.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision Making in Action: Models, Methods*, pp. 138-147. Norwood, NJ: Ablex Publishing.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire-ground. The original study and plus a postscript. *Journal of Cognitive Engineering and Decision Making, 4, (3)*, 186-209.

McIlroy, R. C. & Stanton N. A. (2011). Observing the observer: Non-intrusive verbalisations using the concurrent observer narrative technique. *International Journal of Cognition, Technology and Work, 13, (2)*, p. 135-149.

Miller, J. E., Patterson, E. S., & Woods, D. D. (2006). Elicitation by critiquing as a cognitive task analysis methodology. *International Journal of Cognition, Technology and Work, 8, (2)*, p. 90-102.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, p. 231-259.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The Adaptive Decision Maker. Cambridge University Press.

Polanyi, M. (1967). The Tacit Dimension. New York: Anchor Books.

Rasmussen, J., Pejtersen, A. M. & Goodstein, L. P. (1994). Cognitive Systems Engineering. New York: Wiley.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). The think-aloud method: A practical guide to modeling cognitive process. London: Academic Press.

Vicente, K. J. (1999). Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, G. & Ayton, P. (1987). Eliciting and modelling expert knowledge. *Decision Support Systems, 3*, p. 13-26.

Deciding to design better user interfaces

Daniel P. JENKINS[a], Andrew WOLFENDEN[b], David J. GILMORE[b] and Malcolm BOYD[a]
[a]*DCA Design International Ltd*
[b]*Elekta*

ABSTRACT
This paper describes an approach developed to establish information requirements for human machine interfaces within complex systems. The approach, rooted in decision making, is not limited to the design of digital interfaces; instead, it encourages the consideration of information that is distributed across the physical, digital and social environment. The technique has been successfully applied to the design of radiotherapy equipment, which is used here as a case study. The process starts with a semi-structured interview, around Rasmussen's decision ladder, designed to elicit the information that could be used at each stage of the treatment process (rather than is used, or should be used). The identified information elements are then coded to indicate when the information may be needed, where the information is required, and who may need it. The resultant model has been designed to create an explicit link between analysis and the design of physical and digital artefacts.

KEYWORDS
Practical application; decision making; interface design; medical; radiotherapy

INTRODUCTION

The link between the quality of a user interface and system performance is now almost universally accepted. For very simple interactions, such as an alarm clock app for a mobile phone, developing an interface may be an intuitive and straight forward process. The adoption of a style guide and consideration of a set of heuristics (e.g. Nielsen & Molich, 1990) may be enough to ensure a useable design. However, the challenge is proportional to the complexity of the product or service being designed. The consequence of system failure is also an important consideration in the approach adopted, while the failure of an alarm clock may result in missed appointments or even flights, it is unlikely to cause a fatality. Conversely, in safety critical environments, such as radiotherapy treatment, the cost of failure may be much higher.

Most interface designs start by establishing the information requirements. More often than not, these information requirements are communicated as a text-based document. The resultant document typically forms the bridge between systems architects, or engineers, and the interface designers. Perhaps, unsurprisingly, the quality of these information requirements has a direct relationship with and impact on, the quality of the resultant interface and the performance of the systems in which they are used. Thus, in order to ensure the safety, efficacy, efficiency, usability, and resilience of products and services, it is important that the process for developing information requirements is fit for purpose.

Thus, ostensibly at least, the foundation for a well designed interface design lies in establishing: (1) What information is required? (2) When it needs to be displayed? (3) Where it should be displayed? (4) Whom it should be displayed to? And (5) How – in what format it should be delivered?
This paper aims to illustrate how a highly structured approach to analysis, based on the consideration of decision making, can inform the design of commercial products.

Decision making as a starting point

Decision making is at the heart of all control tasks. There have been many attempts to model decision making activity. Most rational models involve some form of observation of information, orientation to the current situation, a choice as to which action to adopt, and finally an action. The decision-ladder (Rasmussen, 1974; see Figure 1) is the tool most commonly used within Cognitive Work Analysis to describe decision-making activity. Unlike some of its counterparts, its focus is on the entire decision-making activity, rather than the moment of selection between options.

It has been adopted here as it is capable of representing more linear 'rational' or knowledge-based decision making activity, as well as more naturalistic rule and skill-based decision making activity. For more rational, or knowledge-based, decision making, where decision makers are responding to unfamiliar situations, a more linear path through the decision-ladder is expected. Following activation, users are expected to observe available information, determine a system state, consider options, and relate this to a chosen goal. The context specific interpretation of the goal is then used to determine a target state, task and procedure. Thus, the left side of the decision-ladder represents the observation of the current system state, whereas, the right side of the decision-ladder represents the planning and execution of tasks and procedures to achieve a target system state.

The key utility of the decision making model; however, is its ability to also represent more naturalistic decision making activity. 'Recognition-primed decision making' can be represented with 'shortcut' links between nodes (most notably between the two legs; see Figure 1). Another key distinction from other decision making models is that the decision ladder does not have to be limited to a single decision making entity. The decision ladder model can also be used to represent collaborative decision making, distributed between a range of human and technical decision-makers. The concept of the approach described in this paper is to create a structure for capturing the information requirements that could be required and to code them based on when, where they could be needed as well as to whom and in what format.

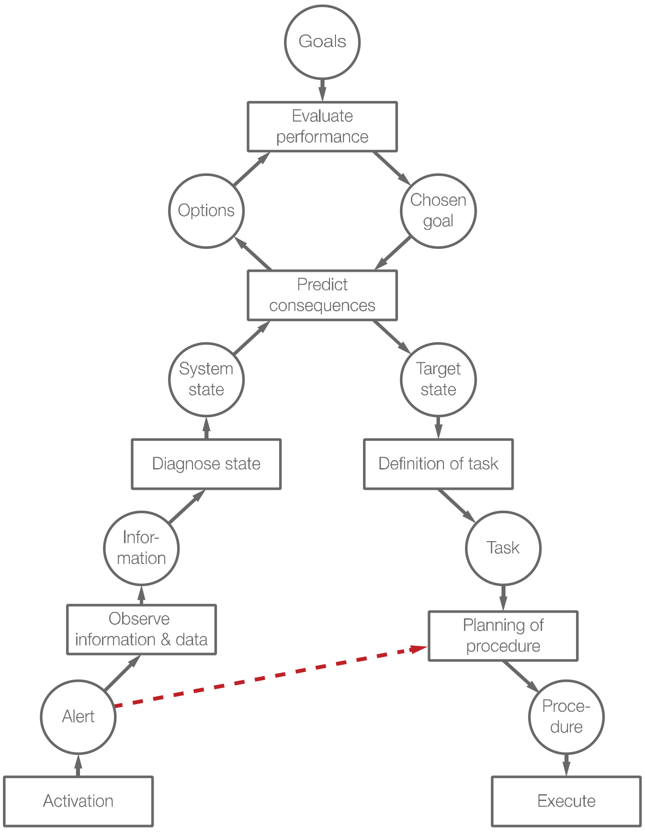


Figure 7. Decision ladder template

Radiotherapy as a case study

External radiotherapy involves targeting specific parts of the body with radiation delivered by high energy x-rays. External radiotherapy is most commonly used to treat cancerous tumours and is typically delivered by a large medical device called a linear accelerator (LINAC). LINACs have been optimised to focus the radiation on the identified volume (normally containing the tumour), while minimising the exposure to surrounding health tissue. LINACs use microwave technology to generate a beam of radiation which is shaped to fit the patient's tumour. This beam is then rotated around the patient allowing the radiation to be delivered from different angles (reducing the damage to surrounding tissue). Each patient has a unique treatment plan that involves specific beam intensities and shapes and, as well as angles of delivery. Due to the size and complexity of the LINAC, the machine tends to be in a fixed location. The patient is aligned to the machine on a movable table. Traditionally, this alignment is done to reference marks drawn onto their body (referred to as tattoos), more recently LINACs are equipped with imaging technologies that allow the position of the tumour (or given volume) to be verified.

DCA supported Elekta in designing and developing a next generation suite of radiotherapy equipment at an early stage of the design process. While conceptual, these 'visions for the future' were grounded through collaborative technical review and based on an extensive body of evidence collected from visits to seven treatment sites worldwide (including sites in Belgium, Brazil, Canada, USA; see Figure 2), over 90 hours of observations (approximately 360 treatment sessions), and over 50 in-depth interviews with health care professionals, thought-leaders, and system stakeholders.

A detailed audit and understanding of the information required to support treatment delivery was fundamental to these concepts. The overarching philosophy was to provide the right information, at the right time, in the right place, to the right people in the right format. This involved splitting information into three categories.

1. Typically required (at given time, location, for given user)
2. Could be required (at given time, location, for given user)
3. Not required (at given time, location, for given user)

Figure 2. Observation of radiotherapy treatment

APPROACH

Interface designers require an understanding of what information is needed in different contexts. . Where the requirements for information are difficult to predict one simplistic approach is to provide all of the information all of the time, or allow users to navigate to the information they require as they need it, which may indeed be the best option for design. However, where the information requirements change predictably based on situation it may be advantageous to change the way information is presented.

The approach involves establishing a long list of all of the information requirements that could be needed by the system, which is coded to provide additional detail and constraints, such as when, where, to whom and how information should be displayed. The approach for eliciting the systemic information requirements is based on a series of semi-structured interviews with system experts and/or stakeholders. These interviews are constructed around a template with a decision ladder at its centre for each key situation. The process involves capturing the questions that decision makers pose themselves and the system at each stage of the decision making process. An example of one of these decision ladder models is shown in Figure 3.

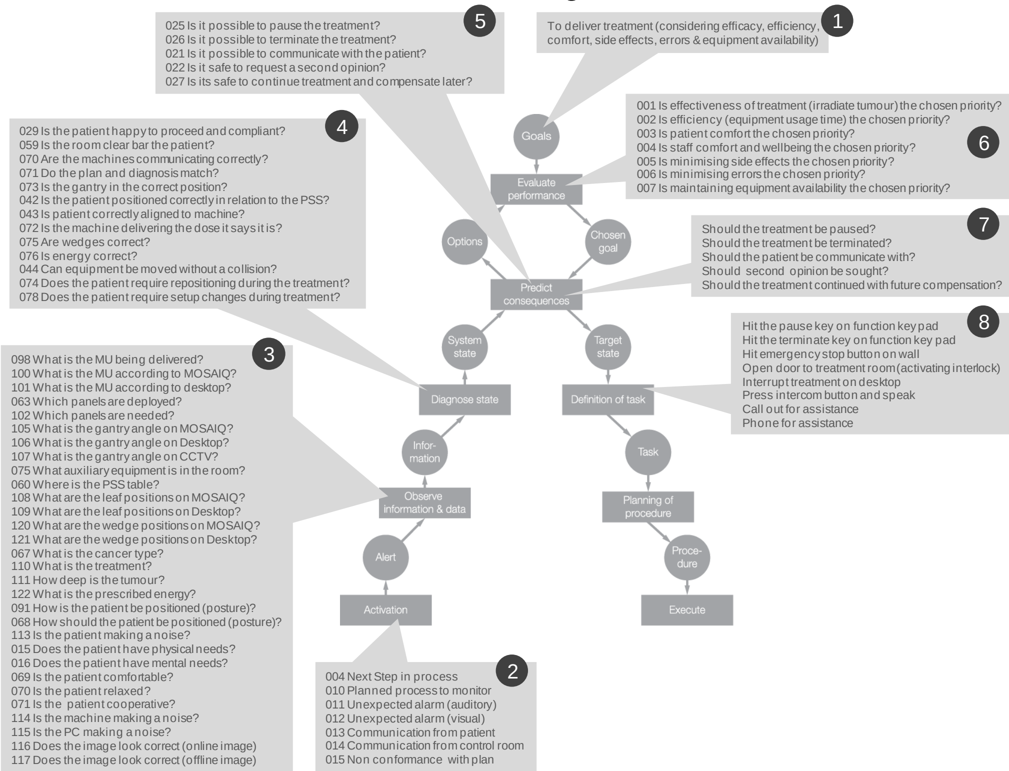


Figure 3. Model for ‘beam on’ (large numbers relate to task phase in Table 2)

Dividing up the activity

A separate decision ladder model should be created for each of the key situations. These situations are typically identified through a contextual activity template (another tool from CWA) or a hierarchical task analysis (HTA) model. For the case of radiotherapy, the treatment process can be broken down into ten discrete phases (see Table 1) each requiring different decisions and thus different information elements. Task phase 8, (Beam on; illustrated in Figure 3), will be explored in greater detail in this paper to illustrate the process. The same process was applied to each of the ten phases.

Table 1. Task phases

Task phases	Description
1 Patient registration	Identify the patient and relate them to the treatment database
2 Manage patient	Explain the treatment process
3 Machine preparation	Set up the machine to receive the patient, add set up aids
4 Patient loading	Sit the patient on PSS and lay them down
5 Patient set-up	Configure fixation / immobilisation devices, position the patient
6 Image	Image the patient (if required)
7 Prepare for beam delivery	Adjust the position of the patient, retract panels (if required)
8 Beam on	Treat patient
9 Unload patient	Remove fixation / immobilisation devices, help patient up
10 Clean up	Wipe down machine, reset ready for next patient

A semi structured interview is the basis of building each of the models. The broad structure for the semi-structured interviews with radiotherapists is presented in Table 2. The output of the interview is presented in Figure 3.

Table 2. Interview process

Step 1 – Determining the goal The first stage of the interview process for each model is to structure the goal of the system. The expert should be asked to provide a high order goal, along with a number of constraints affecting it. The expert should be reassured that the constraints could possibly be in conflict. The information works well placed in the format "To (insert goal) considering (insert constraints)". For the case study, the goal at this phase of the process is simply to 'deliver the treatment', the caveat is that it must also consider the system values of efficacy, efficiency, comfort, side effects, error and equipment availability.
Step 2 – Alert The expert should be asked to begin the walk-through at the chronological start of the process. Alerts capture the events that first draw them to the need to make a decision. During the delivery process alerts to a system state change include monitoring the process, alarms (visual and auditory), and communications from the patient and communications from other members of staff in the control room.
Step 3 – Information The expert is asked to list the information elements they would use to gain an understanding of the situation. The information elements are the nuggets of information that can be brought together to understand the state of the system. In this case, they include information about the physical equipment (e.g. the gantry angle, the equipment in the room, the position of the table) along with information from the HMI (e.g. the dose being delivered, the beam shape), information from patient records (e.g. treatment type and location), information from the patient (e.g. are they comfortable, relaxed).
Step 4 – System state The system states represent a perceived understanding of the work system based upon the interpretation of a number of information elements. The key distinction between an information element and a system state is that system states are formed of more than one quantifiably different element of information. In short, information elements are processed and fused to form system states. Questions such as 'is the patient positioned correctly?' can be assessed by considering the treatment type and the current position of the patient.
Step 5 – Options The options within the ladder can be described as the opportunities for changing the system state in an attempt to satisfy the overall goal. The options are structured as questions in the form; "is it possible to (…)?" The number and type of options available will be informed by the system state. It is anticipated that in certain situations there may only be one option available. At a high level, during treatment there are five main options available to the operator: To pause the treatment, terminate the treatment, communicate with the patient, request a second opinion, or continue treatment and compensate later.
Step 6 – Chosen goal The chosen goal, at any one time, is determined by selecting which of the constraints receives the highest priority. This does not have to be an absolute choice per se, rather, one takes a higher priority than the other does in the given situation. The system values of efficacy, efficiency, comfort, side effects, and error and equipment availability are likely to be in conflict.
Step 7 – Target state The target states mirror the option available; once a particular option is selected, it becomes the target state. The options are rephrased in the form "Should (option) take place?"
Step 8 – Task The listed tasks relate to the tasks required for achieving the target state while maintaining the overall goal (e.g. hit the pause button, press intercom button and speak).
Step 9 – Procedure The procedure lists questions that will inform the choice of task procedure.
Step 10 – Validate model Once a first pass of the decision ladder has been completed, the expert is then asked to repeat the process adding additional or alternate elements. The interviewer can support this by posing the question what if (information element) were unavailable, are there any other cues that you could use? Another useful technique is to cross-check the information elements against the system states to see if system states can be informed by new information elements, or if information elements could inform new system states.

ANALYSIS

What information could be needed?

The analysis approach starts with a combined list of the information requirements that could be needed by the system. This list is aggregated from each of the decision ladder models (in this case the ten task phases). These are listed in a spreadsheet allowing coding to provide additional detail and constraints, such as when, where, to whom and how information should be displayed. At this stage, it is useful to give each element in the model a unique identifier. For example Alerts (AL001), Information elements (IE001), System states (SS001), Options (OP001), Goals (GL001), Target states (TS001), Tasks (TA001) and Procedures (PR001). Where appropriate, similar elements can often be combined and reworded to reduce the total number of elements.

When could the information be required?

The level of consistency between the decision ladder models can be a very useful cue to design. Information elements can often be divided into two groups (1) persistent and (2) context specific. As the name suggests, persistent information elements should be presented regardless of the situation or task, while context-specific information elements should only be displayed during the tasks or situations where they are relevant. For interfaces that predominately contain persistent data, an argument could be made for showing all information elements as this reduces the complexity of a moded display. A matrix can be created listing each of the alerts, information element, system states, option, goals, target states, tasks and procedures. This can be coded to show which elements are present in which situations or tasks. An example of this is provided in Table 3. The matrix can be coded to show when the elements are typically needed (dark grey cell) and when they may be needed (light grey cell). As illustrated in Table 3, some of the information elements may be required nearly all of the time, such as the name of the patient, whereas other elements are only required in specific situations (e.g. the dose being delivered during phase 8).

Table 3. Example elements coded by task phases (dark grey cells indicate typically needed, light grey cells indicate may be needed) – this is a cut down list of elements for illustrative purposes

ID	Description	1. Patient registration	2. Machine prep	3. Machine preparation	4. Patient loading	5. Patient setup	6. Image	7. Prepare for beam delivery	8. Beam on (Deliver radiation)	9. Unload patient	10. Clean up
AL005	Patient appears agitated										
AL011	Unexpected alarm (auditory)										
AL013	Communication from patient										
IE001	What is the name of the patient										
IE008	What is the weight of the patient										
IE015	Does the patient have physical needs										
IE098	What is the MU being delivered										
IE067	What is the cancer type										
IE025	Does the patient have multiple appointments										

Where could the information be required?

The 'where' question can be addressed in two ways, firstly a decision needs to be made on where information should be displayed in the environment. This may mean different sites (e.g. maybe in different countries), different rooms within a site (e.g. control room, plant room, treatment room), or different locations within a room (e.g. wall mounted display, equipment display, indicator lamp, hard-copy manual, whiteboard, poster).

The second way of addressing the question is to consider the arrangement within each of these locations (e.g. the location on the poster, or the screen). There are a number of applicable approaches for grouping information elements. The output of the analysis approach provides a useful means of structuring the interface. By explicitly mapping which information elements relate to which systems states. By adding a column to the matrix for each location the information elements can be coded to indicate the relationship. An example of this is presented in Table 4.

Table 4. Example elements coded by location (dark grey cells indicate typically needed, light grey cells indicate may be needed)

ID	Description	Environment			Feedback		Documentation				Comms	
		Reception area	Treatment room	Patient appearance	Equipment feedback	Imaging outputs	Patient documentation	Treatment description	Workflow description	Schedule description	Staff communication	Patient communication
AL005	Patient appears agitated											
AL011	Unexpected alarm (auditory)											
AL013	Communication from patient											
IE001	What is the name of the patient											
IE008	What is the weight of the patient											
IE015	Does the patient have physical needs											
IE098	What is the MU being delivered											
IE067	What is the cancer type											
IE025	Does the patient have multiple appointments											

To whom should the information be displayed?

In a similar way, different actors in the system may need different access to information. Actors may include: Digital agents, Operators, Supervisors, Administrators, Maintenance staff, etc. The matrix of information elements and system states can also be coded to indicate which actors the information should be displayed to. This can help to inform and document decisions relating to whether separate system views are required and whether actor types can be combined to reduce the number of views required. An example of the coding for the radiography system is presented in Table 5.

Table 5. Example elements coded by role (dark grey cells indicate typically needed, light grey cells indicate may be needed)

ID	Description	Greet and manage patients (typically	Treatment delivery staff (radio-therapists /	Oncologists	Treatment planning (Physicist / Dosimetrist)	Equipment maintenance	Porters
AL005	Patient appears agitated						
AL011	Unexpected alarm (auditory)						
AL013	Communication from patient						
IE001	What is the name of the patient						
IE008	What is the weight of the patient						
IE015	Does the patient have physical needs						
IE098	What is the MU being delivered						
IE067	What is the cancer type						
IE025	Does the patient have multiple appointments						

How

The decision as to how information should be displayed will be informed by a consideration of the factors above. Once the information elements for each situation, location, and actor have been defined, the decision on representation needs to also consider the appropriateness of the representation.

Table 6. Example elements indicating potential formats

ID	Description	Format
AL005	Patient appears agitated	High quality image and videos of patient
AL011	Unexpected alarm (auditory)	Unique sounding alarm louder than background
AL013	Communication from patient	High quality audio
IE001	What is the name of the patient	Text
IE008	What is the weight of the patient	Numerical with units
IE015	Does the patient have physical needs	Numerical with units
IE098	What is the MU being delivered	Text field
IE067	What is the cancer type	Numerical with units
IE025	Does the patient have multiple appointments	Text field / map of body
AL005	Patient appears agitated	Schedule
AL011	Unexpected alarm (auditory)	High quality image of patient

TRANSITION TO DESIGN

As stated in the introduction, the objective of this approach is to create a far more explicit link between analysis and design. The first stage of the design process is to convert the information, represented in the matrix, into a graphical representation that can be passed to interface designers. Figure 4 shows an example of the information requirements for the control room displays during the ‘manage patient’ phase. A full list of information elements is presented and these are coded to show which are needed, which may be needed and which are not needed. It is proposed that these graphics provide a much more usable representation than the solely textual descriptions that they are intended to replace. It is also proposed that they replace some of the subjective ‘black art’ of the interpretation of these documents.

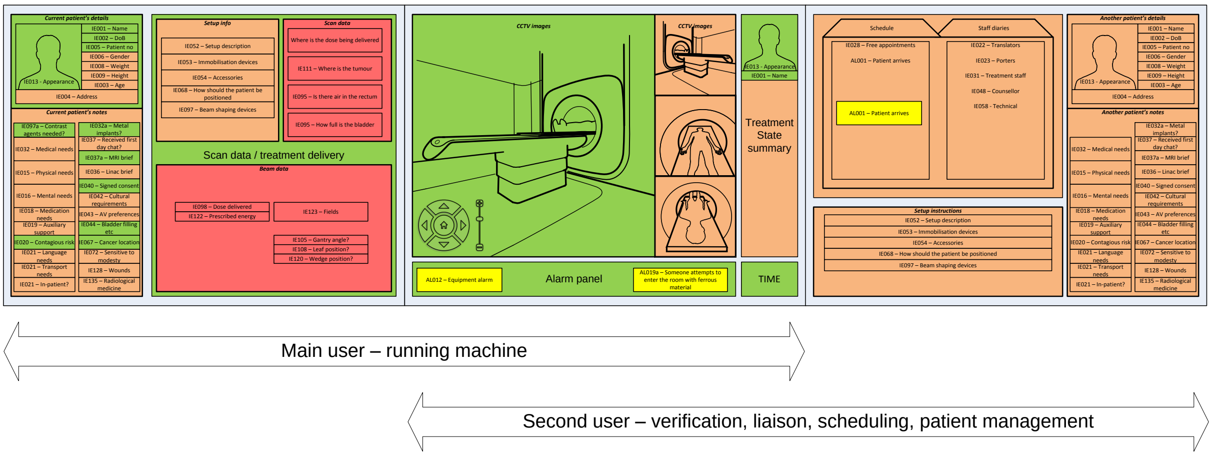


Figure 4. Example of control room display information requirements for the manage patients phase (Green – Typically required at the current phase; Amber – Could be required at the current phase; Red – Not required at then current phase; Yellow – Alerts to be displayed as required).

Similarly, Figure 5 shows a representation for the information that may be required within the treatment room during the patient loading phase of the treatment process. Here a distinction is made between which information elements are presented on/by the patient themselves, those in the general physical environment, those on some form of graphical display and those that are auditory alerts.

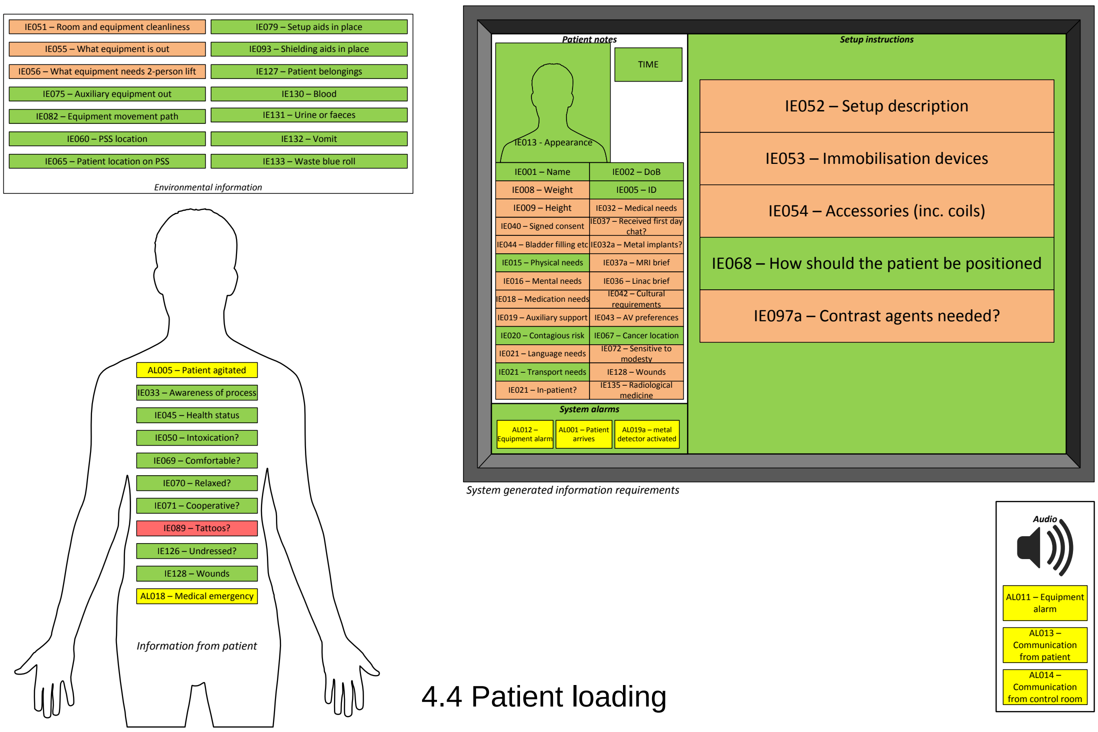


Figure 5. Example of treatment room information requirements for patient loading phase (Green – Typically required at the current phase; Amber – Could be required at the current phase; Red – Not required at then current phase; Yellow – Alerts to be displayed as required).

DISCUSSION AND CONCLUSIONS

The case study described in this paper formed the first step of the information requirements capture and of the interface design process. This activity described in this paper was completed in 2012. Since then, the development of the product has continued and is approaching clinical trials. The approach described here for identifying information requirements was repeated at a later date to validate the findings. A variety of other user interface design approaches have also been adopted to fit the fidelity of the concepts as they move through paper prototypes, to digital wireframes and eventually to fully coded interfaces.

For complex systems, it is highly advantageous to consider system information requirements at an early stage of the project. A structured approach is needed to ensure, firstly that all the required information elements are considered, and secondly that they are included in the optimal way to ensure an appropriate balance of system values (e.g. safety, efficacy, efficiency, usability and resilience). Most critically the output of this analysis needs to be presented in a format that can be shared and be well understood across the design team.

The approach described in this paper has proved to be effective in a wide range of situations. It has been applied to the design of unmanned aerial vehicle (Elix & Naikar, 2008; Jenkins, 2012), a military command and control system (Jenkins et al, 2010), a policing command and control system (Jenkins et al 2011a), a tank training simulator (Jenkins et al, 2011b), an automotive interface (McIlroy &Stanton, 2015), and a number of medical devices.

It provides welcome structure to the process of eliciting and structuring information requirements that focus on end users and stakeholders. One of the clear strengths of the approach is that it provides a very explicit link between the data collection, the analysis and the design of early interface concepts.

Due to the focus on user information requirements, it is contended that it results in more useable interfaces that will have a positive impact on multiple system performance metrics (e.g. efficacy, efficiency resilience).

REFERENCES

Elix, B., & Naikar, N. (2008). Designing safe and effective future systems: A new approach for modelling decisions in future systems with Cognitive Work Analysis. Proceedings of the 8th International Symposium of the Australian Aviation Psychology Association.

Jenkins, D. P. (2012). Using Cognitive Work Analysis to describe the role of UAVs in Military Operations. Theoretical Issues in Ergonomics Science. 13(3)335-357

Jenkins, D. P., Salmon, P. M. Stanton, N. A. Walker, G. H. and Rafferty. L. (2011a). “What Could They Have Been Thinking? How Sociotechnical System Design Influences Cognition: A Case Study of the Stockwell Shooting.” Ergonomics 54 (2): 103–119.

Jenkins, D. P., Stanton, N. A., Salmon, P. M. & Walker, G. (2011b). A formative approach to developing synthetic environment fidelity requirements for decision-making training. Applied Ergonomics. 42 (5) 757-769

Jenkins, D. P., Stanton, N. A. Salmon, P. M. Walker, G. H. and Rafferty. L. (2010). “Using the Decision Ladder to Add a Formative Element to naturalistic Decision-Making Research.” International Journal of Human-Computer Interaction 26 (2–3): 132–146.

McIlroy, R.C., Stanton, N.A. (2015) A decision ladder analysis of eco-driving: the first step towards fuel-efficient driving behaviour, Ergonomics,

Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces, Proc. ACM CHI'90 Conf. (Seattle, WA, 1-5 April), 249-256

The complexity of *skillscapes*: Skill sets, synergies, and meta-regulation in joint embodied improvisation

Michael KIMMEL
University of Vienna, Cognitive Science Platform[2]

ABSTRACT
This paper reports on projects researching sophisticated interaction and improvisation in five domains of bodywork, pair dance, and martial arts. A general taxonomy of the requisite competencies in proficient experts is worked out, which spans multiple timescales and ecological scales (individual, interpersonal). Taxonomy alone, however, cannot explain how concrete, rich tasks unfold: Only the micro-genesis of interrelations between component skills across cognitive, motor, and interpersonal levels provides an explanation, i.e. ways in which functionally distinct mechanisms coalesce or compete. Qualitative methods are proposed to map the diachronic and/or synchronic synergy-building whereby complex behavioral arcs arise. We found that many agent resources used for synergy-building reflect posits by *embodied, embedded, enactive, and extended* cognitive science. The "4Es" highlight mechanisms of interactivity, direct perception, and structural interpenetration. Complexity-related terminology, popular in "4E" debates, allows us to capture how component practices are metacognitively managed and how cross-catalysis between mind, body, interaction, and environmental parameters is stimulated.

KEYWORDS
Embodied interaction; micro-skills; empirical phenomenology; skill integration; synergies; comparison

INTRODUCTION

This contribution will survey micro-genetic research on embodied interaction expertise in *Tango argentino, Contact improvisation* (pair dance), *Aikido* (martial arts), *Shiatsu* and *Feldenkrais Method* (bodywork), all domains that feature multi-stage tasks, a complex interaction space, and an improvisational ethos.

Firstly, sophisticated joint improvisation through touch and kinesthesia is of intrinsic interest to motor, communication, and creativity pedagogy. Secondly, these arenas highlight how adaptivity and creativity impose necessities for orchestrating an array of skill sets in real time. Our sample showcases how elements of a "skillscape" coalesce, while responding to interaction dynamics. Thirdly, studying improvised interactions micro-genetically can enrich theorizing on embodied, enactive, embedded, and extended ("4E") cognition. A micro-genetic approach can be of considerable service here by testing, nuancing, and – where necessary – hedging 4E claims about the multiple timescales of embodied skill, the power of process, interactivity, and physical-structural interpenetration, and how cross-catalysis between micro-elements scaffolds performance.

Our overall explanandum is the ongoing interplay of skilled practices. As a first step, a taxonomy of practices identifies requisite sub-functions of the "skillscape". Our research targeted skills for *self- and interaction-management*, e.g. patterns of posture, breath, muscle tone, inter-body geometry, functional chains between bodies, complemented by how repertoire is cognitively organized and deployed. We also investigated recognition-primed cues for choosing action continuations, as well as perceptual feedback used in micro-coordination, task repair, transitions, and rerouting. Subsequently, we targeted *improvisational creativity in duets*, as agents chain modular action elements or *softly assemble* elements from their embodied "toolbox". We charted (self-created) creativity constraints, constrained exploration, interactive sources of creative inspiration, and looked at the emerging stream of micro-actions (transition points, perceived continuations at junctures, etc.).

Coregulated interaction

Many team sports, dances, circus activities, horseback riding, or field sports involve continual improvised interaction. Communication via direct perception of co-present individuals is termed *coregulation* (Fogel, 1993) (aka *participatory sense-making:* Di Paolo & DeJaegher, 2007). This contrasts with scripted social performance where individual contributions do not cross-trigger one another or where the (pregiven) task outcome needs only

[2] Address: Universitätsstrasse 7, A-1010 Vienna; Contact e-mail: michael.kimmel@univie.ac.at

minor, often timing-related adjustments between agents. Coregulation thrives on direct embodied coupling and, frequently, *kinesthetic interconnectivity*. From the bidirectional information and force flow reciprocal causation results; agents simultaneously shape and are shaped. Sensitivity for *interactional contingencies* cues them to one another and allow closely anticipating, modulating, and blending in with other. This allows them to micro-coordinate actions, negotiate choices, and exploit sudden options. Hence, coregulation is continuously adaptive, often improvised. For a task to emerge interactively, rapid modulation skills are needed, while a well-orchestrated fit between sub-skills must be ensured by meta-skills. The wider challenge is that (1) coregulated interaction streams are difficult to "discretize", (2) that role-specific perception-action skills need to be mutually calibrated to ensure togetherness (e.g. Tango leaders and followers have different abilities), (3) and that the open task structure generates myriad action trajectories, describable by dynamic systems mathematics as complex action spaces or as a set of principal interaction components (e.g. Torrents, Coterón, Ric, & Hristovski, 2015).

METHODS

Microgenetic methods with exceptional grainsize – from several seconds down to sub-second scale – were employed for interaction analysis of selected short events. A team of researcher-practitioners each investigated their "native" skill (Aikido, Tango argentino, Contact Improvisation, Feldenkrais, and Shiatsu), guided by shared research questions. The expert practitioners acted as principal interviewers in their domain, and contributed ethnographic observations as well as practitioner diaries.

Our micro-genetic elicitation method draws on tried-and-true paradigms of *empirical phenomenology* (*Explication Interview*: Vermersch, 1994; Petitmengin, 2006; Høffding & Martiny, 2015, cf. pico-level *Event analysis* Steffensen, 2013). We applied Explication Interviews and Think-alouds as follows: In all domains save Contact improvisation, 24 dialogical interviews were done in a first study. Specifically, we interviewed six learners thrice across one year for longitudinal analysis; additionally six one-time interviews with teachers focused on complex heuristics and concepts. In a second, ongoing study, this time with teachers only, Contact improvisation was added as a domain (6 interviews and 20 think-alouds so far), to be extended with Aikido and Tango think-alouds in the immediate future.

To analyze interactions, we dissect selected events to unpack their micro-process management. The events are of short duration and selected somewhat like in the Critical Decision Method (Klein, Calderwood, & MacGregor, 1989). We focus on interesting moments of choice, transition, critical adaptive response, failure, risk taking, or interaction "magic". Participants first chart the micro-event in its global structure and slice it at their natural junctures (notably decision points or motor transitions) for extended scrutiny. 20-80 minutes are spent on 2-5 seconds of interaction, with both persons contributing their view. Two modes are applied: In dialogic ***Explication Interviews***, the interviewer acts in a maieutic functions; she creates a continuous attentional flow and stabilizes the respondent's awareness to enter into a "thin slice" of interaction. She strictly guides away from associations and explanations. "How" or "what" questions with strong sensory grounding are used, never "why" questions. Specifications may be requested as regards the locus and quality of sensations, the scope of action planning, timing, and any triggers provided by the other person. Secondly, in ***Think-alouds*** experts' subjective experience is elicited during or right after improvised interaction in one of two ways. (a) *Natural sparring* with an agreed thematic focus: Participants are provided with video-feedback, asked to select key moments, and discuss them. The maieutic question techniques match those of solo interviews, but are augmented by special items that focus on transitions, motor control, contextual adaptivity, softly assembled solutions, as well as creativity (e.g. how known elements are combined into something new). (b) *Quasi-experimental variation tasks:* Agents are asked to perturb the interaction dynamic in specific ways to detect how the interpersonal whole adapts. We give instructions to gear up speed, to become less reactive or tense, increase risk or difficulty, alter the micro-timing, or simply change attentional focus, cooperativity, speed, geometry, or cognitive task load. Repeated variations reveal how creativity is impacted by this. Also, by rescaling or blocking relevant interaction parameters the interplay of subskills can be explicated ("When does individual motor difficulty become so taxing it stifles creativity?", "When does task repair begin to fail and rerouting become a more attractive option?")

SKILLS ACROSS TIMESCALES – A TAXONOMY

In all domains, action and perception skills span multiple timescales. Timescale interdependencies occur notably around enduring and deeply entrenched habits, which embody the domain's "grammar". These background constraints provide a framing and functional enablement for the various actions of shorter duration.

A prepared body

General socio-cognitive orientations shape the general flavor of joint activities, which "trickle through the system". They express consensual interaction *frames* (Fogel, 1993) of the domain, e.g. respect, mindfulness, or a focus on creativity (or efficiency). Moreover, enduring somatic modes are created to be ready for interaction:

- General enablers. A conspicuous fact is that a general "grammar" kicks in when practitioners start to dance, to grapple, or to touch a bodywork client. Patterns of posture, tone, gaze, breath, or attention form a permanent backdrop. Firstly, foreground action is hereby enabled, as body structures and attention are subtly preactivated to contribute synergistic strengthening. Secondly, "grammar" limits what is considered meaningful movement in the domain. Thus, Tango is geometric, crisp, and separates linear from circular.

- Poisedness. Bodies must be poised for action-readiness and improvisational fluidity. They cultivate what complexity theory calls "metastability", i.e. hovering in states that make all possible directions equally accessible, and prepare for surprises. Dancers and martial arts experts stay "in axis", for example.
- Modulation of extended patterns. Agents recognize higher-level interaction dynamics, e.g when an interaction partner is nervous or misjudges what has priority. Experts may subtly "tweak" dynamics (beyond their sole control) via multiple enabling actions or by relaxing someone through trustful rapport.

Perceptual micro-skills

In embodied interaction domains, perceptual skill largely works in recognition-primed ways and by immediate transduction into action (Ross, Klein, Thunholm, Schmitt, & Baxter, 2004), at least to the extent that time-pressure exists. Our research confirms the 4E claim that cue recognition is mostly not passive, but depends on skillful microscopic probing actions and in configuring one's body for receptivity:

- Enactive perception. Experts know where to direct embodied attention and use micro-actions in gaze (O'Regan & Noë, 2001), dynamic touch (Turvey & Carello, 2011), etc. For example, a bodyworker can variously scale touch to target a client's skeleton, muscles, or fascia, depending on intention.
- Epistemic probing actions. Agents employ feedback generating epistemic actions in addition to pragmatic actions (Kirsh & Maglio, 1994). Practitioners may start with epistemic probing actions and scale them up into pragmatic actions, whereas the hallmark of experts is to act right away and probe as one does so.
- Background attention. However, we also found passive sensory vigilance for particular "gatekeepers" cues that activate more focal subroutines or signal a need for task switching, e.g. alarm signals.
- Recognition of invariants. "Smart perception" skills (Runeson, 1977) designate perception of complex informational arrays at a glance or touch. This happens by scanning for high-level invariants in perceived interpersonal dynamics, space, balance, muscle tone, or the like. Thus, Tango dancers are simultaneously receptive to relative weight between partners, spatial geometry, distance, axis differentials, and chest-hip twist. This captures the gist even of novel situations. Important for improvisation theory, experts needn't memorize continuation paths of key situations, they can feel open paths in any situation.

(Inter)action skills

Besides the discussed "grammar", coordination as a couple demands *situated* enablement, ongoing responsiveness to contingencies, and skills for micro-coordinating interaction down to the minutest details:

- Preparatory and enabling actions. Immediately prior to actions, agents use self-modulation, e.g. of breathing habits, to get ready. They also report attention guiding inner verbalizations (Sutton, 2007) and other attentional techniques. Inversely, agents may physically mobilize or invite partners.
- Informational enablement. Before interaction can start, swift and unambiguous communication channels must be prepared. Experts create rapport and maximize interbodily resonance as well as interconnectivity, through particular ways of touching, geometry, and inner organization of the body (e.g. muscle chains that transmit signals). This encompasses skills for getting information as well as for making oneself readable by the other, and in martial arts the opposite – concealing information as long as possible!
- Communicative actions for micro-coordination. A lot of communication happens through push and pull, especially in Aikido; Shiatsu and Feldenkrais may also use subtle effective physics. Note, however, that in dance a vestige of conventional signals may blend with this. Tango leaders issue "invitations" that followers read. Leaders may scale signals up or add further synergists when a partner is slow to respond.
- Task-specific micro-coordination. To complete tasks jointly agents must ensure proper relative timing, geometries, and force deployment. The above communicative actions occur in conjunction with pragmatic actions like pushing, walking, rolling, rotating, jumping, blocking, diverting force, pressing or palpating. Proper micro-coordination requires utmost feedback sensitivity. To achieve this feat in multi-phasic actions, agents react to cues signalling sub-action onsets so that partners never lag behind or overshoot.
- Task-protective modulating. Agents ongoingly monitor task vectors and geometry, force, etc., to correct movements when deviating from the appropriate sensorimotor yardsticks. (Familiar interpersonal tasks can be guided by distributed action concepts, which specify motor and feedback control for oneself as well as idealized perceptual feedback from others.) Tasks are always fine-tuned interactively. Sometimes, agents even renounce fixed action concepts and find novel interaction solutions (see below).
- Global constraint management. In multi-phasic tasks, both parts and wholes constrain action. E.g., an Aikido defensive technique, even when improvised, has a global logic of toppling opponents step by step.
- Transition management. At the junctures between action components agents must preserve continuity, which is difficult in improvised chains of elements. (At junctures, especially dancers may also "punctuate" interaction for some physical reconsolidation and for "resetting" into a perfect state of metastability.)
- Interpersonal compensatory modulating. Agents strive to create optimal interpersonal synergies, i.e. ideal macro-patterns such as breastbone opposition in Tango or an ideal defensive geometry in Aikido when the

defender has moved out of line. In cooperative interaction domains, agents frequently compensate in real time for partner glitches to preserve a relational ideal (Riley, Richardson, Shockley, & Ramenzoni, 2011), e.g. by cancelling out the partner's imbalance through subtle weight shifts.

Improvisation skills

Growing mastery leads to a growth of repertoire, but also a decomposition of its structure to allow for fluid improvisation. Able improvisers develop a *hierarchically organized memory* (Pressing, 1998) that can variously draw on scripts, modular interaction elements, or small dynamic primitives (based on perceptually analyzing all distinct control-relevant dimensions). Masters extract principles and categories of functional situations. Creativity itself (Barrett, 1998; Sawyer, 2003) may consist of "idea-driven" creation of new elements or new combinations. Alternatively, interactive resources are used: calculated risks, "playing", exploring, as well as adaptive responses to, or nudges from, the partner. A set of micro-skills support improvisation:

- Affordance "surfing". In much improvisation modules are selected and chained in real time. In a wide-cast attentional mode, agents remain open to emergent affordances (Gibson 1979) of whatever kind, which they select (and act on) without prior deliberation. They may either follow suggestive, "easy" paths or factor in intentions like leaving the beaten path and seeking novelty.
- Non-enforced micro-scripts. Agents may also plan ahead chains of multiple modules, a mini-script. Scripting in interaction, however, is inherently risky due to what the partner or others (on a dancefloor) are doing. Therefore, hybrid cognitive modes arise in which agents complete a tentative script when all is well, but do not enforce it: They remain attentionally poised for immediate switching.
- Continuation selection. Embracing a new trajectory presupposes monitoring perceptual invariants (see above) that specify available continuations. Agents select from the set of affordances by matching invariants with their intention at junctures known to be suitable for transitions or rerouting on-the-fly.
- Creating options by scaling and at a remove. Options may not just arise, but require subtle preparatory actions (scaling up geometry or force variables, etc.). E.g., Aikido defenders actively perfect body-front and arm geometries. Agents also employ strategic micro-actions that produce usable reactions or configurations at a remove ("I do X for my partner to do Y, which let's me do Z").
- Recontextualizing errors to convert some unintended glitch to a new higher-level pattern (notably possible in bodywork sessions: due to their extended duration, the sequential structure as a whole counts).
- Dynamic repairs and "morphing" of action within a given aim occur when task-correction is too costly.
- Soft-assembly allows (1) customizing synergies to a situation and (2) creating novel interaction patterns at unprecedented levels of innovation. This presupposes mastery of multiple control dimensions and dynamic primitives. The challenge is to mix multiple control dimensions in real time; agents must know how dynamic primitives constrain/enable one another to generate a fitting, yet novel mix. Mastery of constraints and principles helps here (e.g. "initiate action from the body center", "blend with the other's movements").

The power or process and interactivity

Confirming another 4E hypothesis, we document the power of interactive processes. In joint improvisation agents use the coupling dynamics and its processual unfolding as a source of scaffolding:

- Exploiting interactivity. Experts exploit solution probing (Steffensen, 2013) and the interaction dynamic (Kirsh, 2014): It is easier to think by manipulating interpersonal structure, because interacting clarifies constraints on the go. This simplifies choices and leads to the discovery of unexpected affordances.
- Dynamic structural specification. Experts generate options while acting, rather than thinking up solutions in advance. They trust that ongoing action will supply structural specification underway.
- Dynamic immediacy. High-level experts are capable of refined, "thin-sliced" perception and of blending with the dynamic. This level of awareness allows staying in tune with micro-changes, dyadic emergence, and unexpected opportunities; it also prevents overshooting or delayed actions (Kimmel et al., 2015) while one "stays in the zone". Subtle and new creativity options may arise if both partners do this.
- Distributed agency and emergent dynamic scaling. In Contact improvisation, in particular, interpersonal dynamics occur in which the individual contributions are so tightly interspersed in reciprocal causation loops that both partners feel they are not the actual author of what happens. Their phenomenology of agency changes, perhaps pointing to downward causation from interaction dynamics to individuals.
- Extra benefits: Rapport and technical perfection can "invite" interpersonal cross-catalysis (see below).

SKILL MATCHING

Ultimately, we must go beyond a mere list of component skills and ask how "coherent constellations of mutually supportive component practices" (Hutchins, 2014, p. 40) form a "system of intermodular integration" (Carruthers, 2002). Interaction tasks demand a matching up of the various sub-skills in commensurate ways. Notably, dexterity and interaction skills must be integrated with cognitive-intentional expertise. In Tango, for instance, leaders need to (a) perceptually analyze the configuration and available space, (b) search the repertoire for a preferred technique that fits this, and (c) communicate lead signals fluidly and with correct micro-timing, (d) while coordinating each minute part of the action with the partner's co-action.

Behavioral arcs and their component synergies

To model skill integration, we investigate how micro-skills coalesce in *behavioral arcs* (Kimmel, 2016). That is, micro-skills unfold in cascades of overlapping, functionally co-specifying skill components. These components (a) entertain **co-dependency relations** of a synchronic or diachronic kind, (b) answer to **constraints** on multiple timescales (including prospective and retrospective constraints), and (c) display **synergies** that exceed the sum of their parts, but are needed for a good macro-ordered behavior. To explain this, the twin notions of synergy and self-organization from dynamic systems research (Turvey, 2007; Kello & Van Orden, 2009) state that arrays of micro-players can create coordinative solutions (i.e. macro-performance variables), from dynamically recruited ensembles. When micro-players are linked in a web of excitatory and inhibiting relations, non-linear effects may result as the emergent macro-structure reverberates back on the parts, dubbed *self-organization.* That is, how certain key parameters – like quality of rapport – behave controls the macro-patterns that emerge from the interplay of micro-processes. These macro-ordered states, as they stabilize, may align micro-components in return. Speaking of self-organization thus highlights self-amplifying and other non-linear effects in multi-component networks of, both, individual motor control and body ensembles behaving as a synergistic unit.

To track these processes across different interaction scenarios, we chart the micro-actions and micro-percepts in one agent over a few seconds and how these intersperse with the partner's micro-percepts and -actions. This reveals which actions co-occur, when sub-element set in and fade out, how partners mutually trigger each other, and how transitions are prepared. All this can be visualized on interaction scorelines of 2-5 seconds length with perceptional and actional subscripts for each phase. To illustrate some outcomes, in Aikido bouts (Figure 1) several insights arise: First, micro-actions are cumulative. Defenders first evade the attack, then compromise the attacker's axis, etc. (3-4 phases). Second, defenders may use ad hoc repairs if needed; they may also obviate unwanted contingencies by picking up prospective signals and accordingly add actions (like a quick distracting punch) or re-scale incipient actions (like giving a lever less play). Third, tricky transitions phases arise where the defender must preserve energetic continuity and perfectly control the attacker's degrees of freedom. At these points alternative techniques can set in, if needed. Fourth, in terms of feedback we found a general orientation of both Aikido practitioners to balance-related cues (and balance differentials), besides many – in part technique-specific – cues for sequencing. Each phase has specific visual, kinesthetic, and tactile signatures.

Figure 1: Scoreline of an Aikido technique (Subscripts for micro-percepts, micro-actions, and intentions may be added).

This analysis extracts task components in their functional interplay, e.g. the relative timing of leg and torso action. To identify synergies we (1) chart macroscopic performance parameters, (2) networks of elements giving rise to them, (3) their trade-offs, and (4) dynamic progressions. In terms of progressions, synergies may begin with key elements, which prepare the ground and help latter ones to fall into place. The strategy of using core properties and scaling up others if needed is also frequent, e.g. a martial artist adding a slight rotation to an interception to gain the upper hand. Also, specifics of interpersonal timing, spacing, or force at t_1 elicit different synergistic strategies at t_2. Path-dependent dynamics can thus be specified, depending on initial approach timing, etc. Finally, *simplexity* relations (Berthoz, 2012) pack multiple functions into one component. E.g., axial alignment in Tango or Aikido benefits uprightness, motility, efficiency, and interpersonal signal transmission.

Metacognition and stimulated self-organization

Skill matching requires metacognitive guidance (Cohen et al., 1996 MacIntyre et al., 2014). Agents have task-specific knowledge of how to create a good fit of means and strategies for dicey situations. Metacognitive guidance triggers items from one's "embodied toolbox" when these are not reflexes.

Yet, skill matching goes well beyond centralized control. To begin with, metacognition merely directs attention to required resources, without appropriately solving the real-time matching task yet. Full solutions arise only as the engagement in interactive micro-cascades continues. Moreover, processes extend over brain-body-social-environmental realms and exploit "extended cognition" (Clark, 2008): Agents benefits from the intelligence of tendons and fascia (*tensegrity* networks, Silva, Fonseca, & Turvey, 2010). Limbs self-organize through preset tension, elastic tissue effects, or maximal movement ranges for "passive" control. In *kinesthetically interconnected body ensembles* similar functions can also be inscribed into, say, a dance embrace that creates stable musculoskeletal chains between bodies. Physical-causal properties of *dynamic mutual incorporation* (Froese & Fuchs, 2012) are thus exploited. Superior performers learn to skillfully configure a wider system for cross-catalysis, so benefits "come for free" to them. They habitually exploit advantageous self-organizational processes between body, others, and space. At the same time, experts must also develop a particular type of meta-awareness concerning which *potentially steerable* things to leave alone and where to steer. As flexible managers of their own cognition experts sensitively combine "auto-pilot mode" and conscious guidance. They give self-organization its due and exploit intrinsic system dynamics, while also transforming it through well-timed interventions and decisive manipulation when needed. Much agency is not automatic, but strategic. Metacognitive competency steers when needed, while stimulating processes indirectly, and exploiting higher timescale "pre-settings" in addition to one time micro-actions. These observations make positions questionable which claim that coping is non-conceptual (Dreyfus, 2002) or that awareness "chokes" performance (Beilock, Wieringa, & Carr, 2002). Note that expert sensitivities for synergistic couplings across minds, bodies, and space are compatible with a related claim of 4E theory. Intention is often seen as one constraint on behavior among several, rather than a direct cause. Contact improvisation supports this best; here cumulative, distributed micro-causalities dominate. Meanwhile Tango, Aikido, and bodywork where one agent "leads" temper this claim.

Complexity-related expertise

Metacognition also guides responses to contingencies when strategic alternatives exist, especially in Shiatsu and Feldenkrais, where bodyworkers diagnose a client's state and gradually help transform the bodily interplay into a healthier, equally sustainable state. Metacognition is complex because bodyworkers think of the human body as self-organizing network of functions and reckon with non-linear processes. Accordingly, they learn to use cumulative intervention strategies (Kimmel, Irran, & Luger, 2015) and stimulate the cross-catalysis of multiple interacting components. Besides encouraging general mindfulness and trust, this stimulation draws on much *strategic* expertise: One key resource is functional anatomy knowledge, i.e. ideas about synergies between muscles, bones, fascia, etc. in networks of interlocking components. Different portions of the expert's mental anatomy model can be contextually cued in order to select focal points of intervention, decide on the sequencing of action modules, and whether stimuli should be "wide bandwidth" or target specific body areas.

CONCLUSION

We have looked at *complex competency systems* – skillscapes – in which resources must be selected and meshed in real time to lend agents improvisational flexibility in response to, both, adaptive demands and their own creative needs. Micro-genetic methods can reconstruct this tight interplay of skills, which is ideally synergistic, but can be subject to cognitive resource competition. Explication interviews and especially micro-genetic think-alouds are granular enough to capture this emergent causal interplay. Meanwhile, quasi-experiments allow experts to discover and verbalize normally reflexively deployed and often subtle causal trade-offs between parameters. This qualitative paradigm enriches the motor skills and interaction literatures in several ways:

Firstly, sophisticated interaction clearly resists experimental reduction. What skills get deployed jointly responds to interaction dynamics, situated boundary conditions, agent resources and intentions, and non-linear parameter trade-offs. Scholars need to grasp how sensorimotor, interaction, and decision skills cross-catalyze or constrain one another (i.e. agents prioritize aims, e.g., by curbing creativity for the sake of rapport). Secondly, modeling the micro-structure of synergy-building processes (Latash, 2010; Latash, Scholz, & Schöner, 2007) direly needs qualitative grounding. Little is known as to how agents create synergies, how many and which elements these involve, what makes for a balanced mix, which elements are central, and how all this is subject to cumulative information streams criss-crossing between bodies. Thirdly, our take on micro-synergies makes creativity amenable to qualitative study, quite beyond recombined basic forms: Experts (a) combine and coordinate dynamic primitives, while (b) exploring openings in the constrained task space, (c) incorporating partner nudges, (d) and cultivating precision, poisedness, and rapport to prepare the ground for cross-catalytic benefits.

In sum, this 4E-based approach highlights traditionally underestimated resources of skilled coping. With embodied interconnectivity, interpenetration and joint physics, the need for planning and anticipation or fixed representations is mitigated (Marsh, Johnston, Richardson, & Schmidt, 2009): In particular, interpersonal micro-coordination (i.e. *execution*) requires no cognitive posits like anticipatorily shared goals (Knoblich & Sebanz, 2008). In experts, real-time skills and direct perception are rapid and very robust, provided that bodies interpenetrate each other informationally and structurally. Internalistic cognition may explain approximate action *selection* processes and strategic process management, albeit only with updatable action concepts that respond to emerging constraints and allow for fine-tuning. What's more, fixed forms are questioned by 4E: While learners mostly use "readymades", masters validate the claim of – at least as one possible mode – softly assembled, self-organizing solutions. In this, body precalibrations, passive dynamics, and interactivity are equal partners to internalistic resources and bound together in massively interlocking cascades. Our micro-genetic paradigm explicates the deployment of skill sets with sensitivity to all these interdependencies and suggests that finely honed qualitative tools hold promise in explaining how experts regulate complex skillscapes.

REFERENCES

Barrett, F. J. (1998). Creativity and Improvisation in Jazz and Organizations: Implications for organizational learning. *Organization Science*, *9*(5), 605–622.

Beilock, S. L., Wieringa, S. A., & Carr, T. H. (2002). Expertise, Attention, and Memory in Sensorimotor Skill Execution, *55*, 1211–40.

Berthoz, A. (2012). *Simplexity*. New Haven & London: Yale UP.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674.

Clark, A. (2008). *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*. Oxford: Oxford UP.

Di Paolo, E., & DeJaegher, H. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

Dreyfus, H. L. (2002). Intelligence Without Representation–Merleau-Ponty's critique of mental representation the relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, *1*(4), 367–383.

Fogel, A. (1993). *Developing Through Relationships. Origins of Communication, Self, and Culture.* Chicago: The University of Chicago Press.

Froese, T., & Fuchs, T. (2012). The extended body: a case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences*, *11*(2), 205–235.

Høffding, S., & Martiny, K. (2015). Framing a phenomenological interview: what, why and how. *Phenomenology and the Cognitive Sciences*.

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philosophical Psychology*, *27*(1), 34–49.

Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*(1), 57–78.

Kimmel, M. (2016). Embodied "micro-"skills in tango improvisation – How a collaborative behavioral arc comes about. In F. Engel & S. Marienberg (Eds.), *Out for a walk. Das Entgegenkommende Denken* (pp. 57–74). Berlin: DeGruyter.

Kimmel, M., Irran, C., & Luger, M. (2015). Bodywork as systemic and inter-enactive competence: Participatory process management in Feldenkrais Method® & Zen Shiatsu. *Frontiers in Psychology for Clinical Settings*.

Kirsh, D. (2014). The importance of chance and interactivity in creativity. *Pragmatics & Cognition*, *22*(1), 5–26.

Kirsh, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*, 513–549.

Klein, G., Calderwood, D., & MacGregor. (1989). Critical Decision Method for Eliciting Knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(3), 462–472.

Klein, G., Moon, B. M., & Hoffman, R. R. (2006). Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems*, *21*(4), 70–73.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2021–2031.

Latash, M. L. (2010). Stages in learning motor synergies: A view based on the equilibrium-point hypothesis. *Human Movement Science*, *29*(5), 642–654.

Latash, M. L., Scholz, J. P., & Schöner, G. (2007). Toward a new theory of motor synergies. *MOTOR CONTROL-CHAMPAIGN-*, *11*(3), 276.

Marsh, K. L., Johnston, L., Richardson, M. J., & Schmidt, R. C. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology*, *39*(7), 1217–1225.

Petitmengin, C. (2006). Describing one's subjective experience in the second person, an interview method for the Science of Consciousness. *Phenomenology and the Cognitive Sciences*, *5*, 229–269.

Pressing, J. (1998). Psychological Constraints on Improvisational Expertise and Communication. In B. Nettl & M. Russell (Eds.), *In the Course of Performance: Studies in the world of musical improvisation* (pp. 47–67). Chicago: University of Chicago Press.

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, *2*(38).

Ross, K. G., Klein, G. A., Thunholm, P., Schmitt, J. F., & Baxter, H. C. (2004). *The recognition-primed decision model*.

Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, *18*, 172–79.

Sawyer, K. (2003). *Group Creativity. Music, Theater, Collaboration*. Mahwah, NJ: Erlbaum.

Silva, P. L., Fonseca, S. T., & Turvey, M. T. (2010). Is tensegrity the functional architecture of the equilibrium point hypothesis. *Motor Control*, *14*(3), e35–e40.

Steffensen, S. V. (2013). Human Interactivity: Problem-Solving, Solution-Probing and Verbal Patterns in the Wild. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition Beyond the Brain* (pp. 195–221). London: Springer.

Sutton, J. (2007). Batting, Habit and Memory: The Embodied Mind and the Nature of Skill. *Sport in Society*, *10*(5), 763–786.

Torrance, S., & Froese, T. (2011). An inter-enactive approach to agency: Participatory sense-making, dynamics, and sociality: *Humana.Mente*, *15*, 21–53.

Torrents, C., Coterón, J., Ric, A., & Hristovski, R. (2015). Emerging dance movements under ecological constraints in Contact Improvisation dancers with different background. In *CS-DC'15 World e-conference*.

Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, *26*(4), 657–97.

Turvey, M. T., & Carello, C. (2011). Obtaining information by dynamic (effortful) touching. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1581), 3123–3132.

Vermersch, P. (1994). L'Entretien d' explicitation en formation continue et initiale. Paris: ESF.

Planning Studies: Empirical investigation of using team sensemaking for early problem formulation

Andrew LEGGATT[a], Nicola TURNER[b] and Paul PEARCE[c]
[a]*Trimetis Ltd*
[b]*BAE Systems*
[c]*Dstl*

ABSTRACT

The UK Ministry of Defence conducts much of UK Government's analysis and research into complex problems. It does this with many techniques but often uses a branch of applied mathematics called Operational Analysis to address complex and/or complicated problems which have important implications for future operations, force development, acquisition, tactics and doctrine. Various studies on these topics produce evidence to inform decision making; therefore, it is important that these studies are well planned to ensure they provide the most suitable evidence possible. To do this Dstl wanted to assess whether a new planning approach, the Initial Analysis Estimate, improves sensemaking which in-turn improves study quality. This paper reports on the application of TSAM (Team Sensemaking Assessment Method) to empirically assess whether the Initial Analysis Estimate, which drew heavily on Cognitive Edge's Cynefin sensemaking framework, improves team sensemaking. The study indicates that compared to a control group the instances of team sensemaking were greater using Dstl's early problem investigation techniques.

KEYWORDS

Sensemaking; Military Analysis; TSAM (Team Sensemaking Assessment Method); Study planning; Initial Analysis Estimate;

INTRODUCTION

Operational Analysis (OA) within the UK Ministry of Defence (MOD) supports decision makers and their commissioners in making informed decisions for a range of problems, from relatively simple well bounded problems to complex unbounded or wicked problems. Regardless of the type of problem to be analysed there are many evidence challenges faced by MOD, for example Levene (2015) identified issues with evidence based on advocacy, MacPherson (2013) addressed the appropriate use of models and methods and more recently Chilcot (2016) has considered issues with underlying assumptions. As a response to these challenges the Defence Science and Technology Laboratory (Dstl) has developed the Evidence Framework Approach (EFA). The EFA provides 'handrails' concerning problem formulation, the assessment of evidence quality and the appropriate use of models and methods. It is consistent with one of the key Government products emerging from the MacPherson review, the UK pan-Government-Department Aqua Book (2015) which provides guidance on accomplishing analysis with evidence that is fit for purpose, i.e. of an appropriate quality.

One key aspect of analytical quality assurance is effective exploration of the problem. The EFA supports problem exploration through enhancing sensemaking at the problem formulation stage of an analytical study, using a process called the Initial Analysis Estimate (IAE). The IAE process is centred around a rationale based on complexity thinking using heuristics that are also proposed within the Cynefin framework (see Figure 8) as a vehicle for improving sensemaking and ultimately improved evidence evaluation and assessment.

The Cynefin framework was developed by David Snowden (Cognitive Edge Limited, 2016) and was designed to support sensemaking and decision making in complex social environments (Kurtz & Snowden, 2003). The framework enables users to interpret contexts and decide on an appropriate type of action to take given the situation. It is made up of five domains: obvious, complicated, complex, chaotic and disorder. The first four domains are shown in Figure 8 below. Disorder is the space between the four other domains, where the user does not know their context, potentially leading to misinterpretation and incongruous action.

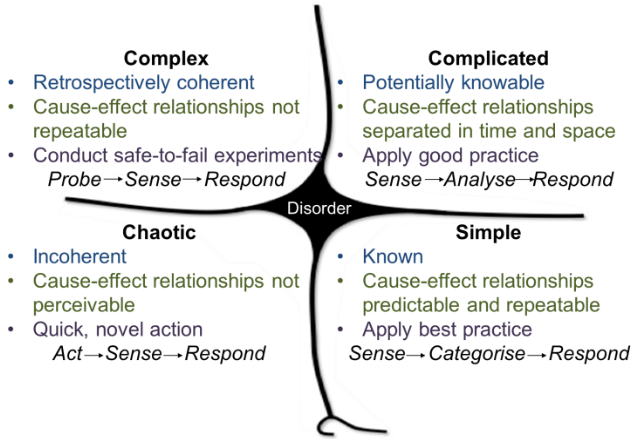


Figure 8: Cynefin Framework

Cognitive Edge Limited have developed a variety of methods based around the Cynefin framework, complexity thinking, and storytelling, including "Future, Backwards", "Four Points Contextualisation" and "Ritual Dissent" (Cognitive Edge, 2016). These methods promote discourse between people with the aim of arriving at a shared understanding of the problem, its context and the future action to be taken via capturing and sharing narratives. The Future, Backwards method enables the group to discuss perspectives on past events and possibilities for the future, working backwards from the current scenario and then future scenarios one step at a time. Four Points Contextualisation supports the group to build a Cynefin Framework from data points by highlighting four extreme examples; and Ritual Dissent enhances ideas by forcing groups to thoroughly scrutinise them.

The cornerstone of the IAE is enhanced sensemaking at the problem formulation stage, therefore, it is particularly important to be able to assess the quality of team sensemaking during the IAE. There are few known techniques developed to assess team sensemaking, one of which is "TSAM" (Team Sensemaking Assessment Method) (Hutton, Attfield, Wiggins, McGuinness and, Wong, 2011). The approach was developed from empirical evidence and scientific theory drawn from various disciplines that have contributed to the understanding of collaboration in complex sensemaking environments. It makes reference to heuristic evaluation methods (Neilson, 1993) and Cognitive Performance Indicators (Wiggins and Cox, 2010) type tools which allows assessors to inspect the sensemaking environment and from this make judgements about whether the environment supports sensemaking. In addition, TSAM is a principled inspection technique to help Human Factors practitioners determine whether newly designed systems were likely to assist or inhibit collaborative working. The technique was adapted to be used as a group working assessment method (Hutton and Leggatt, 2013) during group intelligence development. During this application it was judged to be a useful method for describing team sensemaking and its findings were corroborated by independent observations. TSAM identifies the following elements of collaborative sensemaking to be; information evaluation, information exploration, information organisation, process transparency, assessment sharing, collaborative assessment development, collaborative critiquing, and collaborative resolution.

The aim of this research is twofold, firstly to investigate the effectiveness of using a Cynefin framework type of approach for complexity thinking when conducting the sensemaking phase of the IAE, compared to a traditional approach; and secondly to explore how TSAM can be adapted for use in this context.

Hypotheses

It is predicted that using a set of Cynefin methods when conducting the IAE will result in:

H_1: Significantly better team sensemaking compared with using traditional methods.

H_2: Participant preference over using traditional methods.

H_3: Significantly increased quality, richness and depth of output compared with using traditional methods.

METHOD

Design

The experiment design was a repeated measures crossover design where all participants experienced all conditions and received both OA problems, and it was counter balanced to minimise order effects (see Table 6 below). The two example OA problems were provided by Dstl, one addressed the concept of Information Manoeuvre and the other concerned the concept of Armoured Infantry.

	Day 1	***Day 2***	***Day 3***	***Day 4***
Condition	BOGSAT (2)	Cynefin (1)	BOGSAT (2)	Cynefin (1)
OA Problem	Information Manoeuvre	Armoured Infantry	Armoured Infantry	Information Manoeuvre
Participant Group	A	A	B	B

Table 6 – Experimental Design

The independent variable (IV) was the method used to conduct the sensemaking stage of the IAE (Initial Analysis Estimate). Condition 1 entailed an assembly of Cynefin methods (Future Backwards; Four Points Contextualisation; and Ritual Dissent) with an additional step titled 'Shared-view', placed before Ritual Dissent, which was designed to bridge the conceptual gap between thinking about the problem space and producing an appropriate output. Condition 2 was the control condition, using the 'traditional' approach often referred to as the BOGSAT (Bunch Of Guys Sitting Around Talking) approach.

The IV was manipulated to determine the effect on the dependent variable (DV), the effectiveness of using complexity thinking for OA sensemaking. The DV was measured based on:

- Team sensemaking performance measures: Team Sensemaking Assessment Method (TSAM) is a principled method specifically designed to assess whether a system, process or technology supports team sensemaking (Hutton et al., 2011; Hutton & Leggatt, 2013).
- A questionnaire: This was given to participants following each condition, with a 1 to 5 rating of strongly disagree to strongly agree. The questions related to elements of team sensemaking and approach preference, it included items on engagement, usefulness, problem understanding, generating discussion, and questioning of ideas.
- Richness of IAE output: Information from IAE proformas and written material (i.e. post-it notes, flip charts, etc.) was transcribed and assessed by independent, experienced Dstl analysts using a questionnaire. The questionnaire assessed quality, depth and richness of output; positives and negatives; and qualitative feedback associated with the approach.

There were several variables that were predicted to have a potential effect on the DV, these were held constant to the extent possible and included: condition time; facilitation; group size; knowledge of techniques; experience of participants; size of room; materials available; and similarity of OA problem provided.

Participants

Twenty-Four Dstl and Army personnel with OA backgrounds and a variety of experience participated in the experiment. The median number of years of relevant experience was 16 and their ages ranged from 26 to 65.

Materials

The materials used for the experiment included three flip charts, many different coloured paper sticky notes, and three paper instruments for collecting data: a feedback questionnaire completed individually each day, a proforma IAE form, and a questionnaire to assess the outputs. A camera was also used to capture images of the outputs to enable them to be shared and discussed during and after the sessions.

Procedure

Two weeks before the trial date potential participants were identified and provided with limited information on the purpose of the trial. Those who volunteered were placed into two groups based on years of experience, with the intention of making the groups as equal as possible.

Each condition took place over two consecutive days from 0900 to 1300. Participants undertaking the Lead Analyst role were made responsible for delivering the IAE output at the end of each session using the proforma provided. This was followed by a short group interview about the participants' experience, including what they thought went well and what went less well; participants were also asked to complete a short questionnaire individually. During all sessions two observers made detailed notes with which to make their TSAM ratings after the event. This experimental procedure was repeated for the following two days but with a different order of OA question presentation and with twelve new participants.

On the first day of the two-day block participants were asked to take part in a standard BOGSAT condition with a military commissioner providing background to the question and a Lead Analyst who chaired the IAE and was tasked with the written summary. The chair was given freedom to conduct the BOGSAT in a realistic fashion and address the question within the given time, including taking two coffee breaks.

The following day, the Cynefin condition was led by an expert facilitator from Cognitive Edge Limited. The facilitator focused on ensuring the techniques were followed and instructing the participants at each stage of the process; he was not involved in the content of the session and did not ensure that participants remained focused on the problem. The participants were allowed approximately 45 minutes to work with each of the methods in the Cynefin condition (Future Backwards; Four Points Contextualisation; Shared View; and Ritual Dissent). During this process the groups of 12 participants were split up into two smaller groups of six people for the Future Backwards and Four Points methods, the groups were then combined to work in plenary for the Shared-view method, and then split again into two groups to conduct Ritual Dissent.

The outputs from each session were typed out and given to three Dstl OA practitioners to read and assess. They were asked to rate the outputs (using a five-point scale from poor to excellent) for information quantity, information quality and the usefulness of the content to help move the project forward. The results of this assessment are not reported here as they were only completed by three experts which was insufficient to provide statistical significant comparison.

The quantitative data from the questionnaire were analysed with multivariate statistical methods. The analysis started with a General Linear Model repeated measures MANOVA followed by T-tests when multivariate statistics were not applicable, a significance level of 0.05 was selected for all tests. A qualitative analysis was also conducted on the data using a thematic analysis approach.

RESULTS

Team Sensemaking

TSAM Analysis from Observation

In this study TSAM was applied by two observers who observed all four conditions and made detailed notes concerning the team working, collaboration, leadership, behaviours and communications. The observers summarised their observations and made independent ratings of each distinct period of activity during the IAE. For example, for the Cynefin approach separate ratings were made of the Future Backwards, Four Points, Shared-view and Ritual Dissent activities. For the BOGSAT condition ratings were undertaken for each of the three 1 hour sessions. This produced over four days of data, with a total of 22 different opportunities for ratings against nine different TSAM elements 198 different cells or intervals of observation (Cynefin sessions were each split into 2 different groups). A simple three-point rating scale was applied to each of the cells based on behaviour not being evident, evident at least once, or frequently evident (Hutton & Leggatt, 2013).

Observer's ratings were then compared. There was a high level of concordance between observers (in the region of 0.7 for inter-rater reliability). Where there were differences these were reconciled via discussion of the written observations and their interpretation and this was moderated by the principal investigator.

The TSAM analysis created a large amount of data but this section will only summarise some key findings necessary to make comparisons between BOGSAT and Cynefin groups. It will also provide insight into what elements of TSAM were particularly supported to aid future method development. Figure 2 below shows the relative percentage of team sensemaking elements supported by the two different IAE approaches. This summary indicates the proportion of the observation periods where none, some or many of the elements were observed. The observation periods were converted into percentages as there was one more observation period for the Cynefin approach which used a four-step process. Figure 2 shows that there were considerably more observation periods where none of the elements were identified during the BOGSAT approach compared to the Cynefin approach (37% compared to 10% respectively). There were similar percentages of occasions when there were some TSAM elements observed during BOGSAT and Cynefin (46% to 56% respectively). However, there were considerable differences with the number of occasions where many TSAM elements were identified (34% of the occasions in the Cynefin and only 17% for BOGSAT). These data broadly suggest that the Cynefin approach provides greater support for team sensemaking than the BOGSAT approach.

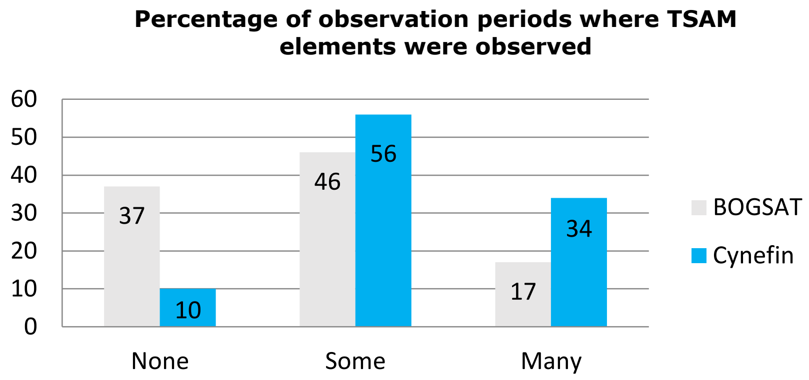


Figure 2 – Comparison of the percentage of TSAM elements observed in the IAE for the different IAE approaches

Figure below shows the summarised TSAM ratings for both conditions. The red elements indicate where there were no TSAM elements observed, the amber indicates where on at least one occasion a TSAM element was observed and the green indicates where on many occasions TSAM elements were observed. The Cynefin approach has two sets of observations because the participants split into two smaller groups during the Cynefin approach. The nine elements of team sensemaking described by TSAM would not be expected to be present in all stages of the IAE. For example, it would not be possible to conduct collaborative resolution without first developing a collaborative assessment. Therefore, it was entirely expected that there would be different TSAM profiles throughout the IAE and it was not anticipated that there would be a green profile throughout. However, the greater the green and amber ratio to red, the more likely it is that greater opportunities for team sensemaking were supported by the methods.

As shown in Figure 3 there are many more red sections in the BOGSAT condition compared to the Cynefin condition. Looking at the data in more detail there are other potential patterns, for example,there are many more red observation periods at the beginning of the IAE during BOGSAT than at the end of the IAE. This is characterised with 10 red sections during observation Period One (the first 1 hour session) and one red section during observation Period Three across both groups. This suggests that at the outset of the BOGSAT there are few team sensemaking opportunities as characterised by the TSAM approach.

The TSAM ratings data from the early BOGSAT sessions contrast clearly with those from the early (Future Backwards) sessions in the Cynefin approach. These sessions are characterised by many of the TSAM elements being observed and as a consequence there are many green and amber ratings (on average there are 5 red TSAM ratings in the BOGSAT sessions and 2.25 in Cynefin groups (to make a fair comparison an average needs to be used as there were two Cynefin groups and only one BOGSAT group) in the first session. These data suggest that at the outset there were many opportunities for team sensemaking which were supported by the Cynefin approach.

Questionnaire data

There were 24 responses to the questionnaire, however the same participants did not attend on consecutive days as intended. This meant that, although the right number of participants arrived each day, the within subjects statistical analysis was not as powerful as expected.

A repeated Two Factor General Linear Model MANOVA was conducted to test for the two main effects of sensemaking treatment and type of OA problem. There was no main effect found for type of OA problem, this suggests that the types of OA problem (Information Manoeuvre vs. Armoured Infantry) did not have any significant effect upon the participants' ratings. This general finding suggests that the two OA problems could be considered as broadly equivalent and suitable for the experiment design.

The participants were presented with a series of questions derived from the TSAM model of team sensemaking. A few of the elements identified significant differences between the BOGSAT and Cynefin methods. In particular participants rated the BOGSAT method more highly for, "the group drilled down into information to uncover the assumptions, gaps and limitations" than the Cynefin method (BOGSAT M=4.07 SD=0.70; Cynefin M=3.20 SD=0.86) **$F(1, 13) = 7.099$, $p<0.05$, $\eta_p^2=0.353$ (a medium size of effect)**. These data suggest that BOGSAT assisted participants by helping them examine relevant information and consider the assumptions with these issues more effectively than with the Cynefin methods. Another element which was rated as being different between BOGSAT and Cynefin was, "the group compared multiple ideas and revised assessment". In this instance the participants rated that the group compared multiple ideas and revised assessments more effectively in the Cynefin group (BOGSAT M=3.20 SD=1.01; Cynefin M=3.73 SD=0.70), **$F(1, 13) = 7.341$, $p<0.05$, $\eta_p^2=0.361$ (a medium size of effect)**. These data suggest that participants perceived that the Cynefin approach assisted the team in comparing multiple ideas and revising assessments.

Figure 3 – TSAM area plot for all conditions

Participants were asked to provide ratings concerning their experience in the conditions when using the Cynefin or the BOGSAT approaches, in order to investigate whether the experimental conditions were comparable to standard Dstl practice. For questions concerning levels of preparation and time allotted for the IAE there was no significant difference between sensemaking approaches. However, when answering the question, "My usual approach engages the group more" there was a significant difference between sensemaking groups with a medium effect size (**$t(12)=2.50$, $p<0.05$ $d=0.55$**) (a medium size of effect). In this instance participants rated that BOGSAT engaged them less than their usual approach (M= 2.92, SD=0.95); and that Cynefin engaged them more than their usual approach (M= 2.39, SD 0.96). These data suggest that the Cynefin approach promoted more engagement than their normal IAE. However, the participants rated the BOGSAT approach as more useful than the Cynefin approach (BOGSAT M=3.67, SD=.90 Cynefin M=3.00, SD=0.91; **$F(1, 13) = 7.031$, $p<0.05$, $\eta_p^2=0.351$**) (a medium size of effect).

There were no other significant differences between the sensemaking approaches regarding generating discussion, facilitating understanding, helping members of the team listen to each other or increasing the questioning of ideas. These findings suggest there were relatively few differences in function between BOGSAT and Cynefin.

Qualitative free response questionnaire data

The respondents suggested that a key benefit of the BOGSAT was that it provided the participants with the opportunity to analyse the question and think about the problem significantly before addressing the solution. This, combined with the presence of the commissioners and the ability to discuss some of the framing assumptions, meant that the participants using this technique were particularly keen to use this type of approach in the future.

These findings suggest that any approach which provides these functions (i.e. addressing the problem, looking at the framing assumptions, group working and access to the commissioner (and explore the commissioner's requirements)), would be welcomed and would support the group in reaching an acceptable output. However, these issues are generally described as hygiene factors, or generic assumptions, which should be expected in any well run meeting. The BOGSAT approach did not attempt to uncover the wider context or address which aspects of the problem were complex (as opposed to being merely complicated or even simple). Arguably, this requirement was only addressed by the Cynefin condition.

The qualitative feedback from the Cynefin groups was more varied than the BOGSAT groups. Many participants liked the structure of the sessions with different techniques which enabled the participants to consider wider issues and context and a small number of participants mentioned that it was helpful to address complexity issues. Also participants noted that the approach was lively and engaging. However, many participants were critical that the approach did not address the exploration and deeper understanding of the problem to any great degree and many participants would not be keen to use it again unless this aspect of the approach was modified.

DISCUSSION

Sensemaking is widely considered to be important for dealing with complex problems (Klein et al, 2003). The Cynefin approach suggests a number of methods that purport to support team sensemaking and therefore support better decision making in complex environments. Dstl often faces these complex type problems when planning and conducting studies to support commissioners and decision makers in the UK MOD. TSAM is a tool which was developed to assess team sensemaking behaviours; in this study TSAM was applied to a group working task called the IAE to help us identify the presence, or otherwise, of team sensemaking behaviours.

The findings from TSAM suggest that the Cynefin approach does indeed support more aspects of team sensemaking than the BOGSAT condition, especially during the earlier periods. The TSAM observation data were supported by the quetionnaire data, suggesting particiapnts did indeed consider more aspects of the situation and context in the Cynefin condition than in the BOGSAT condition. This may be partly attributed to a common phenomenon that proportionally fewer people speak when working in larger groups. Many problem structuring techniques begin by ‘broadening’ the consideration of the problem space in the initial stages. This encourages consideration of linkages to other ‘systems’ and the potential impacts of the factors that affect the system being considered. These are important factors in complexity thinking and this was not seen in the BOGSAT teams. Hence, it can be argued that engaging the whole team (not splitting into groups) may have a negative impact on early stages of the IAE by preventing team sensemaking.

It should be noted that this does not mean the BOGSAT approach was not undertaking important activities during the initial stages, but these were not activities known for increasing team sensemaking, and therefore BOGSATs were characterised by the commissioners describing the question and bringing new information to the group, this was not undertaken in the Cynefin treatment groups. This was not a particularly interactive period for most of the participants in the BOGSAT treatment and therefore they were generally quietly listening to the briefings. These sessions appeared to be very useful to the groups to help them appreciate the question but they were largely not interactive therefore scored low for team sensemaking.

Of note again, although these data suggest that there are many more team sensemaking opportunities in the early stages of the Cynefin approach it does not say anything about the quality of their outputs, only the process by which the output was generated. The quality of the overall output was assessed by three independent Dstl practitioners, although this did not provide any conclusive evidence. It may be that the quality of the outputs of the earlier, intermediate stages may have been better in the Cynefin approach but that later stages did not permit these to be exploited in the final output.

CONCLUSION

On balance, the Cynefin approach provided a very different approach to the BOGSAT and it has the potential to uncover wider issues associated with the problem at hand and to open discussion as to aspects of complexity in an engaging manner. However, in its current state it does not meet the participants’ needs to address the problem analysis and lead to a well considered output. As the Cynefin approach made little use of the commissioners to explore their requirement, this suggests that it would be helpful if some sort of Cynefin approach could be applied which kept the participants focused on problem analysis with the benefits of both understanding the wider context and obtaining a more detailed understanding of the commissioner’s requirement whilst exploring complexity.

TSAM is a tool which provides Human Factors practitioners with the structure and language to describe and analyse team sensemaking. This case study provides an example of its application in a planning and analysing context whereby the TSAM observation data was supported by responses from participants. The authors would welcome other Human Factors practitioners to try the technique to determine whether the method is a useful approach to assess the quality and quantity of team sensemaking when groups are tackling complex problems.

ACKNOWLEDGMENTS

The authors would like to thank CORDA for their kind permission for allowing this work to be reported in the open literature, we would particularly like to acknowledge Judith Rawle, Rachel Purkess, Stuart Burdett and Shannon Geer. Also we would like to thank Cognitive Edge Limited for providing a facilitator for the sensemaking techniques.

REFERENCES

Aqua Book (2015) https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government - Accessed 12 April 2016.

Chilcot, J., (2016) "The Report of the Iraq Inquiry Executive Summary" https://www.gov.uk/government/publications/the-report-of-the-iraq-inquiry - accessed 21 September 2016.

Cognitive Edge. (2016). Retrieved from: http://cognitive-edge.com/ - accessed 15th December, 2016

Hutton, R. J. B., & Leggatt, A. P. (2013). An evaluation of Ex ARRCADE FUSION Understand Intervention using Team Sensemaking Assessment Method (TSAM). Command, Inform Battle Management (CIBM) Task 8: Planning and Decision Support. TES Report 110635.Filton, UK: BAE Systems.

Hutton, R., Attfield, S., Wiggins, S., McGuinness, B., & Wong, W. (2012). Team Sensemaking Assessment Method (TSAM): An inspection method for evaluating collaborative analysis technologies. In M. Anderson (Ed.) Contemporary Ergonomics and Human Factors. Institute of Ergonomics and Human Factors.

Klein, G., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., Hollnagel, E. (2003): Macrocognition. IEEE Intelligent Systems (Volume: 18, Issue 3, May-June 2003).

Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. IBM systems journal, 42(3), 462-483.

Levene Report (2011)https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/27408/defence_reform_report_struct_mgt_mod_27june2011.pdf - accessed 13th April 2016.

MacPherson Report (2013) https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206946/review_of_qa_of_govt_analytical_models_final_report_040313.pdf - accessed 13 April 2016.

Putting the Recognition-Primed Decision (RPD) model into perspective: The revisited RPD model

Anne-Claire MACQUET[a] & Liliane PELLEGRIN[b]
[a]French Institute of Sports (INSEP), EA 7370,
[b] French Forces Centre of Epidemiology and Public Health (CESPA), Marseille, France

ABSTRACT

This paper addresses two issues of the Recognition-Primed Decision (RPD) model developed thirty years ago and enriched twenty years ago. These issues arose from the challenge of studying decision-making in the domain of Naturalistic Decision Making including high uncertainty and time pressure, and stress. Firstly, we are interested in how experts cope with high time pressure and make rapid decisions. Secondly, we are interested in how they filter information from the environment. As a corollary to these issues, we have revisited the RPD model and made it more dynamic by joining the three levels into a single element. We illustrated the revisited RPD model using empirical support from studies conducted in different areas. The revisited RPD model enables us to gain a richer understanding of the way experts use their experience to make decisions in dynamic settings including stress and uncertainty.

KEYWORDS

Decision-making; Education and training; Health; Transportation; Expertise; Dynamic situations.

INTRODUCTION

Thirty years ago, Klein, Calderwood and Clinto-Cirocco (1986) developed the Recognition-Primed Decision (RPD) model to explain how experts use their experience to make their decisions in natural settings and provide an alternative to classical models of decision-making. The present article aims to put this model into perspective in order to revisit it from empirical support. Classical decision-making theories claim that the decision-maker selects an option after comparing possible options on the basis of subjective expected maximization (e.g., Tversky & Kahneman, 1974). These theories were developed from laboratory studies. They do not explain decisions made in natural and dynamic environments, in which people do not have time to compare the trade-offs (Klein, 2008). Dynamic environments present key contextual factors: (a) ill-structured problems; (b) uncertain, dynamic environments; (c) changing, ill-defined, or competing goals; (d) decision loops (more than one decision decision); (e) time pressure; (f) high stakes; (g) multiple players; (h) organizational goals and norms (Zsambok, 1997).

The RPD model (Klein, 1997; Klein et al., 1986) was built from a study on how fire-fighters commanders made their decisions concerning the needs to look for and save people, fight the fire or adopt a fall-back position, as well as the place to undertake actions. A decision tree was not used in such decisions. The commanders reported that they did not make decisions meaning they did not make choices among several options, nor did they assess options. They stated that they undertook courses of action to cope with situational constraints. They said they seldom chose between two courses of action. They did not seek for an optimal choice among several options which could lead them to lose control of the situation. Rather, they identified a course of action that they had enough time to implement and which they considered could be satisfactory for them. They were seen to use their experience to identify a workable course of action as the first action they considered. To check if it would work, they could mentally simulate the possible effects on the situation and themselves.

This model was developed in relation to the paradigm called Naturalistic Decision-Making (NDM, e.g., Zsambok, 1997). Different NDM models have been developed. The main models concern cognitive control (Hoc & Amalberti, 2007; Rasmussen, 1983) and Recognition-Primed Decision (Klein, 1997). Rasmussen (1983) focuses on skill-based, ruled-based, and knowledge-behaviours placed on a decision ladder in relation to the use of heuristics paths. Hoc and Amalberti (2007) characterized the cognitive control modes from their level of abstraction (symbolic, subsymbolic) and their origin (internal-anticipative, external-reactive); metacognition enables the cognitive control to be organized using these dimensions to ensure control of the situation. All these naturalistic decision-making models reached similar conclusions. People did not often generate or compare options. Rather, they committed themselves to a course of action from a number of possible options (Klein, 2008).

Althought the results of Klein et al. (1986) did not show that experts did not compare options, they nevertheless highlighted the unlikely use of options comparisons strategies in less than one minute. The fire-service ground commanders reported alternative strategies. They used situational recognition in the same way chess masters did (De Groot, 1965). When they recognized that the situation was typical, meaning it was associated with a similar situation stored in memory, they knew what to do. Each stored situation was connected to a specific course of action. They adapted the course of action to the specific situational constraints. When they could not link the current situation to one stored in memory, they continued to analyse the situation until they could find similarities between the current situation and a situation stored in memory. If, and only if, they considered that the course of action could achieve negative outcomes, they identified another course of action. Recognition strategies appeared efficient. The experts usually identified only one option, which they took. If they had assesses the outcomes of many options, the fire would probably have become out of control before they had made any decision.

The RPD model addresses three assertions. The first one is that experts can use their experience to generate a possible course of action as the first course of action they consider. The second one is that time pressure does not affect the performance of the decision makers, because they use their experience and pattern matching. They recognize the situation in which they are committed and consider a typical course of action. Pattern matching is central to the RPD model and enables experts to cope with time pressure and implement efficient decisions. A third assertion is that experts can generate an option without comparing different options. Decision-making is based more on recognition procedures than "hypervigilant" procedures, concerned with a checklist. Recognition Primed-Decision is an intuitive process that enables the expert to know what course of action will work. Decision-makers scan only the information needed to make a satisfisfactory decision and rapidly assess the situation.

The RPD model presents three levels: (a) simple match; (b) diagnosis of the situation; and (c) evaluation of a course of action (Fig. 1). The first one refers to rapid situation recognition and the implementation of the typical action corresponding to a typical situation. The second one is used when the situation cannot be recognized rapidly, because some information is missing or the situation understanding presents inconsistencies and anomalies. The expert takes more time to diagnose the situation in order to recognize it as typical; once he/she recognizes it, he/she undertakes the typical action, after adapting it to the current situation. The third level refers to an evaluation of the workability of a course of action. If the expert simulates that it could work, he/she implements it, if not, he/she changes it (Klein, 1997; Klein et al., 1986, see Figure 1).

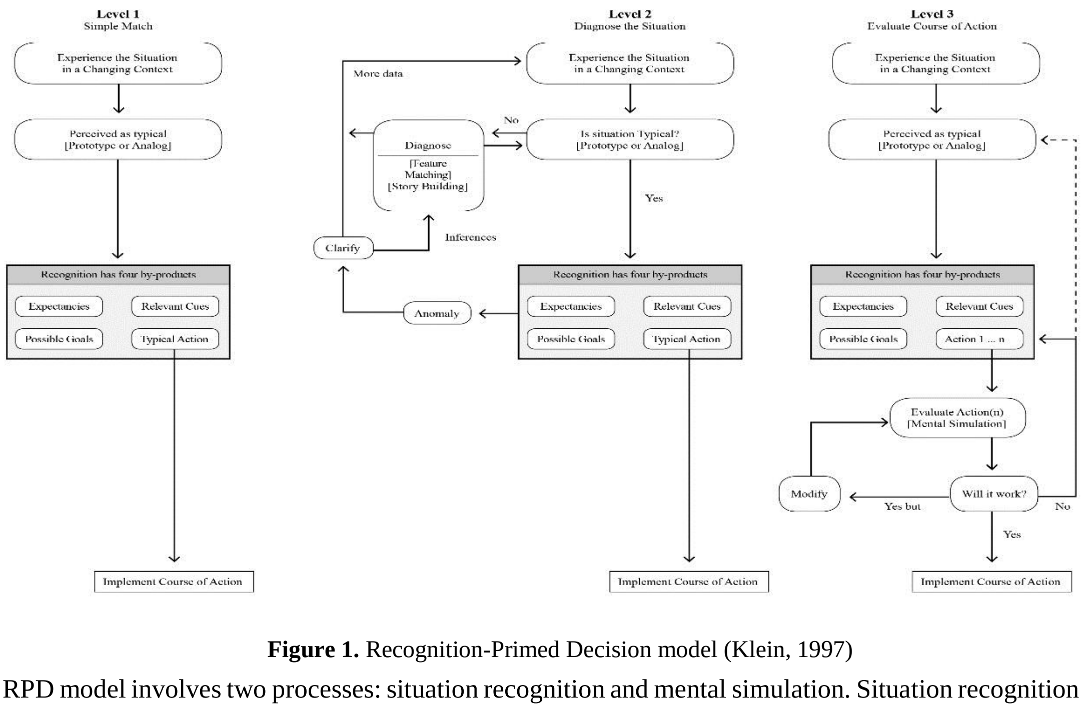


Figure 1. Recognition-Primed Decision model (Klein, 1997)

The RPD model involves two processes: situation recognition and mental simulation. Situation recognition enables the expert to classify the situation as familiar versus unfamiliar or atypical. Recognition is achieved from four by-products: (a) possible goals (e.g., reduce the fire); (b) relevant cues (e.g., how far the fire can spread); (c) expectations (e.g., fire-service ground commanders expect to be able to put out the fire); and (d) typical action to be implemented (e.g., send the troops towards the first floor). These by-products are used to make sense of the situation. Fire-service ground commanders also use a repertoire of patterns built up from experience and their knowledge of other situations. Patterns are the basis of intuition and enable experts to assess what type of situation they are facing (Klein, 2009). Mental simulation provides the means to assess the situation in time-pressure settings that could prevent the use of analytical strategies of decision-making and problem solving (Klein & Crandall, 1995). It is carried out using information perceived within the coupling between the individual and his/her environment. The RPD model is concerned with intuition and analysis. Situation recognition or pattern matching is intuitive and diagnosis and mental simulation are deliberate and analytical. Recognition strategies are adaptive (Klein, 1997). They could be considered to be a source of power (Klein, 1997, 2003).

The RPD model accounts for slow decisions such as those made in operating rooms (e.g., Baber, Chen, & Howes, 2015), as well as rapid decisions such as those taken by fire-ground fire-fighters (e.g., Klein, 1997) and elite athletes (e.g., Macquet, 2009). Studies from different areas conducted worldwide have provided a growing body of empirical support for the RPD model (for example, see NDM 2015 conference proceedings, Klein, 2015). Despite this considerable attention and support, the model does not account for important features of Naturalistic Decision-Making. Firstly, McLennan and Omodei (1996) argue that in dynamic environments like ball sports, time pressure is so high that it does not allow people to engage in the re-framing process and get time for reassessment, as claimed by the RPD model. Decisions must be made almost instantaneously. Baber et al. (2015) claim that (very) rapid decision-making is more a matter of filtering (perceiving affordances) rather than framing (developing schemata and applying them). To Baber et al. (2015), perception-action couplings as described by Gibson (1969) would support selection action at level 1. They wonder whether it is possible to focus less on the structure of schemata and more on the way in which perception is tuned to the environment.

Secondly, the RPD model implies that the decision-making process starts when the decision-maker assesses the situation in progress and he/she makes a decision on the course of action. However, a small amount of time occurs between the moment the decision-maker first becomes aware that a situation has occurred and the point at which he/she is actually faced with this situation (e.g., several minutes elapse between the fire alarm and arrival at the fire site). This raises the possibility that decision-makers simulate possible courses of action from the moment they first become aware of the situation they will be required to deal with, before they have access to situational data. Thus, they anticipate the potential situation evolution and plan decisions before the situation develops. This makes it possible to undertake an appropriate course of action in time. Klein's model focuses on the decisions made during the course of action. As McLennan and Omodei (1996) suggest, in support of Weick's (1995) studies, the RPD model gives insufficient importance to the human continual engagement in sense-making from available and anticipated information. From a practical perspective, it does not account for planned actions. In order to highlight the continuing role of sense-making on decision-making, Klein and colleagues developed the Data-Frame theory (e.g., Klein, Philipps, Rall, & Peluso, 2007). To further explain how experts adapt their plans while executing them in a situation with ill-defined and conflicting goals, Klein (2007) conceived the flexecution model. Revisiting the RPD model enables us to keep the three initial functions of decision-making, underlines the continual engagement in sense-making and accounts for planned actions.

We propose firstly, that planned actions have a substantial impact on the rapidity and appropriateness of the action undertaken by the decision-maker, and secondly, that the adaptive value of such situation activation depends on the extent and appropriateness of the anticipation process. These considerations suggest that the RPD model needs to be revisited to take into account the mental processes that precede the situation and are mobilized during the situation development. It is hypothesized that in high time pressure situations, decision-makers anticipate the potential situation development and possible action prior to the activation of the pattern-matching process on encountering the actual situation development. From a theoretical perspective, we suggest that a typical situation and corresponding typical action are anticipated and therefore more rapidly activated in relation to the current situation development. The decision-maker compares the pattern of the activated situation and that of the current situation. If they are similar, he/she undertakes the planned action; if not, he/she takes different action. He/she can mentally simulate the situation evolution in order to check whether the current situation is similar to the typical situation corresponding to the planned action. If they are similar, he/she undertakes the planed action; if not, he/she takes different action. As predicted by the RPD model, he/she can also mentally simulate whether the action will be satisfactory. Thus, we propose revisiting the RPD model to account for planed actions as well as (very) rapid decisions. This revisiting is based on results provided from fire-fighting, elite sports, ship handling and medicine.

EMPIRICAL SUPPORT FOR THE REVISITED RPD MODEL

The three previously described assertions can be extended by two new assertions. A fourth assertion is that time pressure does not impair the performance of experts because they anticipate the potential situation development and plan a decision before the situation develops. Then, they use pattern matching to assess whether the situation has developed as expected. If so, they implement the decision; if not, they adapt the decision. McLennan and Omodei (1996) studied decision-making by fire-fighters and soccer referees to explain the extent to which experts engage in mental simulation of possible situations that might occur and possible courses of action prior to encountering the actual incident. In 90% of instances, referees recalled mental simulation of already-considered possibilities on the basis of what was likely to occur based on their experience of the way the game was typically played. Mental simulation enabled them to make very rapid decisions and apply the appropriate rule very rapidly (less than one second). Fire-fighters also engaged in mental simulation when driving to the fire site. They built a picture of the incident from the fire call, based on their knowledge of the way fire spreads, the kind of site (e.g., house, factory), weather conditions (wind direction and strength), the likely traffic situation at the site (pedestrian and vehicle), time and water hydrant location. These pre-conditions were used to anticipate the priority operations to be implemented on site (e.g., rescue people). On arrival at the fire scene, fire-fighters compared the existing situation with the anticipated situation. If anticipated events and actual events matched, they implemented the planned action (i.e., priority operations), if not, they examined different courses of action and possible consequences. Results suggest that when the situation occurs, experts have already used available knowledge to anticipate likely situations and already simulated possible courses of action. They compare what they see with what was simulated mentally. Results suggest that rapid decisions and anticipation are a matter of both mental simulation of possible play and fire outcomes and recognition of a typical situation. Thus, the RPD model should be revisited to account for a planning component.

Macquet and colleagues showed this anticipated process in elite sports. When aiming at blocking an attack, volleyball players reported preparing more for a specific situation development and planning a consequent action, while at the same time being attuned to another situation development and ready to change decision (Macquet, 2009). A world champion orienteer recalled that he prepared for a specific navigation option from the map and was ready to adapt it when faced with the terrain and its possibilities (Macquet, Eccles, & Barraux, 2012). Macquet and Kragba (2015) showed that basketball players systematically anticipated a pattern of players' coordination so that teammates could predict their movement and achieve positive outcomes. Anticipation was based on specific situation development related to the playbook that defined what each player was required to do. Results showed that before undertaking the planned play, players compared information about players' placement and movement on the court, players' competencies, tendencies and roles, and ball trajectory to corresponding information associated with the anticipated situation development. Recognition consisted of reconsidering the anticipated frame created from anticipating possible situation development and enriching it with information that became available in the course of the situation development. Players checked whether the situation developed as expected from the typical situation anticipated using the playbook. Results showed that as the situation developed, experts had already used available knowledge to anticipate likely situations and subsequent actions. They also showed that players anticipated situations and planned subsequent decisions. They suggested that players used situation recognition and mental simulation to make sense of situations. Belling, Suss and Ward (2015) showed that under time pressure, football players generated task-relevant options that were positively and strongly related to skill at generating and accurately rating the criterion of best option. This suggests a greater tendency of use long term memory-working memory-type strategy and recognition processes. Macquet and Lacouchie (2015) and Macquet and Fleurance (2007) showed that elite athletes prioritized the use of their favourite techniques. Situations in which a favourite technique could be used arose in one of two ways: either the situation arose naturally, or athletes decided to manipulate the situation to create the conditions required to implement their favourite technique; then they carried the technique out. In the latter event, experts made a first decision to create appropriate conditions for the effective implementation of their favourite technique (i.e., a second decision). In other words, experts did not change their decision to implement their favourite technique. Rather, they changed the situation to make it possible to implement their favourite technique. For example, a badminton player sent the shuttle behind the net to force his opponent to lift the shuttle and enable him to smash and win the rally. This changing the situation was used in 12% of all situations recalled by badminton players (Macquet & Fleurance, 2007). In judo, this two-step decision was seen to enable judoka to win matches (Macquet & Lacouchie, 2015).

Chauvin and Lardjane (2008) showed similar strategies in ship handling. To avoid collisions at sea, experts used two kinds of rules: a stereotype consisting of adapting their handling to crossing vessels (89.49% of all situations recalled) and an alternative involving forcing the target to change direction and reduce speed, in order to impose on the other vessel the course it should take and thus maintain his/her own vessel direction and speed (10.51% of all situations recalled).

To provide empirical support to the revisited RPD model, Pellegrin, Gaudin, Bonnardel and Chaudet, (2010) re-explored 219 data events resulting from the management of an outbreak alert by an expert team in simulated conditions. Of all the actions elicited, 4% had a "force the situation" function. Before being able to identify the precise pathogenic cause of the outbreak, the team decided on specific public health countermeasures such as patient isolation or the use of boiled or bottled water, with the aim of containing the outbreak.

A fifth assertion of the revisited RPD model is that time pressure does not affect experts' pattern matching, because experts are attuned to relevant cues. Perception is guided by relevant cues in relation to experts' goals, expectancies, knowledge and competencies. Perception is active (Gibson, 1969) and relevant cues are perceived directly. Although Gibson's theory is useful to describe critical cues that are perceived directly, Neisser (1976) stressed that it does not satisfactorily explain what is in the perceivers' heads. The Neisser (1976) schema theory and perception action cycle highlights that schemata guide external and internal information search and specify the information to be perceived and that to be ignored. Relevant cues are used to build a big picture of the situation. In fast sports, Macquet (2009), Macquet and Fleurance (2007), and Bossard, De Keurlaere, Cormier, Pasco and Kermarrec (2010) showed that relevant cues related to the ball or shuttle trajectory (65% in Macquet's study, 2009) and players' actions (100% in Macquet's study, 2009). Players' experience of the game guided perception towards what they needed to notice within the game. Results suggested that perception was guided by experience and schemata stored in memory in relation to the possible game development. Although Macquet (2009), Macquet and Fleurance (2007) and Bossard et al. (2010) failed to provide results demonstrating that experts did not use schemata, they nevertheless highlighted the likely use of information filtering processes anchored based on schemata and experience. Experts in sport reported alternative strategies. They used filtering strategies, in the same way Gibson (1969) stressed in explaining direct perception. Filtering was based on experience and subsequent schemata. As a judoka recalled (Macquet & Lacouchie, 2015), a novice would have noticed different cues from an expert in the same situation (opponent's agitation versus athlete's calm).

REVISITED RPD MODEL

We recently revisited the RPD model by connecting the three levels of the RPD model and adding two additional components: anticipate a possible situation development and force the situation (Macquet, 2016; see Figure 2). Connecting the three levels accounts for dynamic decision-making. Revisiting the model with planned decisions, in addition to decisions made in the course of action, is important for the RPD model because temporality largely determines the course of the action adopted. In time pressure situations, decision-makers often spend more time anticipating what could happen and planning an action than waiting for the situation to develop. Planned actions are based on anticipated events. The purpose of anticipation is to assess a possible situation development when relevant cues are not yet available and time pressure is high. Anticipating enables adaptation to uncertainty about situation development and a short time for undertaking action. Anticipation is a function based on knowledge about possible situation development in relation to relevant and available cues and possible goals and subsequent planned action. McLennan and Omodei (1996) showed that when fire-fighters were en route to the fire site, they anticipated the situation at the fire site from the fire call and planned priority actions. Once they arrived, they checked whether the anticipated situation matched the current situation. This option generation process comes before the situation develops (e.g., sports) or before the decision-maker is at the incident site (e.g., fire-fighters).

We have augmented the RPD model by adding a function referred to as forcing the situation. Forcing the situation is initiated in response to the discrepancy between the current situation and situation required to implement a specific course of action. Forcing the situation enables adaptation to reduce this discrepancy and adoption of an efficient course of action. Decision-making may be one-step or two-step. Either the decision-maker aims to force the situation features to make them match those required for the planned action, and then continues with this action when the situation features fit, or the decision-maker implements an initial action aimed at forcing the situation features to fit those of the planned action and then undertakes the planned action (see Figure 2).

The revisited RPD model is able to account for actions with one of two aims: to adapt the action to the context or to adapt the context to the action. Adapting the action to the context is a one-step decision. Adapting the context to the action is a two-step decision. To adapt the action to the context, the decision-maker compares whether the situation is similar to a situation stored in memory and relates it to a specific course of action. Then, either the typical action is implemented directly, or the decision-maker mentally simulates whether it will work (function 3). If it might not work, the decision-maker adapts the typical action. When the decision-maker anticipates a course of action, he/she mentally simulates whether the situation evolves as expected from anticipated events. To adapt the context to the action, the decision-maker implements an action that will adapt the situation features to match those of the situation corresponding to the planned action. When the actual features match the anticipated ones, either the decision-maker undertakes the second step decision (Macquet & Fleurance, 2007; Macquet & Lacouchie, 2015), or he/she continues with the action (Chauvin & Lardjane, 2008; Pellegrin et al., 2010). The expected result of the planned action is firstly re-evaluated by the expert, leading secondly to him/her undertaking another action or pursuing the same action to move the situation towards the desired states.

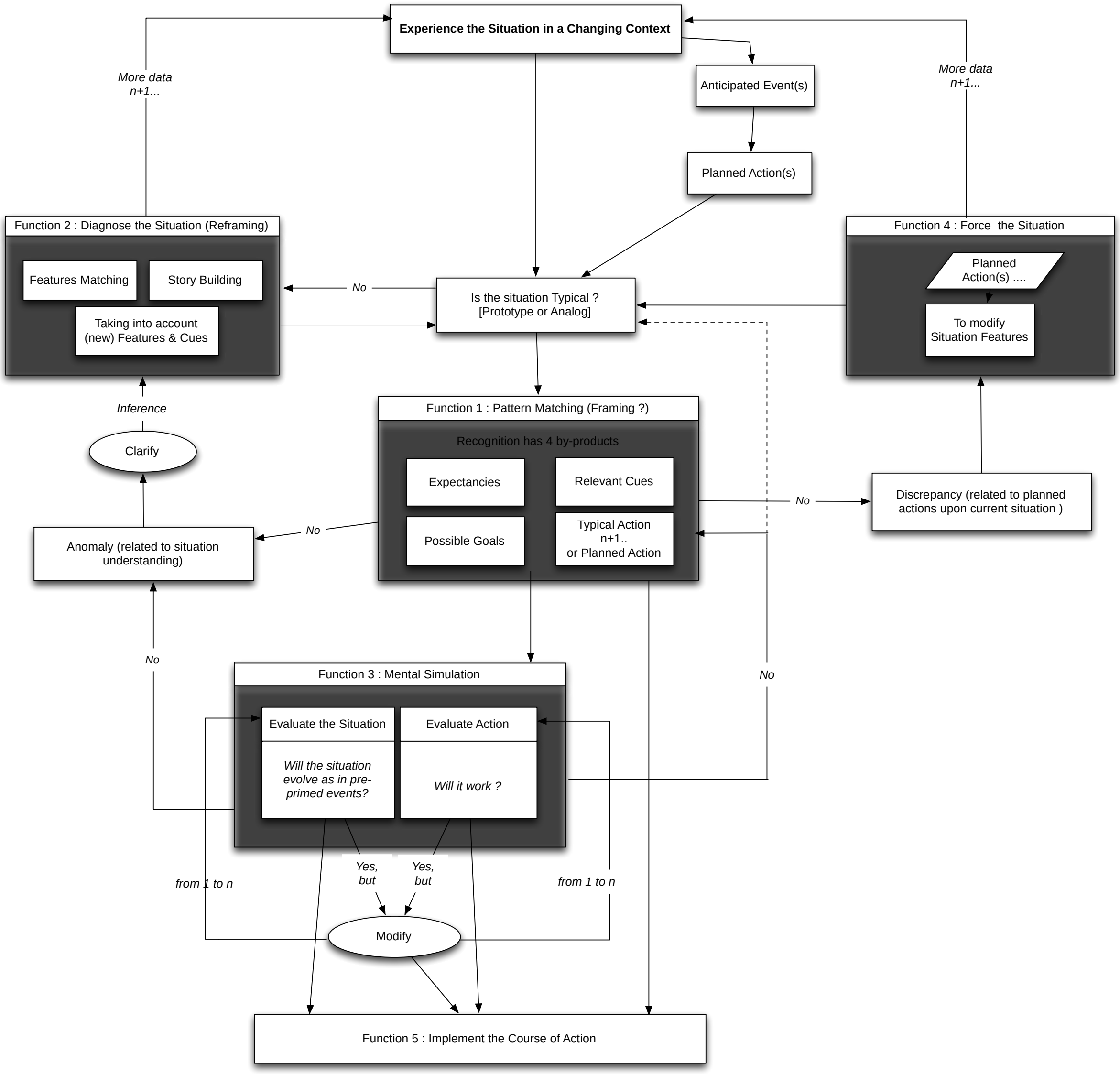


Figure 2. The revisited RPD model

CONCLUSION

The RPD model has evolved during the past thirty years, addressing the issues of situation diagnosis, mental simulation and planned actions, and becoming more dynamic by joining the three levels into a single element. At the same time, it has received empirical support from research conducted in different areas. The revisited RPD model accounts for (rapid) decisions made in dynamic environments. It connects immediacy, temporality and anticipation. This model adds to other macro-cognition models and theories, such as the Data-Frame theory of sensemaking (Klein et al., 2007) and flexecution model (Klein, 2007) related to experts' adaptive skills in complex settings. The continued study of experts' decision-making in dynamic situations will improve our understanding of cognitive processes and performance and enables the NDM community to gain a richer understanding of the way experts use their experience to make decisions in dynamic settings including stress and uncertainty.

REFERENCES

Baber, C., Chen, X., & Howes, A. (2015). *(Very) rapid decision-making: Framing or filtering?* Communication presented at the International Conference on Naturalistic Decision-Making, Mc-Lean, VA.

Belling, P., Suss, J., & Ward, P. (2015). The effects of time constraint on anticipation, decision-making, and option generation in complex and dynamic environments. *Cognition, Technology and Work, 17*, 355-366.

Bossard, C., De Keukelaere, C., Cormier, J., Pasco, D., & Kermarrec, G. (2010). L'activité décisionnelle en phase de contre-attaque en hockey sur glace [Decision-making activity during turnover phases in ice hockey]. *Activités, 7*(1), 42-61.

Chauvin, C., & Lardjane, S. (2008). Decision making and strategies in an interaction situation: Collision avoidance at sea. *Transportation Research Part F: Traffic Psychology and Behavior, 11*(4), 259-269.

De Groot, A. D. (1965). *Thought and choice in chess*. New York: Mouton.

Hoc, J. -M., & Amalberti, R. (2007). Cognitive control dynamics for reaching a satisfying performance in complex dynamic situations. *Journal of Cognitive Engineering and Decision Making*, *1*, 22–55.

Gibson, J. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Klein, G. A. (1997). The Recognition-Primed Decision (RPD) model: looking back, looking forward. In C. E. Zsambok, & G. A., *Naturalistic decision making* (pp. 285-292). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Klein, G. A. (2003). *The power of intuition.* New York NY: Random House Inc.

Klein, G. A. (2008). Naturalistic Decision-Making. *Human Factors, 50*(3), 456-460. DOI: 10.1518/001872008X288385

Klein, G. A. (2007). Flexecution, part 2: Understanding and supporting flexible execution. *IEEE Intelligent systems, 22*(6), 108-112.

Klein, G. A., Calderwood, R., Clinton-Cirocco, A. (1986). *Rapid decision making on the fireground*. Communication présented at The Human Factors and Ergonomic Society 30th annual meeting 1:576–580.

Klein, G. A., & Crandall, B. W. (1995). The role of mental simulation in problem solving and decision making. In P. Hancock, J. Flach, J. Caird, & K. Vicente (Eds.) *Local applications of the ecological approach to human-machine systems* (pp. 324-358). Hillsdale, N. J.: Lawrence Erlbaum Associates, Publishers.

Klein. G. A., Philipps, J. K., Rall R. L., & Peluso, D. A. (2007). A data frame theory of sense-making. In: R. R. Hoffman (Ed.), *Expertise out of context*, (pp.113-155). New York, NY: Lawrence Erlbaum Associates.

Klein, G. L. (2015). Proceedings of the International Naturalistic Decision-Making conference, Mc Lean, VA.

Macquet, A. -C. (2009) Recognition within the decision-making process: A case study of expert volleyball players. *Journal of Applied Sport Psychology, 21*, 64-79. doi: 10.1080/10413200802575759.

Macquet, A. –C. (2016). *De la compréhension de la situation à la distribution des informations : la prise de decision en sport de haut niveau [From situation understanding to information distribution: Decision-making in high level sports].* Unpublished thesis for the ability to conduct researches, Université de Bretagne Sud - INSEP

Macquet, A. -C., Eccles D. W., & Barraux, E. (2012) .What makes an orienteer an expert? A case study of a highly elite orienteer's concerns in the course of competition. *Journal of Sport Sciences, 30,* 91-99. doi: 10.1080/02640414.2011.617774

Macquet A. –C., & Fleurance, P. (2007) Naturalistic decision-making in expert badminton players. *Ergonomics, 50*, 1433-1450. doi: 10.1080./00140130701393452.

Macquet, A. -C., & Kragba, K. (2015).What makes basketball players continue with the planned play or change it? A case study of the relationships between sense-making and decision-making. *Cognition, Technology and Work, 17*(3), 345-353. Doi: 10.1007/s10111-015-0332-4

Macquet, A. -C., & Lacouchie, H. (2015). *What is the story behind the story? Two case studies of decision-making under stress*. Communication presented at the International Conference on Naturalistic Decision-Making, McLean,Va.

McLennan, J., & Omodei, M. M. (1996). The role of prepriming in recognition-primed decision-making, *Perceptual and Motor Skills,82*, 1059-1069.

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive*

Psychology. Freeman: San Francisco.

Pellegrin, L., Gaudin C., Bonnardel N., & Chaudet H. (2010). Collaborative activities during an outbreak early warning assisted by a decision-supported system (ASTER), International Journal of Human and Computer Interaction, 26(2-3), 262-267.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs and symbols and other disctinctions in human performance models. *IEEE Transaction Systems, Man, and Cybernetics, 13*¸257-266.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, Ca: Sage Publications.

Zsambok, C. E. (1997). Naturalistic Decision-Making: Where are we now? In C. E. Zsambok, & G. A., Klein (Eds.), *Naturalistic decision making* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers

Leaders and Ethical Decision Making with Autonomous Systems

Yvonne R. MASAKOWSKI, Ph.D. & Thomas E. CREELY, Ph.D.
U.S. Naval War College

ABSTRACT
The number and range of military operations has expanded on a global scale requring the development of teams of humans and autonomous unmanned systems. Indeed, UAVs, UUVs, etc. have become critical components of the military force. Advances in artificial intelligence and computing power have increased the capability and complexity of information management that shapes situational awareness (SA) and decision making. Given these advanced technologies, today's military leader must understand the ethical and moral consequences of their decisions. This presentation will focus on the importance of validating information received by autonomous systems within the context of the military leader's ethical judgment.

KEYWORDS
Decision Making; Military; Command and Control.

INTRODUCTION

The 21st-century is replete with global security challenges, including counter-terrorism, piracy, human trafficking, immigration, as well as humanitarian and disaster relief crises. This increased demand for military support throughout the world mandates the need for utilizing autonomous, unmanned systems as part of the military force. Advances in artificial intelligence (AI), computing power and cyber technologies provide a means of integrating information on a large scale to ensure global security. Intelligence preparation of the battlespace (IPB) is the primary step for framing and preparing an effective mission plan. These autonomous, unmanned systems provide a plethora of information and enhance a military leader's situational awareness (SA) required for military operations. Indeed, IPB is a critical component of mission planning and essential for mission success. Autonomous, unmanned systems provide a capability for reconnaissance and surveillance necessary for characterizing the physical and security operational environment.

Autonomous unmanned systems gather data from a variety of perspectives whether aerial, undersea and/or above ground. Once collected, this data is further analyzed and distributed to the command and control team to contribute to a shared understanding of the environment. While this information elevates everyone's SA, it is the military leader who must form their decisions based on the integrity of the information received.

The complexity of military operations is amplified by the nature of each mission, whether it is humanitarian, counter-terrorism, or a combat mission. The military mission planning process extends to all commanders, regardless of which service they serve. For the Army officer, there is a need to understand both the physical and the security environment. There is also a need to understand cultural aspects of the environment as cultural factors may negatively impact operational effectiveness. Therefore, it is insufficient to rely exclusively on autonomous, unmanned systems as they merely collect sensor data and target information. These systems are not designed to integrate cultural or socio-political aspects that play a role in command decision making. Autonomous systems may be programmed to capture the topographical information in an accurate manner, however, such systems are not programmed to integrate cultural information that might negatively impact mission outcome. Thus, the military leader must evaluate the information within the context of the operational environment.

There is currently a great deal of discussion surrounding the issue of weaponizing autonomous systems. Advances in AI have equipped autonomous systems with the capability of sensing, evaluating a target and making decisions. Furthermore, these systems are also capable of operating independently, as well as in collaborative teams. Although these systems have been evolving at an advanced rate, the question is whether the consequences for their actions far outweigh the benefits of utilizing them in this capacity. Specifically, there are potential consequences for shifting the responsibility for human decision making to the weaponized autonomous system that raise serious concerns for future warfare and humanity with regard to the increased risk to innocent civilians.

This is the great unknown territory that raises serious ethical issues surrounding weaponized autonomous unmanned systems.

Agencies such as DARPA and ONR have demonstrated that these systems can be programmed to conduct independent missions in an autonomous manner. While there are policies and directives written regarding the design of autonomous systems (DoD, 2012), there is no way of ensuring how society will view the consequences of such systems. We assume that these systems will be used only in foreign military operations, however, we must also consider whether these systems will be used to defend our security on a national and global level.

This raises numerous ethical issues regarding the risks associated with transferring human decision making to an autonomous AI warfighter. For tasks that are routine and harmless, such as rotobic vacuums, there are minimal risks to the human. However, the transformation of autonomous unmanned systems into weaponized systems raises significant risks for the general population with regard to their independent decision making capability. However, for the task of weaponizing autonomous systems, the question remains should we transfer this authority to the autonomous system?

Weaponizing autonomous systems with decision making capabilities raises ethical dilemmas for future warfare. We contend that now is the time to consider the ethical and moral consequences of using such systems in a military operational environment. What if the combat zone is an urban area with a large civilian population still in residence? What are the ethical and moral considerations that we should consider? How might we minimize the risk to the civilian population? What are the potential repercussions of our failure to do so?

Automobile designers, such as Tesla, have been integrating artificial intelligent systems into the design of self-driving automobiles. Tesla recently had its first fatality as a result of their self-driving car due to a “technical failure” of the automatic braking system (Granville, 2017). This accident suggests that there is still work to be done to improve the reliability of autonomous systems. Sheridan (1992) suggests that the human can monitor and maintain a level of control over the system. There are levels of automation that provide decision support to the operator and facilitate a reduction in operator workload. Combat systems designed with embedded artificial intelligent agent architecture provide a means of reducing operator workload and supporting the operator’s ability to manage big data and information management. Porat, Oron-Gilad, Rottem-Hovev & Silbiger (2016) examined the capacity of human operators to manage a number of unmanned systems (up to 15 Unmanned systems) where operator experience and mission complexity played a role in their performance.

Cummings & Guerlain (2007) have also shown that the human’s role as a supervisor of an autonomous system maintains a level of control over weapon systems. Their results indicate that the complexity of the autonomous unmanned system may require less human intervention. However, one could argue that the need for rapid decision making under conditions of uncertainty and high risk may yield a different outcome as a result of human stress. Specifically, the ability to manage numerous UAVs is a function of the complexity of the operational environment (combat setting) vs normal test conditions (Goodrich and Cummings, 2014). These results suggest that there is a need to address the likelihood of unintended consequences related to weaponizing autonomous systems for combat operations.

We anticipate that advances in AI will enhance the development of autonomous systems in the future. However, we posit that there are other challenges involved in the management of autonomous systems. Weaponizing autonomous systems may have a negative impact on the warfighter and society itself.

For the warfighter, there may be ethical consequences related to the fact that once released for the mission, autonomous weaponized systems will be unable to be recalled. As the commander, there may not be an opportunity to reprogram or to override the decision making of the autonomous system during a combat mission. As a result, there is an increased risk for civilian casualities related to the actions of the weaponized AUV that may cause death to innocent civilians and/or moral injury to the soldier, a new type of post traumatic stress syndrome The question is, how will we design these weaponized autonomous systems to ensure that humans can override their decision capabilities? For the 21st-century warfighter, weaponized autonomous systems equipped with embedded artificial intelligence represents a significant change in the character of warfare. Just as the introduction of the tank, airplane and aircraft carrier to the battlespace, weaponized autonomous systems represent a force multiplier on the battlefield that may provide the warfighter with a decision advantage against adversaries. In contrast to earlier technologies, however, there is an even greater likelihood that these systems will present a clear and present danger to innocent civilians and to society as a whole. Like the advent of the aircraft carrier, weaponized autonomous systems will transform warfare and may well also have a negative impact on individual freedoms, privacy, and personal security in ways that we have yet to anticipate.

For society, there are considerable benefits to the development of autonomous systems that facilitate a higher quality in medical care and daily living. Smart homes and digital cities are being designed to provide a higher quality of life with increased medical attention and security. However, there are trade-offs relaled to ethical and privacy issues regarding the intrusion of personal freedom Masakowski, Smythe & Creely (2016) examined the impact of ambient intelligence technologies on society and warfare. They concluded that although advances in artificial intelligence and autonomous systems are inevitable; there are potentially significant risks associated with personal freedom, individual privacy, and for national security.

We anticipate that as the level of sophistication increases in the design of autonomous systems, there will be an increased level of capabilities, such as advances in AI, increased memory capacity, composite materials, and increased sensory modelling capabilities, that will help to make the autonomous system more independent, agile and more independent than ever before. These capacities, wherein the system models human behaviour and cognitive processes, will provide systems with greater independence and decision making capabilities. Thus, we must anticipate the time when such systems will not respond in a directive manner to a human supervisor. We need to consider how to manage systems that will use reason and understanding to respond to the directive put forth by the human to them. Designing AI systems from a human biological perspective will help to develop systems that will be capable of functioning in a human-like manner. Autonomous systems will be transformed from being an extension of human decision making in the battlespace to become an independent agent capable of sensing, targeting, and killing independent of their human supervisor.

AI autonomous systems are capable of logical thinking and learning and are also capable of developing strategies and defense plans that will facilitate decision making that is independent of the human. Given the realm of possibility of these designs at this time and the range of applications, both military and civilian, we need to consider how best to integrate ethics and moral reasoning into the design of these systems. We must also consider the manner in which these systems will be employed both in the military environment and in the realm of daily life.

Self-defense capabilities for autonomous systems in the future is a reality. According to military doctrine, autonomous unmanned systems will be designed to defend themselves in the coming years (DoD, 2012). Specifically, autonomous systems such as the Predator are able to detect and respond to threats. The combination of AI and self-awareness capabilities will facilitate the systems' ability to transform from merely being programmed machines capable of detecting and targeting potential threats to detect, target, and decide to respond to a target independent of the human. A critical component of this capability is the development of proprioceptive awareness of self in relation to a potential target. As humans, we are biologically hard-wired for proprioception that enables us to be aware of ourselves, our posture and our position relative to people and objects in our environment. Autonomous systems, embedded with AI will also be designed to be self-aware and learn to understand their position in the environment as well as use AI cognitive models to inform the characteristics of potential targets in their environment. Thus, systems designed with self-awareness will transform and enhance a system's capability as it gains perspective of self and is capable of evaluating a variety of options and potential outcomes (Dutt & Taher-Nejad, 2016; Dobbyn & Stuart, 2003). While these advances sound promising, they raise several ethical and moral questions regarding their impact on warfighting and society.

For example, how will humans manage autonomous systems with greater levels of awareness and decision making capabilities? What happens if the system makes an error and targets innocent civilians in the combat environment? How can humans trust the autonomous system to make correct, rational decisions? Will the human maintain their right of supervision on the networks and swarms of autonomous systems? For now, the human retains supervisory control. However, that role will change as advanced technology evolves.

We contend that now is the time to have a discussion on the ethical decision making capabilities of autonomous unmanned systems.

We must address how these systems will impact our society. How do we constrain their operational capabilities? Even though the US has military doctrine to define some of the limitations and constraints in the development and use of weaponized autonomous systems; we cannot ignore the potential for adversarial use of such weapons against us. How will society accept the use of autonomous, unmanned systems as part of an overall surveillance system in society?

We are facing a time when autonomous unmanned systems may be used to set up security barriers regardless of the battlefield or home port. The potential for overriding the barriers of security and ignoring the ethical consequences of such actions should be taken as a warning sign (Giordano, Kulkarni & Farewell, 2014). Now is the time to define the limits of these systems. There are known and unforeseen consequences for developing autonomous systems that will control our global security environments.

Today, there are trust issues related to autonomous unmanned systems. One can envision a future in which autonomous unmanned systems will be used for much more than surveillance. Autonomy in unmanned systems, while presenting advantages for the military commander with regard to information gathering, also presents dangers in military decision making in the future. As the design of AI systems become more sophisticated, there is a need to consider the societal impact of these systems .

Imagine a world in which networks of autonomous systems monitor the security of our towns, homes, work places, and airports, etc. We might find some level of comfort as these systems have the capability to monitor public spaces for our security. However, we might equally find such systems invasive to our private lives. Intrusions on our privacy may impede our social and political freedoms in ways we have not experienced before. The constant presence of surveillance would constrain free speech in our democracy as data would be collected in a continuous manner. The presence of systems that monitor our speech and behaviour in a continuous manner will have negative ethical consequences for society as a whole. Our relationships and conduct will be scrutinized in ways unforeseen and impede individual freedom of expression, and potentially negatively impact how we innovate and self-determine our individual destiny.

More importantly, how do we defend our systems against potential adversaries who might reprogram these to be used against us? What safeguards are we developing to ensure this?

There is an ethical dilemma in this regard. For the military leader and decision maker, there is a clear advantage to using a network or swarm of autonomous systems to gather information across the battlespace. There are critical implications for society and for individual rights to privacy that will be overtaken by advances and implementation of these technologies in the future.

Autonomous systems equipped with an embedded AI intelligence may provide the decision advantage for the military commander; however, there is a need to consider the ethical consequences for these technologies. Society, philosophers, and ethicists must argue to what extent should these advanced technologies be integrated with the whole of society. Similar to the Atom Bomb, the development of lethal, weaponized autonomous systems may have negative, long term unanticipated consequences.

The dynamic relationship between autonomous systems and humans will shift over time as these systems advance in their development. The ethical concerns will be weighed within the context of the benefits for each nation's security. Thus, military leaders must remain attuned to each nation's position on the utility and governance of autonomous systems. There are nations which may not be limited by our nation's ethical constraints. We must balance our nation's security goals and defend against our adversaries' potential use of weaponized autonomous systems as a tool for combat vs merely data collection. This will take a level of commitment from each nation, as well as their level of capabilities that may be integrated into the overall mission plan.

Autonomous unmanned systems present a significant challenge to military command and control, especially to a leader's decision making. These systems can enter combat zones and gather information faster, safer and more accurately than a human. Their capacity for pattern recognition and data storage/retrieval exceeds that of the human decision maker. The question is, how will our military leaders best utilize these capabilities while ensuring that these systems do no harm? How can we best prepare our military leaders to make decisions in these situations? The real question is how will our adversaries use this technology against us in the future? Will we be ready? Thus, it is our task to find ways to prepare our leaders to meet these challenges successfully.

A commander's ethical leadership skills in an age of lethal autonomous technology falls short for meeting challenges. There are significant limitations to a human's capacity for processing vast amounts of data which do not constrain AI and autonomous systems. Thus, given the exponential rate of development of these systems, we can anticipate the evolution of systems to surpass human capabilities in the near future. Moore's Law places the human at a distinct disadvantage. It is incumbent on leaders to evaluate advances in these technologies in terms of their utility in military defense as well as achieve an understanding of their impact within the context of ethical requirements for society.

Ethics of technology, a recent discipline of study and scholarship, is unfamiliar to most commanders. With a gap between technology and its ethical implications, military leaders need an understanding of the philosophy of technology.

Today's commander must be flexible to engage conventional and asymmetrical warfare in the Third Offset Strategy landscape. In addition to autonomous weapons, cyber, neuro, nano, ambient intelligence, and biotechnologies will impact military decision making. Traditional and non-traditional adversaries are quickly developing these technologies for their own use which complicates the Just War Tradition and the ethical response. This makes decision making more complex than dealing with a few technology weapons.

Commanders are accustomed to a top-down chain of command for clear lines of leadership. Sometimes silos are built for protection of power, ideas, innovation, or even insecurity. However, technology is reshaping how men and women lead in technology. Cross-functional and self-directed teams that can quickly leverage data and produce turnaround decisions will be the match for technology leadership. How is the military leader to change? Can the decision maker be more adaptive to contextual ethical scenarios. Can military leaders and machines work in tandem with respect to ethical decisions? Or will autonomous unmanned systems have precedence and authority for making ethical decisions?

Developing effective military leaders who will be equipped to handle these challenges requires an education in the Humanities and Ethics. A multi-disciplinary approach will foster the development of leaders who will think critically and effectively. We must also ask whether the the future military leader would be willing to risk their safe zone of command and control to have more control of the battlespace? To what extent will the human decision maker form a symbiotic relationship with the machine? What type of interface will be established to support the human decision maker and allow them to override the system? For the near term, the human will remain in the loop and will have to co-exist with technology in ethical dilemmas.

CONCLUSION

The evolution of autonomous unmanned systems has accelerated rapidly and with little time for consideration of the ethical consequences of their application in the military battlespace. As artificial intelligence and machine learning systems evolve and transform the character of warfare, systems with a capacity for self-awareness, and self-explanation will be designed to make informed, rationale decisions based on logic and reasoning capacities. Indeed, advances in AI, machine learning and computational modeling have moved research in the direction of affective computational models and machine consciousness (Chandra, 2017; Aberman, 2017).

The military has taken an aggressive posture in developing these technologies as a means of augmenting military capabilities while reducing overall design and manning costs. However, there is a need to consider the potential for catastrophic consequences for developing lethal autonomous unmanned systems within the context of potential negative consequences for military decision making and for society as a whole.

As autonomous unmanned systems continue to be developed and evolve, we have an obligation to raise questions regarding ethical and moral considerations related to their impact on society. For the military, these technologies provide significant battlespace and warfighting capabilities that serve as a force multiplier. Technologies such as unmanned aerial vehicles (UAVs) and unmanned undersea vehicles (UUVs) provide critical decision support capabilities that make the environment transparent to the warfighter. Such systems can be deployed in a stealth-like manner and provide surveillance capabilities that would otherwise not be available to the warfighter.

Although the employment of unmanned systems such as UAVs, and UUVs have proven to be critical to our military defense, there are significant societal and ethical concerns associated with this 21st century strategy. The question is, how do we best prepare ourselves to defend against this new age of weaponized autonomous systems? Will we need to consider how we defend our society against weaponized autonomous systems that have been programmed by our adveresaries to use against us? As a society, what are our ethical and moral obligations with regard to designing autonomous systems with artificial consciousness? Will society be prepared to accept such systems in their daily life, as well as on the battlefield?

The 21st-century is a critical time to address these challenges. Just as the Industrial Revolution had its societal challenges, so too, we must address questions related to advances in technology and their impact on society and warfare The Industrial Revolution has taught us that we want to maintain our technologies in support of our lifestyles. As a society, we are unwilling to return to a time without our technologies. We want to keep the advantages that technology has provided as technology evolved from the typewriter to the computer, or the telegraph to the mobile phone One lesson learned is that once we turn systems on whether it's a toaster, typewriter, telephone, computer, self-driving car, or a weaponized autonomous system, we will not turn it off. We do not un-invent but always continue to move forward as a society seeking ways to improve and integrate advances in technology aimed at improving the quality of our daily life and that of society as a whole. The question remains, how will society determine its own course of action. Will society allow autonomous systems to dictate actions in the battlespace? How will the commander trust the weaponized system of the future? These are the ethical dilemmas of 21st-century warfare.

> "The world is very different now. For man holds in his mortal hands the power to abolish all forms of human poverty and all forms of human life."
>
> John Fitzgerald Kennedy

ACKNOWLEDGMENT

We thank Ms. Isabel Lopes for her contribution to reviewing and editing this manuscript. Her expertise and knowledge contribute greatly to the editing and revision of this manuscript.

REFERENCES

Aberman, J. (2017, February 27). Artificial intelligence will change America. Here's how. The Washington Post (Online) Retrieved from https://www.washingtonpost.com/news/capital-business/wp/2017/02/27/artificial-intelligence-will-change-america-heres-how/?utm_term=.3e325159efd9

Arkin, R. C. (1992). Modeling neural function at the schema level: Implications and results for robotic control. In R. D. Beer, R. E. Ritzmann & T. McKenna (Eds.), Biological neural networks in invertebrate neuroethology and robotics (pp. 383-410). Cambridge, MA: Academic Press.

Arkin, R. C. (2007). Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture (technical report GIT-GVU-07-11). Atlanta: Georgia Tech GVU Center. Retrieved from http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf

Chandra, R. (2017). An affective computational model for machine consciousness. Retrieved from http://arxiv.org/abs/1701.00349

Cummings, M. L., Clare, A., & Hart, C. (2010). The role of human-automation consensus in multiple unmanned vehicle scheduling. Human Factors: The Journal of Human Factors and Ergonomics Society, 52(1), 17-27. doi:10.1177/0018720810368674

Cummings, M. L., & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(1), 1-15. doi:10.1518/001872007779598109

Department of Defense. (2012). Autonomy in weapon systems (DoD directive Number 3000.09). Washington, DC: Department of Defense

Dobbyn, C., & Stuart, S. (2003). The self as an embedded agent. Minds and Machines, 13(2), 187-201. doi:1022997315561

Dutt, N., & TaheriNejad, N. (2016). Self-awareness in cyber-physical systems. Paper presented at the 29th International Conference on VLSI Design and 15th International Conference on Embedded Systems (VLSID), Kolkata, India. Retrieved http://ieeexplore.ieee.org/document/7434906/

Giordano, J., Kulkarni, A., & Farwell, J. (2014). Deliver us from evil? The temptation, realities, and neuroethico-legal issues of employing assessment neurotechnologies in public safety initiatives. Theoretical Medicine and Bioethics, 35(1), 73-89. doi: 10.1007/s11017-014-9278-4.

Goodrich, M. A., & Cummings, M. L. (2014). Human factors perspective on next generation unmanned aerial systems. In K. P. Valavanis, & G. J. Vachtsevanos (Eds.), Handbook of unmanned aerial vehicles (pp. 2405-2423). New Delhi, India: Springer.

Hamel, G., & Green, B. (2007). The future of management. Cambridge: Harvard Business Review Press.

Lin, P., Bekey, G., & Abney, K. (2008). Autonomous military robotics: Risk, ethics, and design. . Retrieved from http://www.dtic.mil/docs/citations/ADA534697

Loten, A., & Simons, J. (2017, Jan 04). Leadership evolves amid tech changes - management styles shift to embrace shorter, more frequent data-fueled development cycles. Wall Street Journal. Retrieved from https://search.proquest.com/docview/1855011133?accountid=322

Masakowski, Y. R., Smythe, J. S., & Creely, T. E. (2016). The impact of ambient intelligence technologies on individuals, society and warfare. Northern Plains Ethics Journal, 4(1), 1-11. Retrieved from http://www.northernplainsethicsjournal.com/NPEJv4n1/The%20Impact%20of%20Ambi ent%20Intelligence%20Technologies%20on%20Individuals.pdf

Porat, T., Oron-Gilad, T., Rottem-Hovev, M., & Silbiger, J. (2016). Supervising and controlling unmanned systems: A multi-phase study with subject matter experts. Frontiers in Psychology, 7, 568. doi:10.3389/fpsyg.2016.00568

Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control. Cambridge: MIT Press.

Singhvi, A., & Russell, K. (2016, July 12). Inside the self-driving tesla fatal accident. New York Times. Retrieved from https://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0

Wallach, W. & Allen, C. (2009). Moral machines: Teaching robots right from wrong. New York: Oxford University Press.

ARGUinDSM – A model of argumentation in team problem-solving situation: An application to nuclear control room teams

Johanna MÉRAND[a], Françoise DARSES[a],
[a]*French Armed Forces, Biomedical Research Institute, Brétigny-sur-Orge, France*

ABSTRACT
In this paper, we propose a model - *ARGUinDSM* - that combines a model of decision-making in dynamic risky situations with the theoretical framework of Argumentation. This model was built to investigate the joint problem-solving processes that are implemented within teams, when coping with complex dynamic situations.

We tested the *ARGUinDSM* model in the domain of nuclear power plant supervision. Control room teams were observed in various incident management scenarios. We examined the impact of two different organizational modalities on the argumentative processes developed by the teams during joint problem solving, and on the ensuing decisions.

KEYWORDS
Team problem-solving processes; Argumentation; Organization; Reliability, Nuclear power plant

INTRODUCTION

Models for analysing problem-solving processes in dynamic situations

The supervision of most complex and risky systems requires teamwork. Each member of the team contributes to achieving objectives by carrying out sub-tasks that are specific to their role and skills. Team members are mutually dependent, and must integrate what they know, coordinate what they do, and cooperate to achieve shared goals. This is the reason why extensive research in the Naturalistic Decision Making community has addressed the cooperative dimension of supervision (Cannon-Bowers, Salas, & Converse, 1993). But most contributions are not focused on joint problem-solving processes themselves.

The model proposed by Burke, Stagl, Salas, Pierce & Kendall (2006) focuses on how adjustments are performed by the team to adapt their behaviour to the current situation. These adjustments are described through an adaptive cycle, unfolded in four phases: (i) assessment of the situation, (ii) formulation of a plan, (iii) execution of the plan, and (iv) team learning. Each of these phases is supported by socio-cognitive representations, such as shared mental models, or team awareness. But the processes involved in each phase of the cycle are not described in detail by the authors.

Another model, known as the Common Frame of Reference - COFOR (Hoc, 2001), describes the team decision making processes into three levels: *action*, *planning* and *meta* levels. The lowest level (*Cooperation in Action)* emphazises the interferences, either negative (e.g. workload increasing) or positive (e.g. mutual control) which are created, detected and solved between members during the execution of actions. The highest level (*Metacooperation*) is dedicated to the elaboration of compatible representations and to the elaboration of a model of oneself or of the others (goals, knowledge). This level thus improves the cooperative activities at the two lower levels. This level is more developed for expert teams in which their members used to work together. The intermediate level (*Cooperation in Planning)* applies both to the state of the process, its environment, the activity of each partner, the available resources and means for action. Thus, it supports the management of shared plans and goals and accounts for the elaboration of *shared representation* of the situation.

However, none of these models addresses in detail the problem-solving processes that the team must draw upon when they have to manage a situation that is ambiguous, or a situation where operating procedures are not optimal. In these circumstances, the team must develop a solution that ensures the reliable operation of the installation. Vicente, Mumaw and Roth (2004) describe the cognitive supervision strategies that precede the detection of a problem in nominal conditions. But their model does not shed much light on problem-solving processes, when it is necessary to develop a representation or an action plan.

We argue that the DSM model (Dynamic Situation Management model, Hoc & Amalberti, 1995) provides the most detailed description of the cognitive processes involved in dynamic problem-solving. This model was adapted from the well-known Rasmussen's Decision Ladder model (1986). It provides a detailed cognitive architecture, with three levels of abstraction describing cognitive controls. These different levels of abstraction (knowledge-based, rule-based and skill-based) can be either triggered for problem-solving. There are more or less expensive in terms of symbolic attentional resources. But more importantly, the model is built around the occurrent representation of the situation, which is constantly adjusted as the situation unfolds. It relates to both how the process unfolds, and the evaluation of the means, and the external and internal resources available for the action. The occurrent representation is therefore closely intertwined both with the processes of state assessment and the processes of defining and planning a task.

Although the DSM model provides a very detailed cognitive architecture of problem-solving processes, it has an important limitation. It only focuses on individual problem-solving processes. It is therefore necessary to adapt it for studying team problem-solving processes. This is what we are proposing with *ARGUinDSM.*

The theoretical framework of Argumentation: an innovative blueprint to describe cooperative problem-solving processes in dynamic situations

According to Schmidt (1991), cooperation has three functions for the team members: augmentative, integrative and debative. The integrative and debative functions are most often mobilized when elaborating a shared representation of a situation or an action plan. These functions are implemented through interactional processes. In the context of the supervision of a nuclear power plant, these interactions mainly rely on verbal communication. In this domain, time spans are usually long enough to allow in-depth joint understanding and problem solving.

Many studies have investigated the relationship between verbal communication and team performance in dynamic and non-routine situations. Both their objectives and the associated methods are varied. Some studies aim to analyse processes of cooperation and mutual understanding (Fischer, McDonnell, & Orasanu, 2007; Lee, Park, Kim, & Seong, 2012; Waller, Gupta, & Giambatista, 2004). Others focus on the structural and organizational properties of communication networks (Barth, Schraagen, & Schmettow, 2015; Schraagen & Post, 2014). But very little work has looked at the argumentative mechanisms in dynamic and non-routine situations. Lu and Lajoie (2008) examined them in the medical field, while Bourgeon (2013) studied them in the field of aeronautics. We believe that the theoretical framework of Argumentation, which has a long history in psycholinguistics, is particularly relevant for examining the integrative and debative functions at play in cooperation (Schmidt, 1991). This frame provides theoretical and conceptual elements that are relevant to the study of verbal interactions in problem-solving situations, particularly when the focus is the reliability of decisions.

According to the Argumentation framework, to argue is to convince another person of the merits of a thesis, drawing upon assertions or reasonable arguments in a context of uncertainty (Breton, 2006; Oléron, 1996; Perelman & Olbrechts-Tyteca, 1958/2008). Argumentation has a dual function (Champaud, 1994; Darses, 2006; Nussbaum, 2008). It is *cooperative,* because the pursuit of a shared goal encourages convergence towards new solutions (Baker, 1999). It is also *dialectic* when two views conflict. Refuting, objecting, contradicting, casting doubt, evaluating by providing supporting evidence all make it possible to test the validity of the propositions and arguments put forward by other group members (Rieke, Sillars, & Peterson, 2005). This dual function makes argumentation a potential process for improving the reliability of decision-making process.

ARGUINDSM: A MODEL TO ANALYSE ARGUMENTATION IN DYNAMIC ENVIRONMENTS

We hypothesize that the performance of teams during the collective resolution of problems depends on the argumentative processes developed by its members when elaborating the decision. However, the deployment of argumentative processes will be influenced by organizational modalities (e.g., the number of team members or/and their functions). These modalities have an impact on the reliability of joint decision-making. To test these hypotheses, it is necessary to draw upon on a model of argumentation in dynamic situations. We therefore develop the *ARGUinDSM* model (Figure 9). It aims to describe and predict how argumentation impacts the problem-solving processes. The *ARGUinDSM* model combines the theoretical framework of argumentation (Perelman & Olbrechts-Tyteca, 1958/2008; Toulmin, 1958/2003) and the DSM architecture (Hoc & Amalberti, 1995). The *ARGUinDSM* model is based on two key concepts: the *argumentative aim* and the *argumentative function*.

Argumentative aims. – We consider four aims to characterize the interactional position that interlocutors adopt in problem-solving. Three of them are: *Feeding*, *Building*, and *Critiquing* a position about the problem. Each of them is linked to one of the three levels of abstraction of the DSM model. The fourth one is a cross-sectional aim, which consists in *Approving/ Disapproving* the interlocutor position. These argumentative aims are defined in Table 1.

Argumentative functions. – These functions are the engines that make it possible to construct the aims. Each argumentative function contributes to a particular aim. These functions are listed in Table 1.

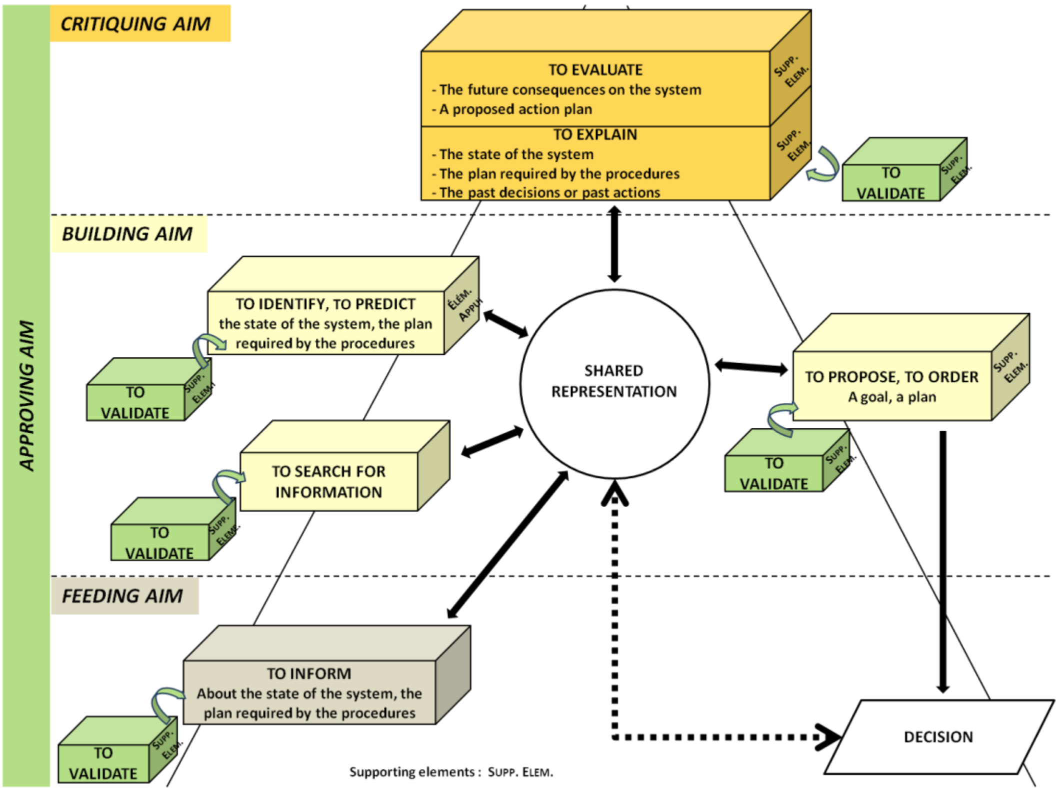


Figure 9. Model of argumentation in dynamic situations – *ARGUinDSM*

Table 1. Argumentative functions that support argumentative aims

Argumentative aims	Argumentative functions supporting the aim
Feeding aim: It brings together exchanges aimed at bringing or asking for information that will feed the co-construction of the problem and the positions of each team members.	Inform – Request information
Building aim: It brings together exchanges aimed at constructing and developing a representation of the state of the problem and/or the action plan in order to solve the problem.	Search for information – Ask for information – Identify – Propose – Order
Critiquing aim: It brings together exchanges aimed at explaining and evaluating the state of the system or the action plan.	Evaluate – Request an evaluation – Explain – Ask for an explanation
Approving aim: It brings together exchanges aimed at reinforcing, questioning, disapproving the argumentative functions belonging to each of the aims.	Validate – Request validation – Invalidate (an assertion)

Each argumentative function that expresses a position, such as *Identify*, *Propose*, *Evaluate* or *Validate*, can be accompanied by *supporting elements* that serve to justify it. These *supporting elements* are the simplified equivalent of the *data*, *guarantees* and *foundations* found in Toulmin's (1958/2003) model of argumentation. In this paper, we will not report results about these supporting elements. Thus, the *ARGUinDSM* reflects the progressive development of the shared representation, which evolves during the interaction through the articulation of different individual points of view, for which there is more or less support.

METHODOLOGY

Context

This study fits in with a large-scale Human Factors evaluation program of a new nuclear reactor (Labarthe & De La Garza, 2011). The aim is to contribute to the organizational specifications related to the operating teams and the means of the new control room (operating procedures, human-system interface features). Our study thus took place in a sociotechnical system which had not yet been stabilized. Its purpose was to ensure the system reliability, especially on its organizational side. In this context, two team organizations were tested: an organization composed of five members *(ORGA-5),* similar to those already found at nuclear power generation plants, and a new organization consisting of four members *(ORGA-4)* (Figure 10). The new team organization (task allocation, team coordination, required practices and workload management) and its technical and regulatory environment have been the subject of a detailed analysis which is not presented here. The functions performed by the operators in these two organizations are similar. One of the main differences is that the process control actions are handled by a single operator in *ORGA-4* teams, compared to two in *ORGA-5*.

The functions performed are as follows. The *OP* carries out actions that control the process and checks the result. The *SUP* carries out a second check of these actions, anticipates what might happen next, based on procedures, and ensures that current actions are appropriate to the status of the installation. The operation's manager (CE), team leader, checks the consistency between the current control strategy and the status of the installation. He/she forms an interface with the internal hierarchy and external experts who are not in the control room and who are regularly updated about current operations. The safety engineer (IS) carries out various independent checks of the status of the installation. He/she has a human-machine interface and follows specific operating procedures. If an incident occurs, the key operational challenge is to guarantee the safety of the installation. The objective of the team is to bring the plant back into a stable and controlled state by ensuring constant core cooling, and the control and containment of reactivity. Operators follow written, step-by-step operating procedures contained in manuals that indicate the actions to be taken as a function of the status of the installation. Despite this high level of proceduralization, operators sometimes have to adapt to unforeseen or ambiguous situations where the operating procedures are not fully optimized into a system whose design is not completely finalized. These problematic situations may be due to, for example, a perceived difference between the status of the installation and the current operational strategy or a lack of understanding of the behaviour of the technical system (automatisms, regulation). When these differences require operators to develop or readjust their representation of the system state and/or the action plan, they implement joint problem-solving processes.

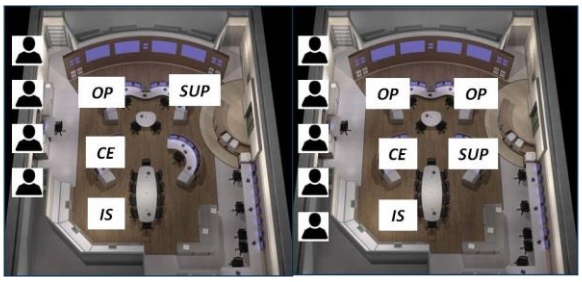


Figure 10. Configuration of control room teams: *ORGA-4* (left) and *ORGA-5* (right)

Data collection

Data collection consisted of 12 simulations performed by three *ORGA-4* teams and three *ORGA-5* teams in a full scale simulator. These simulations were organized as shown in **Table 2**. Recordings (audio/video) were made, leading to the creation of four complementary corpuses: An in-depth view of each workstation in order to visualize how written procedures were implemented by the operator, screenshots of workstations and overviews, a general view of the whole control room, audio recordings of each operator. The 36 hours of audio-video recording were supplemented by 36 hours of post-simulation debriefings, real-time note-taking, and 15 post-simulation personal interviews with the *ORGA-5* team members.

Table 2. Data collection design

Accident scenario	*ORGA-4* teams (3)	*ORGA-5* teams (3)
Thermo-hydraulic event potentially leading to radioactive release	Scenario 1a	Scenario 1b
Accumulation of events (thermo-hydraulic and electrical problems)	Scenario 2a	Scenario 2b

Data analysis

Verbal communication was transcribed in order to characterize the *EVENT–PROBLEMS* defined as *all exchanges related to solving a particular problem.* For example, an *EVENT–PROBLEM* begins when the team is wondering about the proper functioning of an automatism or the relevance of actions required by operating procedures. Each *EVENT–PROBLEM* is then divided into *UNITS OF MEANING.* Inspired by the theory of speech acts, a *UNIT OF MEANING* is an utterance which contains one idea or piece of information and whose content can be encoded in predicate/ argument form (Darses, Falzon, & Robert, 1993; Debanne & Chauvin, 2014).

Table 3. Coding of an EVENT–PROBLEM extract

		PREDICATE	ARGUMENTS[5]			
Unit of meaning	Transmitter	Argumentative function	Nesting	Purpose of the Communication	Time of this object	Supporting elements[5]
SUP: We'll get there (next manual action) the page just after for me *CE*, but hey, there's still some work to do between now and then!	*SUP*	EVAL[1]	-	TPS/BUTr[2]	future	no
CE: Yes, we have to inject the boron, complete isolation.	*CE*	VAL[3]	EVAL–SUP[4]	TPS/BUTr	future	yes

[1]EVAL: evaluate; [2]VAL: validate; [3]TPS/BUTr: time to achieve a goal required by written operational procedures; [4]EVAL-SUP: assessment made by the *SUP*. [5]In this paper, results about arguments are not reported.

The *ARGUinDSM* model was used to develop the coding scheme. **Table 3** shows the overall structure of this coding scheme and illustrates it using two examples. 2451 *UNITS OF MEANING* were coded. We compared the number of occurrences (here, *UNITS OF MEANING)* between teams or operators. As the number of *EVENT–PROBLEMS* managed by each team is specific, the number of occurrences is always based on the total number of occurrences observed for a team or an operator. We use the nonparametric Pearson Chi-squared statistic to test for the independence of qualitative variables. The results are expressed as percentages.

RESULTS

Effect of the organization on the argumentative activity of the team

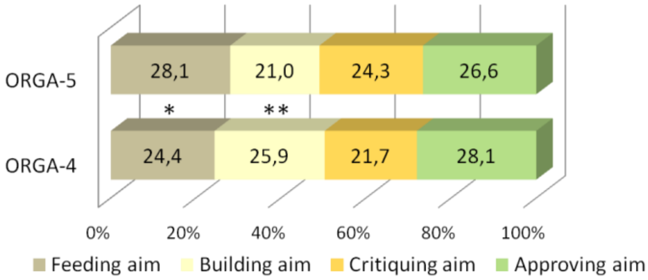


Figure 11. Use of argumentative aims as a function of the organization (*p<0,05 ; **p<0,01 ; ***p<0,001)

The results show a significantly different distribution of exchanges between argumentative aims as a function of the team organization (χ^2 (df = 3) = 11.78; p = 0.008) (**Figure 11**). *ORGA-5* teams are more involved in the *Feeding* aim (χ^2 (df = 1) = 4.29; p = 0.038) and less invested in the *Building* aim (χ^2 (df = 1) = 8.19; p = 0.004). However, the effect is moderate: the proportion of exchanges differs by a maximum of 4% from one organization to another. The *Critiquing* and *Approving* aims are, in turn, equally represented in both organizations. Further analysis of communications within the *Building* aim show that *ORGA-5* teams formulate more statements to *Search for information* (SEARCH: χ^2 (df = 1) = 19.42; p < 0.001), fewer statements to *Identify* the status of the system (IDEN: χ^2 (df = 1) = 10.27; p = 0.0013), and to *Propose* solutions to problems (PROP: χ^2 (df = 1) = 4.89; p = 0.027). On the other hand, no significant difference was observed between the two organizations regarding the distribution of the argumentative functions underlying the *Feeding*, *Critiquing* and *Approving* aims.

Impact of organization on operators' involvement and contribution, as a function of their role

We also investigate the impact of the organization on the argumentative activity of team members, according to their function in the team. To investigate this influence, we distinguish 'involvement' from 'contribution'. The *involvement* of an operator refers to his investment in the four argumentative aims. The *contribution* of an operator is the support he provides to the team in developing each argumentative aim. The way we measure both *involvement* and *contribution* is described in **Figure 12**.

$$\text{Involvement} = \frac{\textit{Number of meaning units transmitted by the operator for an argumentative aim}}{\textit{Total number of meaning units transmitted by the operator}}$$

$$\text{Contribution} = \frac{\textit{Number of meaning units transmitted by the operator for an argumentative aim}}{\textit{Total number of meaning units transmitted by the team to support the aim}}$$

Figure 12. Involvement and contribution measures

The results presented mainly concern *OP*, *SUP* and *CE* team functions, as the *IS* contributed very little to solving the *EVENT–PROBLEMS*. Between the *OP*, *SUP* and *CE* functions, the involvement of the *OP* function varied the most depending on the organization (Figure 13). The *OP* is more involved in the *Critiquing* aim in the *ORGA-5* teams (χ^2 (df = 1) = 4.49; p = 0.025), but less involved in the *Building* aim (χ^2 (df = 1) = 4.92; p = 0.026). Whatever the organization, the *OP* is very invested in the *Feeding* and *Approving* aims. The units of meaning issued by the *SUP* and *CE* are fairly similarly distributed across the four argumentative aims. No effect of the organization is observed on the aims used by the *SUP*. Finally, the *CE* in the *ORGA-5* teams is less involved in the *Approving* aim (χ^2 (df = 1) = 4.19; p = 0.040).

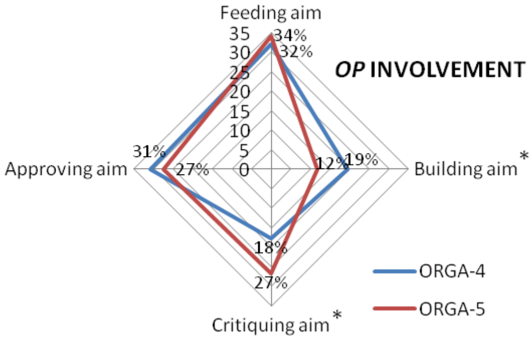


Figure 13. Involvement of the *OP* in the argumentative activity (*p<0,05 ; **p<0,01 ; ***p<0,001)

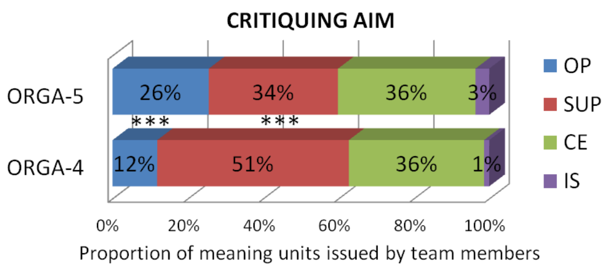


Figure 14. Operators contribution to the *Critiquing* aim according to their function in the team (*p<0,05 ; **p<0,01 ; ***p<0,001)

The clearest differences in the contribution of team functions relate to the *Critiquing* aim, notably for the *OP* and the *SUP* (Figure 14). In *ORGA-5* teams, the contribution of the *OP* to this aim is significantly higher than in the *ORGA-4* teams (χ^2 (df = 1) = 17.42; p < 0.001), while that of the *SUP* decreases (χ^2 (df = 1) = 15.71; p <0.001). The *OP* pays more attention to *Assessing* and *Explaining* the status of the problem, or its solutions. While it remains lower, the contribution of the *OP* to the *Critiquing* aim tends to be similar to that of the *SUP* and the *CE*. To a lesser extent, a similar phenomenon is observed for the *Feeding* aim. The contribution of the *OP* to this aim is greater in *ORGA-5* teams (χ^2 (df = 1) = 9.15; p = 0.002), while that of *SUP* decreases (χ^2 (df = 1) = 6.48; p = 0.01). Finally, concerning the *Approving* aim, the results show a significantly greater contribution of the *OP* in the *ORGA-5* teams (χ^2 (df = 1) = 7.01; p = 0.008) while that of *CE* decreases (χ^2 (df = 1) = 4.90; p = 0.026). The contributions of all team members to the *Building* aim are the same, regardless of the organization.

DISCUSSION

Empirical contribution: the effect of the organization on argumentative activity

Our analyses help to support or confirm organizational choices (*ORGA-4* vs. *ORGA-5*). Regarding the effect of the organization on the argumentative activity of the whole team, the results suggest that *ORGA-5* teams have a better understanding of problems. On the one hand, these teams draw more upon the *Feeding* aim, and much less upon the *Building* aim. On the other hand, within the *Building* aim, the argumentative function *Search for information* is drawn upon more in the *ORGA-5* teams, while there are fewer identifications and propositions than in the *ORGA-4* teams. *ORGA-5* teams appear to have greater latitude to feed the resolution of the problem, then to build a representation of it and the action plan, based on the behaviour *Search for information.* Identifications seem to become more accurate and propositions more appropriate, as they do not need any further reformulation. The time taken to build a representation of the problem and an action plan is consequently reduced. The organizational modality (*ORGA-4* vs. *ORGA-5*) has an effect on how the team members involve themselves and contribute to the argumentative activity, notably for the *OP* and *SUP*. *ORGA-5* encourages greater involvement and contribution by the *OP* in the *Critiquing* aim. This operator is more likely to express their opinion and offer a judgment on the situation. In the *ORGA-5* teams the *SUP* is no longer the major contributor to this aim and the contributions tend to be more divided between the *OP*, *SUP* and *CE* functions. Because the workload of the *OP* function is split between two members in the *ORGA-5* teams, operators take a more critical stance, which gives them a better overview of the situation. We argue that this improves the reliability of the decisions that are taken in the control room.

Theoretical Contribution: the *ARGUinDSM* model as a model of team decision-making

The *ARGUinDSM* model was built to conceptualize the argumentative activity that manifests in teams in dynamic and complex environment. In our study, we used it to describe the argumentative activity of team members when they collectively solve a supervision problem in a context of incident-related operation. Our study showed that this model was operationalizable on real-world dialogue, gathered in ecological simulated situations related to operations in a complex risky system. Thus, the *ARGUinDSM* model offers a new understanding of control room decision-making processes, supporting reliability improvement. Firstly, we assume that it may also be relevant in the context of normal and abnormal operation. In these contexts, operators must also solve problems that may even prove to be more difficult, as the situations are not completely covered by less detailed operational requirements. Secondly, the *ARGUinDSM* model can be generalised to all dynamic environments in which the timespan is long enough to enable operators to work together to solve problems. In this respect, it could be interesting to test its application in others areas and for larger teams. Thirdly, the prospect to use *ARGUinDSM* model in more routine situations (not only in problem solving situations) must be tested, provided the team members can exchange sufficiently in the situations examined.

CONCLUSION AND PERSPECTIVES

In addition to these results, we studied the link between argumentative activity and the performance of teams. Performance was measured on the basis of *diagnosis accuracy* and *decisions relevancy* during and/ or following the *EVENT–PROBLEM*. Two independent experts (a Human Reliability expert and an Ergonomics expert) judged it according to the potential safety impact of the decisions made by the team. Two groups were formed: two *PERF−* and four *PERF+* teams. The results, which are not presented here, did not find a link between argumentative activity and team performance. However, these results must be treated with caution due to a potential sample size bias (unequal number of teams in each group). Additional factors may explain performance, such as level of experience and level of knowledge specific to the new process. These points should be investigated in future studies, involving more teams.

ACKNOWLEDGMENTS

This work was supported by EDF Research and Development (R&D) and The French Armed Forces Biomedical Research Institute. We particularly thank Cecilia De La Garza (EDF R&D) who supervised this study.

REFERENCES

Baker, M. J. (1999). Argumentation and constructive interaction. *Foundations of Argumentative Text Processing*, *5*, 179–202.

Barth, S., Schraagen, J. M., & Schmettow, M. (2015). Network measures for characterising team adaptation processes. *Ergonomics*, *58*(8), 1287–1302. https://doi.org/10.1080/00140139.2015.1009951

Bourgeon, L., Valot, C., & Navarro, C. (2013). Communication and Flexibility in Aircrews Facing Unexpected and Risky Situations. *The Internat. J. of Aviation Psychology*, *23*(4), 289–305. https://doi.org/10.1080/10508414.2013.833744

Breton, P. (2006). *L'argumentation dans la communication*. Paris, France: Editions La Découverte.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: a conceptual analysis and model. *The Journal of Applied Psychology*, *91*(6), 1189–1207. https://doi.org/10.1037/0021-9010.91.6.1189

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making: Current issues*. Hillsdale, N: Erlbaum.

Champaud, C. (1994). L'argumentation. *Psychologie Française*, *39*(2), 193–203.

Darses, F. (2006). Analyse du processus d'argumentation dans une situation de reconception collective d'outillages. *Le travail humain*, *69*(4), 317–347. https://doi.org/10.3917/th.694.0317

Darses, F., Falzon, P., & Robert, J. M. (1993). Cooperating partners : investigating natural assistance. In G. Salvendy & M. J. Smith (Eds.), *Human-Computer Interaction: Software and Hardware Interfaces, Proceedings of the 5th Internl Conf. on HCI* (Vol. 2, pp. 997–1002). Orlando, Florida, USA: Elsevier.

Debanne, T., & Chauvin, C. (2014). Modes of Cognitive Control in Official Game Handball Coaching. *Journal of Cognitive Engineering and Decision Making*, *8*(3), 283–298. https://doi.org/10.1177/1555343414538819

Fischer, U., McDonnell, L., & Orasanu, J. (2007). Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. *Aviation, Space, and Environmental Medicine*, *78*(Supplement 1), B86–B95.

Hoc, J.-M. (2001). Towards a cognitive approach to human–machine cooperation in dynamic situations. *International Journal of Human-Computer Studies*, *54*(4), 509–540. https://doi.org/10.1006/ijhc.2000.0454

Hoc, J.-M., & Amalberti, R. (1995). Diagnosis: Some theoretical questions raised by applied research. *Cahiers de Psychologie Cognitive*, *14*(1), 73–101.

Labarthe, J.-P., & De La Garza, C. (2011). The human factors evaluation program of a control room: The French EPR approach. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *21*(4), 331–349. https://doi.org/10.1002/hfm.20227

Lee, S. W., Park, J., Kim, A. R., & Seong, P. H. (2012). Measuring situation awareness of operation teams in NPPs using a verbal protocol analysis. *Annals of Nuclear Energy*, *43*, 167–175. https://doi.org/10.1016/j.anucene.2011.12.005

Lu, J., & Lajoie, S. P. (2008). Supporting medical decision making with argumentation tools. *Contemporary Educational Psychology*, *33*(3), 425–442. https://doi.org/http://dx.doi.org/10.1016/j.cedpsych.2008.05.005

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, *33*(3), 345–359. https://doi.org/DOI: 10.1016/j.cedpsych.2008.06.001

Oléron, P. (1996). *L'argumentation.* (4e éd). Paris, France: Presses universitaires de France.

Perelman, C., & Olbrechts-Tyteca, L. (2008). *Traité de l'argumentation. La nouvelle rhétorique. 6ème édition.* (6e éd). Bruxelles, Belgique: Editions de l'Université de Bruxelles.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York, NY, USA: Elsevier Science Inc.

Rieke, R. D., Sillars, M. O., & Peterson, T. R. (2005). *Argumentation and critical decision making.* (6e éd). Pearson.

Schmidt, K. (1991). Cooperative work: A conceptual framework. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed decision-making: Cognitive models for cooperative work* (pp. 75–110). Chichester, UK: Wiley.

Schraagen, J. M., & Post, W. (2014). Characterizing naval team readiness though social network analysis. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (Vol. 58, pp. 325–329). Chicago, USA: SAGE Publications.

Toulmin, S. E. (2003). *The uses of argument* (2e éd.). New York, USA: Cambridge University Press.

Vicente, K. J., Mumaw, R. J., & Roth, E. M. (2004). Operator monitoring in a complex dynamic work environment: a qualitative cognitive model based on field observations. *Theoretical Issues in Ergonomics Science*, *5*(5), 359–384. https://doi.org/10.1080/14039220412331298929

Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of Adaptive Behaviors and Shared Mental Models on Control Crew Performance. *Management Science*, *50*(11), 1534–1544. https://doi.org/10.1287/mnsc.1040.0210

Using Sensemaking to Better Understand Chronic Pain Management

Laura G. MILITELLO[a], Shilo ANDERS[b], Sarah M. DOWNS[c], Julie DIIULIO[a], Elizabeth C. DANIELSON[c], Robert W. HURLEY[d], Christopher A. HARLE[c]
[a]*Applied Decision Science, LLC*
[b]*Vanderbilt University Medical Center*
[c]*Indiana University Richard M. Fairbanks School of Public Health*
[d]*Medical College of Wisconsin, Department of Anesthesiology*

ABSTRACT
Chronic pain leads to reduced quality of life for patients, and strains health systems worldwide. This article describes a study exploring how primary care clinicians in the U.S. manage their patients with chronic noncancer pain, using the data-frame theory of sensemaking. Based on a cognitive task analysis of primary care clinicians in the U.S., we identify key macrocognitive functions, common anchors, and clinician factors that influence the frames clinicians use to assess their patients and determine a pain management plan. The findings from this study have implications for training design, guideline development, and technological support.

KEYWORDS
Sensemaking; Health; Primary Care; Chronic Pain; Decision Making; Macrocognition

INTRODUCTION

Chronic pain is a costly condition worldwide, resulting in increased health care costs, lost productivity, and reduced quality of life (Barker & Moseley, 2016; Breivik, Eisenberg & O'Brien, 2013; Gaskin & Richard, 2012; Yeo & Tay, 2009). Adding to this burden, chronic pain frequently co-occurs with mental health conditions, such as depression, anxiety, and substance use disorder (Bair, Robinson, Katon, Kroenke, 2003; Gureje, Von Korff, Simon, Gater, 1998; Rosenblum, Joseph, Fong, Kipnis, Cleland, Portenoy, 2003; Martell, et al, 2007). The management of chronic pain by clinicians is influenced by political and legal issues related to the use of opioid analgesics and access to health services, as well as attitudes and understanding about pain conditions. In the United States (U.S.), the majority patients with chronic pain are followed by primary care clinicians (PCCs) (Committee on Advancing Pain Research Care Education, 2011). Yet, PCCs frequently have little training in chronic pain and struggle to deliver care that overcomes the complex, biopsychosocial, nature of many pain conditions and helps patients achieve improved health outcomes (Upshur, Luckmann, Savageau, 2006, Committee on Advancing Pain Research Care Education, 2011; Leverence, et al, 2011; Breuer, Cruciani, Portenoy, 2010). This article describes an in-depth study of chronic noncancer pain management by PCCs in the U.S. We use the data-frame theory of sensemaking to characterize macrocognitive strategies used by PCCs.

Chronic Pain Management in the U.S.

In the 1990s, the U.S. experienced a shift in medical practice with an increased emphasis on relieving pain for people with chronic noncancer pain. Physicians were encouraged to be more compassionate about pain. Pain was considered a "fifth vital sign" to be assessed routinely, and in this context, prescribing opioids became more accepted and more commonly employed. As access to potent prescription opioids increased, non-medical use became more common. This change in medical practice coincided with the increase in the potency and availability of heroin, a decrease in the cost of heroin (National drug control strategy: data supplement, 2014), and an increase in heroin use (Unick, Rosenblum Mars, Ciccarone, 2013). Non-medical use of prescription opioids is a risk factor for heroin use (Muhuri, Gfroerer, Davies, 2013). In the last decade, the majority of heroin users reported prescription opioid misuse, and declining availability of prescription opioids likely contributed to heroin use. However, the trend of increasing heroin use preceded that of the any policy or statutory change in opioid prescribing, suggesting other factors were at play as well (Compton, Jones, Baldwin, 2016).

Today, the response to this public health crisis falls largely on the shoulders of PCCs who must manage patients with chronic pain, many of whom have an unrealistic expectations and inaccurate understanding of the benefits and risks of prescription opioid therapy. Some patients are dependent on opioids as a result of previous prescribing practices; yet, still experience pain. Compassionate care continues to be a goal for PCCs as they treat patients with chronic pain. However, the evolving understanding of opioid therapy and legal/social changes regarding prescribing practices create challenges for clinicians as they seek to make sense of each patient's situation and create an effective pain management plan.

Because chronic pain encompasses a wide range of conditions that are treated by healthcare professionals from a range of disciplines, many clinical practice guidelines have been developed to help guide clinician decision making (Max et al, 1995; American Society of Anesthesiologist Task Force on Chronic Pain Management and the American Society of Regional Anesthesia and Pain Medicine, 2010; Chou, 2009; Manchikanti et al, 2012). Given the epidemic of opioid misuse and abuse in the U.S., the Centers for Disease Control and Prevention (CDC) recently published a guideline specifically for Prescribing Opioids for Chronic Pain. In addition to the CDC guideline, several U.S. states have created laws or other policies governing opioid prescribing for chronic noncancer pain. A common characteristic of many of these guidelines and policies for chronic pain is that they do not rely on rigid decision rules. Therefore, guideline-based chronic pain care requires access to and interpretation of information that is often scattered, missing, or out-of-date, such as medication history, imaging results, and pain assessments. Moreover, guideline recommendations can be difficult to follow when patients have co-morbid health conditions and PCPs have limited time during patient visits (Østbye, Yarnall, Krause, Pollak, Gradison, Michener, 2005; Abbo, Zhang, Zelder, Huang, 2008).

Data Frame Theory of Sensemaking

We chose to use the data frame theory because chronic pain management is a classic sensemaking activity, characterized by uncertainty, complexity, and changing conditions. PCCs must make choices about a patient's pain management plan, often without a clear understanding of the cause of the patient's pain, complicating factors such as a mental health diagnosis, and the patient's social support. The effectiveness of specific pain management strategies varies from patient to patient. Furthermore, all of these factors, as well as the political and regulatory environment, change over time. We anticipate that the data frame theory of sensemaking will be useful in highlighting the cognitive challenges and information needs of PCCs, and identifying leverage points for clinical decision support.

The data frame theory of sensemaking posits that "Sensemaking enables people to integrate what is known and what is conjectured, to connect what is observed with what is inferred, to explain and to diagnose, to guide actions before routines emerge for performing tasks and to enrich existing routines." (Klein et al, 2007, p 114). Sensemaking is defined broadly and is used to detect problems, connect disparate data elements, form explanations, project future states, find levers (determine how to think and act), see relationships, and identify problems. An important proposition of the theory is that the decision maker's mental representation (i.e., frame) influences what data are attended to and how they are interpreted, while at the same time the interpretation of data drive the formation and elaboration of the mental representation.

The data frame theory views sensemaking as dynamic. Key information or data are described as anchors for a specific frame. Anchors may change as new information becomes available requiring the decision maker to question the initial frame, elaborate the frame, re-frame, or compare frames. In cases characterized by uncertainty and conflicting information, a decision maker may seek a frame by looking for a similar analog or searching for more data. Decision makers may explain away inconsistent data, thereby preserving the frame in the face of contrary evidence. This may be adaptive in situations in which the data are unreliable or transient. Alternatively, preserving the frame may be maladaptive when new information represents a flaw in the initial frame.

METHODS

Overview

We conducted Critical Decision Method interviews with PCCs working in an urban safety net health system in the U.S. Because of the high volume of patients seen by PCCs it is often difficult to recall details of a specific patient. To aid in recall, we chose to conduct interviews within close proximity to the patient encounter to be discussed. Thus, for each PCC who agreed to participate in study, we identified patients with chronic noncancer pain via a computable phenotype that searched the electronic health record (EHR) for repeated diagnoses or medications that indicated chronic pain without a recent cancer history. When eligible patients with upcoming appointments were identified, a researcher met with each patient immediately prior to the patient encounter, confirmed the patient's chronic pain history, and invited him/her to participate in the study. If the patient agreed, the patient would be the topic of a critical decision method interview conducted within 2 days following the encounter. Note: The data presented here are part of a larger study that includes audio recordings of patient encounters that will be analyzed to triangulate with the critical decision method interview data described here. Furthermore, critical decision method interviews at additional sites are planned.

Participants

We report findings from 15 interviews conducted with seven PCCs. PCCs were invited to participate after a presentation describing the planned study. All were full-time physicians . Seven work for a large, urban safety net medical system. Two work for a network of PCCs that serve a more rural population in west central Indiana and eastern Illinois. All had received limited (or no) specialized training in managing chronic pain. Experience level ranged from 2.5 to 26 years.

Interviews

For the first set of PCCs interviews, three interviewers were present to facilitate common ground across the data collection team, and encourage refinements to the interview guide. After the first six interviews, interviewers worked in teams of two, including one lead interviewer and a primary note-taker. Due to schedule constraints, two interviews were conducted by only one interviewer. All interviews were audiotaped and transcribed. Interview notes served as an important backup to transcripts.

Each PCC participated in a of series interviews. Prior to the first interview, each PCC was asked to fill out a brief demographics questionnaire. During the first interview with each PCC, the interviewer asked a series of questions about the PCC's patient population, general approach to chronic noncancer pain care, and tools, instruments and assessment used when taking care of patients with chronic noncancer pain. For the remainder of the first interview and all subsequent interviews with that PCC, we used an adapted critical decision method interview technique (Crandall, Klein, Hoffman, 2006).

The critical decision method portion of the interview focused on the patient who agreed to participate in the study, except in cases in which scheduling conflicts precluded an interview within two days of the patient encounter. In those cases, the interview focused on a recent patient with chronic pain that stood out in the interviewee's memory. Participants were asked to recall the patient of interest, and encouraged to open the patient record in the EHR as a memory aid, if needed. After thoroughly exploring the patient incident, the interviewer asked the PCC to complete a multi-item questionnaire that assessed the clinicians' satisfaction with the most recent patient visit. In the final part of the interview, the interviewer asked a series of questions about information needs and the EHR, exploring how the EHR supports management of patients with chronic pain conditions, and limitations of the current EHR.

Analysis

We used a modified grounded theory (Strauss & Corbin, 1990) approach to analyze deidentified transcripts from each interview. This process emphasizes the value of exploring differences in interpretation rather than reaching coding concordance via inter-rater agreement (Barbour, 2001). We used a process of upward abstraction in which we coded the data to explore specific components of the interviews. To develop a codebook for the upward extraction, each team member (5 behavioral researchers, 2 primary care clinicians, 1 pain physician) reviewed two transcripts and identified topics of interest and potential themes. We compiled responses and met to discuss a draft codebook. Four behavioral researchers (SA, ED, SD, LM) coded one interview using Dedoose qualitative analysis software, and met to discuss points of agreement and disagreement. The codebook was refined and the process repeated for a second and third interview. After the third interview, the interview analysis team had reached consensus on a codebook and the codes for the three interview transcripts. Two researchers then coded an additional 12 interview transcripts, meeting after each to reach consensus on all codes. After 15 transcripts had been coded, two researchers (SA, LM) began the process of upward abstraction. They reviewed data in individual coding categories, exploring themes within individual categories and across related categories. They drafted summary documents highlighting themes and insights from the data. Summaries were shared with the larger team for additional discussion and input from clinician team members (RC, RH, BM). Clinician team members served as a reality check and were often able to provide additional context or interpretation of potential themes.

Because PCC frames were of particular interest to this project, additional analyses were conducted on these data. Two researchers extracted 61 segments containing descriptions PCCs used to characterize their patients. These were entered into an Excel spreadsheet and sorted into four initial categories. The larger team was then asked to conduct their own individual card sort, placing segments into the four categories and creating new categories when a segment did not fit. Individual card sorts were integrated into a single document and discussed until consensus was reached. As a result of this activity we identified subcategories within the initial 4, and an additional category not considered initially (i.e., clinician factors). Further discussion and refinement distinguished macrocognitive activities, anchors, and clinician factors.

FINDINGS

We identified macrocognitive functions that support sensemaking, as well as anchors and clinician factors from which frames are derived (Figure 1). For the purposes of this study, we define a frame as the mental representation used to develop a pain treatment plan for each patient, as depicted at the center of Figure 1. Anchors (shown in the top half of the second ring) are specific patient-related data elements that aid in creating the frame, while at the same time the frame influences which data elements might be considered an anchor. Clinician factors (shown in the bottom half of the second ring) are also important contibutors to the frame. The macrocognitive functions (shown in the outer ring) represent the cognitive processes PCCs use to identify anchors. Each is described in turn.

Macrocognitive Functions

The cases described by our PCC participants emphasized four important macrocognitive functions when managing patients with chronic pain. First, PCCs describe situations in which they interpret patient actions and motives. Because opioids are a controlled substance with potential for abuse and diversion, clinicians must be vigilant for indicators that medications will not be (or are not being) used as recommended. Clinicians attend to a range of things including patient behaviors during the clinician-patient encounter. Patients who are challenging to interview with long, repetitive stories about their pain (‘*she talks about all the different pain that she has and then how careful she is about always taking her medications and that no one ever has access to them, other than her and, … she has a very long script really, and goes through a lot of detail*’), or who want to talk about their pain to the exclusion of all other health issues (‘*I can't talk to this patient about anything except, she just wants her pain meds*’) may raise suspicions about the patient’s motivation. Some patients are aggressive in pursuit of pain medications, requesting opioids specifically or calling often with repeated requests (‘*because just in the 24-48 hours after visit, my nurse had gotten many, many phone calls from her about pain medications*’). Another cue might be the patient’s openness to trying pain management strategies that do not include opioids (‘*saying nothing else works for them…’*).

A second macrocogntive function PCCs described was assessing the patient’s condition with regard to pain. This can be complex as there may be no supporting evidence of pain (i.e, causal pathology on imaging, physical exam or other objective measures such as eletrodiagnostic testing); yet, the patient reports pain (‘*anytime you have someone with depression you wonder how that affects you know their perception of pain, their compliance, you know those kind of things.*’). Thus, the PCC considers factors such as whether the pain is likely related to an acute medical condition or is persistent in nature, how miserable or comfortable the patient is, and how functional the patient is currently and likely to be in the future (i.e., able to manage self-care, able to work, etc.).

A third macrocognitive function is assessing risk. Relevant risks include potential for opioid misuse, abuse, diversion, but also side effects and overdose risks associated with opioid and other pain medications. Above all else, clinicians intend to ‘do no harm ;’ however, with chronic pain care it can be difficult to assess the potential risks and benefits of a specific treatment plan – in part because the PCC must rely on often self-report data from the patient for key information. PCCs consider factors such as whether there is evidence of substance abuse. For example, one PCCs described a patient who came to every appointment inebriated. This made it difficult to address the patient’s pain at all as even non-opiod medications may exacerbate liver damage caused by alcohol. Some patients have a pre-existing dependence on opioids, so the PCC must consider the difficulties in weaning the medication. PCCs report that the patient’s ability to manage other chronic issues may be an important risk factor. For example, if the patient is unable to manage the recommended regimen for bipolar disorder, PCCs may consider whether prescribing opioids will exacerbate existing problems. PCCs also report that they consider the patient’s stability including emotional stability, the patient’s living situation (is the patient homeless ?) and social support (Does the patient have family or friends who can help with transportation and self-care, if needed? Are there other people in the home who might harm the patient and/or steal medication?).

A fourth macrocognitive funtion is identifying goals. Goals are driven by the patient, but often the PCC has a role in helping the patient articulate realistic goals. For example, many patients come to the doctor with goals of being pain-free and returning to a level of functionality they experienced as a younger, healthier person. The PCCs may help patients understand that they may never be pain free, but they can still improve their physical function or quality of life. PCCs often aid patients in articulating concrete goals such as living opioid free, returning to work (perhaps part-time), resuming a course of studies, or increased level of self-care (‘*… the goal is not for his pain to... his pain probably won't be gone forever. He probably won't have the strength he had in his hands when he was 20 years old either. No matter what we do the goal is just to improve as much as we can and have him functioning as much as we can.*’).

Anchors

The macrocognitive functions described above aid the PCCs in identifying anchors that will inform the frame used to develop a pain management plan. Common anchors for PCPs in managing patients with chronic pain include the the patient’s ability to manage opioids, the pain impact, the pain etiology, and opioid management history. With regard to the patient’s ability to manage opioids, the key determinant is whether the PCC believes the patient is capable of handling opioids safely. With regard to the pain impact, PCCs describe patients for whom pain has completely disrupted their lives, greatly reducing level of function, and contributing to depression. Other patients experience pain, and still enjoy a comfortable life. Anchors related to pain etiology focus on whether the underlying cause of the pain can be treated and will resolve. Opioid management history refers to how patterns of opioid use in the past increase likelihood that the patient will maintain the current level of opioids, require an increase in opioids, successfully wean from opioids, or be treatable without opioids. It is important to note that for each of these potential anchors, PCCs descibe patients that are difficult to characterize or place with regard to some anchors. Each potential anchor is reliant on ambiguous and changing information. What initially appeared to be an acute problem may become chronic. The PCC may not have enough information to determine whether the patient will be able to manage opioids safely and legally. It may be unclear whether a specific injury or condition will respond to treatment. For these situations in which the clinician is unsure, s/he is likely to be in a frame seeking mode, identifying what information is needed to resolve the uncertainty.

Figure 15: Sensemaking in chronic pain management

Clinician Factors

We identified four clinician factors that influence framing. A first clinician factor is the experience and training the PCC has in managing chronic pain. One PCC in our study reported that s/he had over 20 years of experience managing patients with chronic pain. In fact, this had been an area of interest, and s/he was well-read on the topic and maintained a panel with a higher proportion of patients with chronic pain than many colleagues. This PCC was able to reflect on the changes in guidelines and regulations regarding opioids and chronic pain management over time, as well as the experiences of his patients in that time period. The PCC had observed first-hand that long-term use of opioids results in decreased function and often is less effective in treating pain over time. As a result, this PCC would not consider long-term opioid use for any patient that still had a reasonable likelihood of functional life.

A second clinician factor is perception of own role. One PCC described a patient who desperately wanted to wean from opioids but was having limited success. The PCC made weekly appointments with the patient, primarily to provide emotional support and encouragement. Others might consider this outside the role of the PCC. In fact, this PCC indicated that these weekly appointments were only possible because s/he was new in the clinic and currently had a relatively small panel of patients.

A third clinician factor is interpretation of regulations and guidelines. Regulations and guidelines relevant to chronic pain are diffuse and require interpretation on the part of the PCC. For example, the U.S. Centers for Disease Control and Prevention guidelines states: ‘When prescribing opioids for chronic pain, clinicians should use urine drug testing before starting opioid therapy and consider urine drug testing at least annually to assess for prescribed medications as well as other controlled prescription drugs and illicit drugs.’ One PCC pointed out that the urine drug test used at his/her facility does not detect all relevant substances and a blood test may better meet the spirit of the guidelines for some patients.

A fourth clinician factor is knowledge of resources available. One PCC reported that there were long wait times to obtain an appointment with the pain clinic, so s/he asked administrative staff to compile a list of all pain clinics in the region, regardless of whether they were affiliated with the hospital system in order to offer patients more options. One PCC indicated that it was difficult to quickly locate and review relevant information about chronic paint patients in the EHR. S/he used a function in the EHR intended for inpatients to compile a summary of each patient scheduled for the day. Knowledge of the resources available and how to best leverage them influences what data elements PCCs examine and what options they consider.

DISCUSSION

This study represents an in-depth exploration of PCC decision making related to chronic pain, a topic that has been primarily studied using survey techniques (Breuer, Cruciani, Portenoy 2010; Green et al, 2002; Turk, Bordy, Okifuji, 1994). This is the first in-depth analysis conducted after publication of CDC guidelines intended to influence PCC decision making. Articulating the macrognitive functions, common anchors, and clinician factors that influence framing in chronic pain management has implications for supporting sensemaking in this environment. By shedding light on the data elements needed to form a frame, elaborate a frame, seek a frame, re-frame, question a frame, and compare frames, potential improvements in EHR design become clear. For example, an integrated pain display might be structured around common anchors. Increased interoperability of EHRs, particularly when it comes to sharing pain related information from specialists (i.e., emergency department, psychiatry) would reduced uncertainty for PCCs. Furthermore, the EHR provides a natural platform for integrating guidelines into practice.

This characterization of chronic pain management also has implications for guideline development. It is important to avoid rigid decision rules given the complexity and dynamic nature of chronic pain management; yet, there may be strategies for guiding clinicians to appropriate anchors. Guidelines that emphasize appropriate anchors may lead to more standardized practice across PCCs. Although interpretation and judgment by individual PCCs would still be necessary, focusing on common anchors might increase quality of care as patients would be less likely to encounter different PCCs with disparate interpretations of the same guidelines. Training to support clinicians in managing patients with chronic pain might emphasize macrocognitive functions such as intepreting patient actions and motives, assessing patient condition as it relates to chronic pain, assessing risk, and working with patients to identify realistic goals. This might include educational materials and continuing education courses that include documentation of effective strategies and case studies.

The data frame theory allows for a more nuanced representation than more decompositional approaches because of the emphasis on the dynamic nature of sensemaking. Next steps for this project include examining patient cases over time to better understand what data elements lead to elaboration, questionning, and re-framing. Furthermore, we plan to expand the data collection to include interviews with PCCs across additional healthcare systems to explore the generalizability of these findings.

CONCLUSION

We found the data frame theory of sensemaking useful for the capturing complexity of chronic pain management in primary care. Findings from this study suggest that interventions to support PCCs in chronic pain management should consider the complexity, as well as the ambiguity and uncertainty associated with pain management. Interventions should focus on supporting sensemaking. Common clinican decision support elements such as algorithm driven decision rules and decontextualized clinical reminders are less likely to be effective.

ACKNOWLEDGMENTS

This project was supported by grant number R01HS023306 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

REFERENCES

Abbo E.D., Zhang Q., Zelder M., Huang E. (2008) The increasing number of clinical items addressed during the time of adult primary care visits. *Journal of General Internal Medicine 23*(12), 2058-2065.

American Society of Anesthesiologists Task Force on Chronic Pain Management. (2010). Practice guidelines for chronic pain management: an updated report by the American Society of Anesthesiologists Task Force on Chronic Pain Management and the American Society of Regional Anesthesia and Pain Medicine. *Anesthesiology*, *112*(4), 810.

Bair, M.J., Robinson, R.L., Katon, W., Kroenke, K. (2003). Depression and pain comorbidity: A literature review. *Archives of Internal Medicine, 163*, 2433–2445

Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: a case of the tail wagging the dog?. *British Medical Journal*, *322*(7294), 1115-1117.

Barker, S., & Moseley, G. (2016). The difficult problem: Chronic pain and the politics of care. *AQ-Australian Quarterly*, *87*(3), 8.

Breivik, H., Eisenberg, E., & O'Brien, T. (2013). The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. *BMC Public Health*, *13*(1), 1.

Breuer B, Cruciani R, Portenoy RK. (2010). Pain management by primary care physicians, pain physicians, chiropractors, and acupuncturists: A national survey. *Southern Medical Journal*, *103*(8), 738-747.

Chou, R. (2009). 2009 Clinical Guidelines from the American Pain Society and the American Academy of Pain Medicine on the use of chronic opioid therapy in chronic noncancer pain: what are the key messages for clinical practice?. *Polish Archives of Internal Medicine*, *119*(7–8), 469-477.

Committee on Advancing Pain Research Care Education. (2011). Institute of Medicine. *Relieving pain in America: A blueprint for transforming prevention, care, education, and research*. Washington DC : The National Academies Press.

Compton, W. M., Jones, C. M., & Baldwin, G. T. (2016). Relationship between nonmedical prescription-opioid use and heroin use. New England Journal of Medicine, 374(2), 154-163.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge : Mit Press.

Gaskin, D. J., & Richard, P. (2012). The economic costs of pain in the United States. *The Journal of Pain*, *13*(8), 715-724.

Green, C.R., Wheeler, J.R.C., LaPorte, F., Marchant, B., Guerrero, E. (2002). How well is chronic pain managed? Who does it well? *Pain Medicine*, *3*(1):56-65.

Gureje, O., Von Korff, M., Simon, G.E., Gater, R. (1998). Persistent pain and well-being: A World Health Organization study in primary care. *JAMA*, *280*, 147–151.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. In Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making (pp. 113-155). New York: Lawrence Erlbaum.

Leverence, R.R., Williams, R.L., Potter, M., Fernald, D., Unverzagt, M., Pace, W., Parnes, B., Daniels, E., Skipper, B., Volk, R.J., Brown, A.E., Rhyne, R.L., on behalf of PNc. (2011) Chronic non-cancer pain: A siren for primary care – a report from the primary care multiethnic network (prime net). *The Journal of the American Board of Family Medicine*, *24*(5), 551-561.

Manchikanti, L., Abdi, S., Atluri, S., Balog, C., Benyamin, R., Boswell, M., Brown, K., Bruel, B., Bryce, D., Burks, P., Burton, A., Calodney, A., Caraway, D., Cash, K., Christo, P., et al (2012) American society of interventional pain physicians (asipp) guidelines for responsible opioid prescribing in chronic non-cancer pain: Part 2-guidance. (2012) *Pain Physician. 15*(3 suppl), S67-116.

Martell, B.A., O'Connor, P.G., Kerns, R.D., Becker, W.C., Morales, K.H., Kosten, T.R., et al. (2007). Systematic Review: Opioid Treatment for Chronic Back Pain: Prevalence, Efficacy, and Association with Addiction. *Annals of Internal Medicine*, *146*, 116-127. doi: 10.7326/0003-4819-146-2-200701160-00006.

Max, M. B., Donovan, M., Miaskowski, C. A., Ward, S. E., Gordon, D., Bookbinder, M., ... & Janjan, N. (1995). Quality improvement guidelines for the treatment of acute pain and cancer pain. *JAMA*, *274*(23), 1874-1880.

Muhuri PK, Gfroerer JC, Davies MC; Substance Abuse and Mental Health Services Administration. Associations of nonmedical pain reliever use and initiation of heroin use in the United States. *CBHSQ Data Review*. http://www.samhsa.gov/data/2k13/DataReview/DR006/nonmedical-pain-reliever-use-2013.pdf. Published August 2013.

National drug control strategy: data supplement (2014). Washington, DC: Office of National Drug Control Policy Executive, Office of the President of the United States, 2014 (https:/ / www .whitehouse .gov/ sites/ default/files/ondcp/ policy-and-research/ ndcs_data_supplement_2014 .pdf).

Østbye T, Yarnall KSH, Krause KM, Pollak KI, Gradison M, Michener JL. (2005). Is there time for management of patients with chronic diseases in primary care? *The Annals of Family Medicine*, *3*(3), 209-214.

Rosenblum, A., Joseph, H., Fong, C., Kipnis, S., Cleland, C., Portenoy, R.K. (2003). Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. *JAMA*, *289*, 2370–2378.

Turk, D.C., Brody, M.C., Okifuji, E.A. (1994) Physicians' attitudes and practices regarding the long-term prescribing of opioids for non-cancer pain. *Pain*, *59*(2), 201-208.

Unick G.J., Rosenblum D., Mars S., Ciccarone D. (2013). Intertwined epidemics: national demographic trends in hospitalizations for heroin- and opioid-related overdoses, 1993–2009. *Public Library of Science One*, *8*(2), e54496.

Upshur C.C., Luckmann R.S., Savageau J.A. (Jun 2006) Primary care provider concerns about management of chronic pain in community clinic populations. *Journal of General Internal Medicine*, *21*(6), 652-655.

Yeo, S. N., & Tay, K. H. (2009). Pain prevalence in Singapore. *Annals Academy of Medicine Singapore*, *38*(11), 937.

Exploring Information Flow: Consultations between Primary and Specialty Care Clinics

Laura G. MILITELLO[a], April W. SAVOY[bc], Brian PORTER[b],
Mindy FLANAGAN[bc], Justina WU[d], Jasma ADAMS[b], Shakaib REHMAN[e],
Hamed ABBASZADEGAN[e], Michael WEINER[bcf]

[a]*Applied Decision Science, LLC*

[b]*Center for Health Information and Communication, U.S. Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service CIN 13-416, Richard L. Roudebush VA Medical Center*[c]*Indiana University*

[d]*VA Palo Alto Health Care System, U.S. Department of Veterans Affairs*

[e]*Phoenix VA Health Care System, U.S. Department of Veterans Affairs*

[f]*Regenstrief Institute, Inc.*

ABSTRACT

Consultations are a critical part of healthcare for many patients. Coordination between primary care and specialty clinics is challenging because information must be shared across clinics, roles, and time. We conducted a study of the consultations process in the U.S. Department of Veterans Affairs. Using interviews, observations, and document review, we identified limitations in current documentation of information flow. We offer a descriptive model of information flow in the consultations process as a more ecological basis for improvements to technological support via the electronic health record, process design, and training.

KEYWORDS

Coordination; Health; Consultations; Information Flow; Ecological Design

INTRODUCTION

Consultations in healthcare impose macrocognitive challenges for individuals, as well as the larger system. Consultations require complex coordination across clinics and over time, involving multiple clinicians and staff at various levels. Primary care clinicians regularly refer patients for consultations with specialists for both standard and critical aspects of care. Although referrals are a common and important component of care, they also represent a common point of failure. For example, a referral may be cancelled or lost so that an appointment is never made, or a patient may be lost to follow-up after a consultant's initial evaluation.

Referrals as a point of failure in the U.S. health system have been in the spotlight as recent news articles highlighting long wait times in U.S. Veterans Health Administration (VHA) medical centers (Oppel, 2015). One report suggests that 30% of referrals in VHA medical centers are cancelled (Singh, et al, 2011). A 2014 survey of physician appointment wait times reveals similar issues in private health systems, documenting average wait times ranging from 5 days to 72 days for specialists (Hawkins, 2014).

Electronic health records have the potential to support information flow and coordination of care in the context of consultations; yet, studies find that electronic health records may fall short in this regard. Cognitive engineering suggests that mismatches between the way in which work occurs (i.e., the natural system) and the way it is represented in the electronic health record (i.e., the technology system) can result in unnecessary complexity and confusion (Miller & Militello, 2015). In this article, we use the VHA consultation process as a case study, highlighting important mismatches between documented representations of information flow and observed information flow, and offer an alternative model of information flow that can serve as a foundation for the design of tools and processes to support consultation management.

METHODS

As part of a larger project to explore barriers and facilitators to effective consultations, we conducted observations and interviews at two VHA medical centers. After the initial interviews and observations were completed and analyzed, we conducted a second set of interviews focusing primarily on tracking consultations from the perspective of primary care. In addition, we conducted a document review to identify documented consultations information flow.

Participants

Participants included VA staff members (physicians, physician assistants, nurses, and other staff who support the consultation process) involved in managing consultations between primary care and subspecialty clinics. Specialty care clinicians were recruited from two VA medical centers, targeting six service areas: mental health, oncology, orthopedics, rheumatology, cardiology, and ophthalmology. The VA medical centers were located in geographically disparate regions. Both included an urban tertiary care center and affiliated primary and secondary care facilities (e.g., outpatient clinics, nursing facilities, and domiciliary rehabilitation facilities).

We interviewed 20 VA staff members at Site 1 and 22 at Site 2. We observed 18 VA staff members at Site 1 and 20 at Site 2. (See **Table 1**.) Participants were purposively recruited to provide diversity among key roles within the primary care and specialty care referral process. Ethics approval was obtained from the VA Central Institutional Review Board prior to recruitment for this study.

Table 1. Participants by specialty and site

Setting	Site	Observations	Initial Interviews	Follow Up Interviews
Primary care	A	8	8	10
Primary care	B	9	8	10
Specialty care	A	10	12	
Specialty care	B	11	14	

Observations

Individual data collectors observed clinicians during clinical encounters, to understand how consultation requests, triage (decision to accept, and prioritization of consultation requests), consultation, and documentation occur in the context of clinic workflow. Each clinician was observed in his or her clinic for a half-day session (approximately four hours). Observers obtained written assent from patients to observe their clinician during the visit. Observers asked clinicians questions opportunistically during breaks to clarify aspects of the consultation process, so as not to disrupt the clinical work. Each data collector recorded handwritten field notes, and these notes were transcribed prior to analysis. Each data collector then revisited his or her own field notes to characterize information flow and identify vignettes illustrating key phenomena related to barriers, facilitators, and workarounds.

Initial Interviews

We conducted semi-structured interviews (Saleem et al, 2009) to understand barriers and facilitators to effective consultations from both the primary and specialty care perspectives. The interviews were designed so that the same core topics were addressed in each interview, while still allowing the flexibility to explore the perspective and experiences of each interviewee. Questions for primary care interviews focused on challenges in entering consultation requests, reasons for rejected consultation requests, and challenges in receiving consultants' findings. Questions for specialty care interviews focused on the process for handling new consultation requests, triage processes, and strategies for meeting time goals for processing consultations. Both primary care clinicians and specialty care clinicians were asked about coordination strategies and challenges, and suggestions for improving the process. Interviews were conducted by either a single researcher or a team of two researchers (an interviewer and a note taker). Participants were interviewed individually except in a few instances, where two participants were recruited at the same time and expressed a preference to be interviewed together. Interviews were conducted in locations chosen by the participants, usually an office, similar work space, or conference room. Interviews lasted approximately 30-45 minutes, and were audio-recorded, as permitted by the participants. Audio-recorded interviews were transcribed and checked for accuracy. Any personally identifying information was removed prior to analysis.

Follow-Up Interviews

We conducted a second set of semi-structured interviews to further understand how consultations are tracked from the primary care perspective. Questions focused on how referrers track the progress of consultations and communicate with the consulting clinician about the referral. Interviews were conducted either in person at a place of the interviewee's choosing or over the telephone. Interviews lasted approximately 30 to 45 minutes and were audio recorded as permitted by the participant. Audio recordings were transcribed, checked for accuracy, and any identifying patient or clinician information was removed prior to analysis.

Document Review

To understand how consultation information flow is represented to VA staff, we conducted a document analysis including *What Every VA Clinician and Resident Needs to Know about Consults (*VA Consult Steering Committee, 2016), and the *Consult Request/Tracking User Manual Version 3.0 (*Department of Veterans Affairs Office of Information and Technology Product Development, 2016*).* Document analysis is a strategy commonly used by cognitive engineers to explore an organization's functional purposes, values and priority measures, specific functions, and tools used to accomplish functions (Naikar, Hopcroft, & Moylan, 2005). For this project, we were particularly interested in documents available to clinicians, administrators, and software developers describing the consultations process. Both of these documents are publicly available.

Analysis

One investigator (LM) reviewed field notes from the observations to develop a representation of observed information flow. She drafted an information flow model based on these data. Three other investigators who had participated in observations (MF, BP, JW) suggested refinements. The model was then presented to two VA clinicians and the broader research team for additional feedback.

A team of four investigators (JA. MF, LM, JW) conducted a qualitative thematic analysis (Green & Thorogood, 2013) of the transcribed interview data. Using a thematic approach, the four analysis team members (non-clinician, behavioral scientists) independently reviewed transcripts constructing categories to describe the data. The analysis team used an iterative consensus-based approach to create a coding list consisting of subcategories of the consultation process (e.g., decision to request a consultation, consultation submission steps, tracking consultations, etc.) Next, the analysis team plus one clinician used the list of categories to code a sample of randomly selected transcripts. This process led to discussion and revisions to the codebook. The final codebook included codes related to the consultation process, the technology used to support the process, and role of patients in the process. Next, the four analysis team members coded remaining transcripts, continuing to refine the categories to capture subtle nuances in the consultation process, and reaching consensus for each code.

FINDINGS

Figure 1 depicts the documented information flow for consultations in the VHA found in the *Consult Request/Tracking User Manual Version 3.0.* In this representation, the clinician orders a consultation in the VA's electronic health record system—the Computerized Patient Record System (CPRS)—and the consultation service receives a printed copy of the consultation order. The consultation service either accepts the consultation request using the *receive* action in CPRS, or *discontinues* or *cancels* the consultation via the CPRS interface. After the consultant evaluates the patient, results from the evaluation are entered and signed in CPRS. The referring clinician receives a notification through the CPRS interface that the consultation is complete and a report is available.

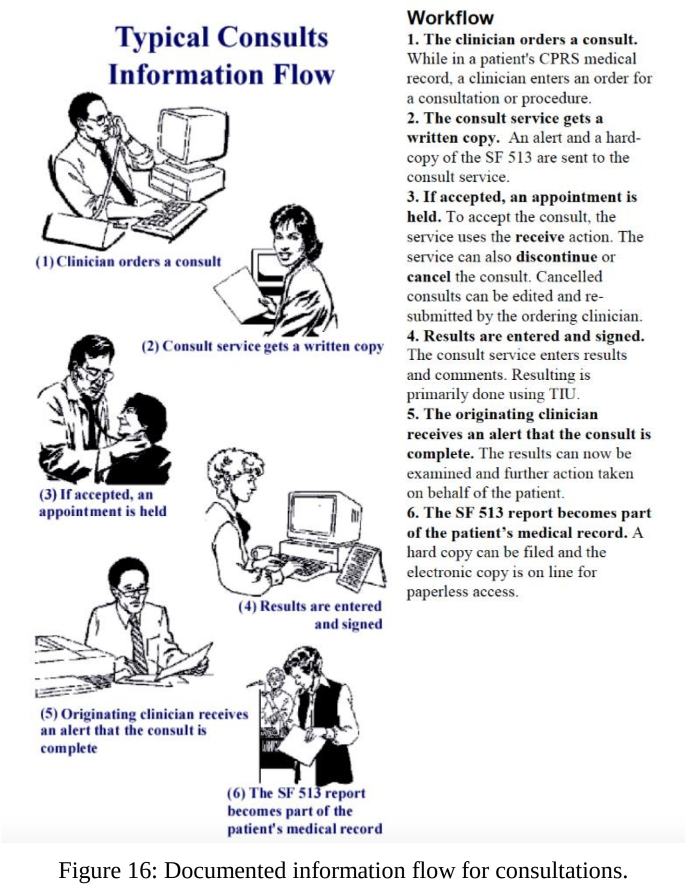


Figure 16: Documented information flow for consultations.

Based on our observations, we found this representation to be an outdated and oversimplified representation of information flow. The representation depicts outdated components such as a paper form (e.g., SF 513) and a TIU, which refers to a set of software tools not mentioned by a single participant in our study. Furthermore, observations highlight notable variation and complexity not depicted in this representation. **Figure 2** depicts a descriptive model representing steps of the consultation process as observed: referral, triage, scheduling, consultation, and follow-up. Below each, we highlight the roles and responsibilities for specific tasks within the consultation process, critical electronic tools and forms, other tools and forms, and common points of breakdown in the process. Note : The VA has two different interfaces that draw from the same underlying database of medical records. CPRS is the newer interface and commonly used to access patient data and take clinical actions (i.e,. order consultations and medications). The older interface, commonly referred to as VistA and/or DHCP, is used to access some information and actions (i.e., reporting functions).

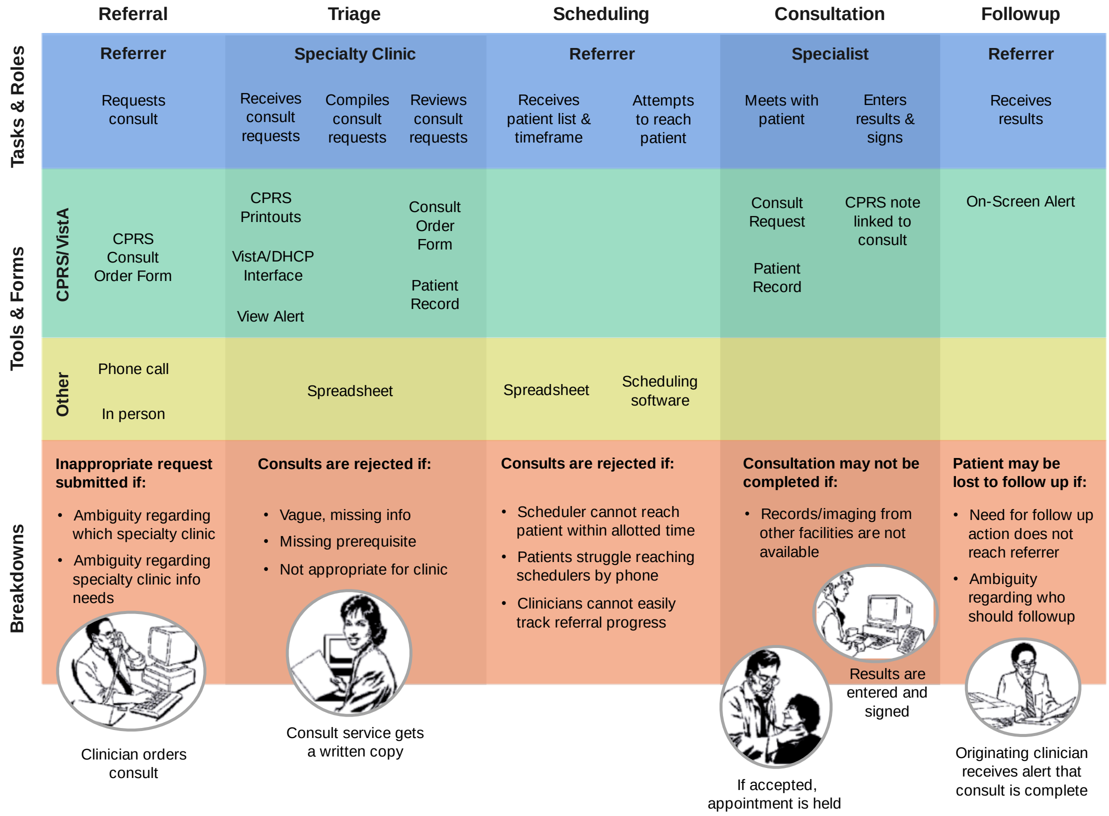


Figure 17: Descriptive information flow model

We note five important limitations in **Figure 1**: scheduling components are missing, information transfers appear to be standardized and well understood, all information relevant to consultations appears to be efficiently tracked from beginning to end, CPRS appears to support critical communication about consultations, and follow-up is deemphasized. We discuss each in limitation in the context of our interview and observation data.

Scheduling is a common point of breakdown

First, and most notably, the scheduling component is not represented in **Figure 1** but is a common point of breakdown in the consultation process. In many cases, the scheduler is unable to reach the patient to schedule an appointment within the allotted timeframe. Even if the scheduler reaches the patient, the patient may not be available to visit the clinic in the desired time frame. In some clinics, patients are unable to reach the scheduler by telephone to return a missed call or to reschedule an appointment. Some interviewees reported that information did not always flow easily between the triage team and scheduling. Because the scheduling and triage functions might fall under different chains of command, and the two functions do not share a software interface, the stage is set for information to fall through the cracks at this point in the process.

Information transfers are not standardized or well understood

Second, **Figure 1** implies that information transfers are standardized and well understood. Information appears to flow from one stage of the process to another in a straightforward and predictable manner. In contrast, we found considerable variation in roles and information ownership from clinic to clinic. For example, referrers reported that knowing which specialty clinic is most appropriate for a patient's specific medical condition can be challenging. When faced with this type of uncertainty, it is not always clear how to obtain additional information or how best to reach someone in the relevant specialty clinics who can help with the decision. Thus, a referral is often ordered using the referrer's best guess as to which specialty clinic is most relevant, relying partly on trial and error to learn the system. Ordering a referral from the wrong specialty can result in delayed care for patients, and cause added workload for the healthcare team. For each referral ordered, the specialty clinic must review medical-records. If a specialty clinic deems a referral request inappropriate for their clinic, the specialty personnel cancel or discontinue the request, and the referring clinician is notified through CPRS. The referring clinician must then determine how to proceed. Common actions are to submit another request to the same clinic with additional clarification about the clinical situation, to manage the patient in primary care, or to request a referral to a different specialty clinic.

A second example of undocumented variability in information transfers is during the triage process. We observed a wide range of variability in how triage was handled and by whom. Specifically, we observed information transfer from referral to triage using paper forms, via CPRS, and by telephone. In some clinics, referral requests appear on a printer, and are then routed by a staff member before being triaged by a clinician. In other clinics, a clinician may obtain the referral requests directly from CPRS or VistA, and then triage them. In yet other situations, the consulting clinic may be contacted by the referrer via telephone, particularly when a consultation is urgent. After the referral is received, some clinics assign one person to conduct triage for the entire clinic, resulting in relatively predictable priorities and responses. In other cases, however, the triage responsibility rotates, often resulting in a less standardized process. Variation in the triage process is notable because this is a common point of breakdown. When primary care clinicians have difficulty anticipating what information a specific specialty clinic will need, they are more likely to submit requests that are missing information or are inappropriate for that specialty clinic.

Tracking is not seamless

Third, **Figure 1** seems to suggest that all consultations are tracked seamlessly from beginning to end. Indeed, some interviewees reported that CPRS effectively tracks consultations, so that they do not need to initiate and maintain a manual tracking process. In spite of this perception, our observations highlighted places where information is stored and shared outside CPRS, introducing opportunities for error and lost information. These types of discontinuities in information flow are exacerbated with non-VA consultations, for which multiple disparate information systems are used.

As noted, information transfer from referral to triage sometimes happens via "non-standard" means outside CPRS. In some situations, referrers contact the consulting clinic via phone or face-to-face interaction. Specialists report that real-time discussion of patients, particularly in urgent or critical situations, is encouraged, and is required for emergencies. However, in some cases, after this initial discussion, the referrer forgets to enter the request into CPRS—a necessary process for documentation and follow-up. Thus, in these instances, some of the most critical consultations are not initiated or tracked properly in CPRS.

Some specialty clinics use an electronic spreadsheet such as Microsoft Excel to compile a complete list of all consultation requests received via paper, electronic communications, and interpersonal communications. Many interviewees reported that this is a labor-intensive process. Although VHA documentation indicates that CPRS users can display a list of all consultation orders for a specific clinic (Department of Veterans Affairs Office of Information and Technology Product Development, February 2016, p. 101), some of our interviewees were unaware of this function.

During the consulting clinic's triage, the spreadsheet is refined to reflect accepted requests. It is then sent to schedulers to contact patients. Many VA facilities use scheduling software that is separate from CPRS/VistA and has a completely different user interface. Although these different software packages share the same back-end database; in practice, the schedulers and clinicians do not have ready access to a clear view of the information available to the other. Interviewees reported a range of problems related to poor information flow with scheduling, including duplicate appointments in the same time slot for specialists, and difficulty tracking the status of a consultation after the initial referral request is submitted.

For consultations outside the VA health system, even more information transitions occur inside and outside CPRS. We observed the information transitions required by health systems that use different electronic health record systems, including the use of faxes and email to request referrals. Another type of discontinuity relates to the interface between private and public health systems. The VA requires that private health facilities be approved to provide specific health services to Veterans, and maintains a list of private health services available. However, this critical information is not available to all clinicians. As a result, care may be delayed when a clinician from a private facility requests permission to provide care that has already been approved, or when a VA clinician is uncertain about which private clinics are available to his or her patients.

In many cases, findings and recommendations from private facilities are not readily available in CPRS. Systems to support referrers in tracking and receiving this information from consultants outside the VA vary. Some clinics have staff dedicated to coordinating and tracking consultations with private facilities. Others rely on general VA processes. Those relying on general VA processes report a significant time lag between the consultation visit and when the consultant's findings are available in CPRS. This is because the consultant's report must be scanned as a portable document formant (PDF) and uploaded to CPRS, then correctly linked to a note in CPRS. If the report is not correctly linked to the note, the clinician may be unaware that it is available or unable to locate it. In addition, scanned PDFs may contain many pages of unsearchable text, hindering identification of important information. Interviewees also report that imaging from clinicians outside the VA may be difficult to find and view, resulting in delays in care, and unnecessary repeated imaging, which creates unnecessary health hazards and inconvenience for patients, and costs for them as well as other payers of care.

Documentation does not equal communication

Fourth, **Figure 1** implies that CPRS supports critical communication about consultations. Our observations and interviews suggest that both primary care clinicians and specialists perceive current approaches to communication to be a primary barrier to effective and efficient consultation. In the current approach, there is an implication that documentation of care and status in CPRS should serve as communication across clinics. Therefore, limited time is apportioned to be available for phone calls, instant messaging, or face-to-face communication with other clinicians. Even when one clinician makes time for a call, a timely response is often not received. Yet, many of the participants in our study noted how much more smoothly consultations work when there are personal relationships and opportunities to communicate real-time outside of CPRS. An opportunity to ask a clarifying question can prevent a cycle of referral request, cancellation, and re-referral. An opportunity to mention a sensitive issue that does not fit neatly into the consult referral template such as dementia or suspected elder abuse can positively influence the care received by a patient. Without the opportunity for direct communication, simple questions end up a part of the patient's medical record, exacerbating the difficulty of finding relevant information when needed. As one interviewee observed:

> "Well, yes. So there is one thing that's kind of a glaring deficiency in CPRS. There is no messaging system. So we use in CPRS … as a messaging system, which is not appropriate. We'll put in what do you think about x, y, z, or please call x, y, z in a patient's chart. That really doesn't need to go into a medical record. So having some sort of communication system that is kind of parallel to CPRS but not part of their kind of official medical record would be helpful."

Thus, even clinicians' recognition of the communication problem, and potential solutions, appear unrelated to their own approaches to communicating with colleagues.

Follow up is a critical component of consultations

Fifth, **Figure 1** provides no suggestion about how the need for follow-up action is communicated to the referrer. Our interview data suggest that many primary care clinicians rely on the "View Alerts" function (on-screen informational message lists) in CPRS to discover that a consultation is complete and a report is available. Many, however, report that the large number of non-relevant messages awaiting review often hide important messages in the list. Therefore, the referrer may not realize that follow-up action is required. The need for action may not be discovered until the patient's next visit for primary care, when the clinician will review the patient's medical record and discover relevant notes and reports from the specialty clinic.

DISCUSSION

Oversimplification of the consultations process can have negative consequences. The documented information flow is likely to be the starting point for clinicians and staff in the VA for learning about the consultations process and forming mental models for working effectively within the system. A misleading view of information flow can hinder troubleshooting when things go wrong. Without a clear understanding of how one's own role fits into the larger flow of information, anticipating problems and discovering the most efficient ways of working within the system becomes difficult. The importance of developing mental models that closely align with the natural world for making good judgments and actions, particularly in a complex socio-technical system, has been repeatedly observed and discussed (Bennett & Flach, 2011).

Perhaps even more troubling, however, are the implications of oversimplification from a systems view. The documented representation of information flow is likely to inform technology design, process improvement initiatives, and even staffing decisions. At a systems level, a distorted view of information flow and the related roles and tasks can hinder the understanding of current challenges, and the pathway towards improvement. The consequences of designing technologies and systems using strategies that do not coincide with the natural ecology of work have been well documented. Cognitive engineering methods and models have been developed over the last 30 years, specifically to aid scientists and technologists in eliciting difficult-to-articulate aspects of work, observing work in situ, and developing representations to support systems level analysis and design.

In the context of this study, we found that the information flow representation (**Figure 1**) de-emphasized information transfer between people and technologies. This idealized flow uses passive voice and focuses only on a subset of actions that *should* occur. This idealized representation leaves out important information about roles and information transfers, and seems to suggest that information tracking and communication are a natural byproduct of the process.

The more ecologically oriented, descriptive model (**Figure 2**) better represents the complexity of the system, providing a foundation from which to begin to explore potential improvements to the system. The Tools & Forms section of **Figure 2** suggests opportunities to support information handoff and perhaps eliminate or reduce transitions inside and outside CPRS. The Common Breakdowns section of **Figure 2** highlights other important leverage points for process improvement, such as strategies to aid primary care clinicians in better understanding specialty clinics' information needs, increased transparency into the scheduling function for both clinicians and patients, improved tracking and notification functions, and the need for communication between and across roles in the system *outside CPRS*.

CONCLUSIONS

Consultations are an important part of healthcare for many patients. Because consultations require information transfer across clinics, strategies for representing information flow are particularly important in monitoring the effectiveness of the system and suggesting improvements. Ethnographic observation and interviews are powerful tools for understanding workflow “as is”, rather than “as supposed to be.” In this study, our observations and interviews revealed five important limitations of the documented workflow: scheduling is omitted, information transfers appear to be standardized and wellunderstood, all information relevant to consultations appears to be efficiently tracked from beginning to end, CPRS appears to support critical communication about consultations, and follow-up is de-emphasized. We offer a more ecologically oriented, descriptive model of information flow and highlight common breakdowns.

Although this research occurred in the context of the VA health system in the U.S., we anticipate that these findings will be relevant for other medical instituations coordinating care between primary care and specialty services.

ACKNOWLEDGMENTS

This work was supported by VA HSR&D Grant No. IIR #12-102 and the Center for Health Information and Communication, Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service CIN 13-416. Views expressed in this article are those of the authors and do not necessarily represent the views of the U.S. Department of Veterans Affairs or the U.S. government.

REFERENCES

Bennett, K. B., & Flach, J. M. (2011). *Display and interface design. Subtle Science and Exact Art*. Boca Raton, FL: CRC Press.

Department of Veterans Affairs Office of Information & Technology Product Development (February 2016). *Consult/request tracking user manual*. Washington DC: Department of Veterans Affairs . https://www.va.gov/vdl/documents/Clinical/CPRS-Consult_Request_Tracking/consum.pdf

Green, J., & Thorogood, N. (2013). Qualitative methods for health research. Los Angeles: Sage.

Merritt Hawkins (2014). Physician appointment wait times and Medicaid and Medicare Acceptance Rates. Irving, TX: Merritt Hawkins.

Miller, A., & Militello, L.G. (2015). The role of cognitive engineering in improving clinical decision support. In A.M. Bisantz, C.M. Burns, R.J. Fairbanks (eds.) *Cognitive Engineering Applications in Health Care*. New York: CRC Press.

Oppel, R.A. (June 20, 2015). Wait lists grow as many more veterans seek care and funding falls short. *New York Times. https://www.nytimes.com/2015/06/21/us/wait-lists-grow-as-many-more-veterans-seek-care-and-funding-falls-far-short.html?_r=0*

Naikar, N., Hopcroft, R. and Moylan, A. (2005) Work domain analysis: Theoretical concepts and methodology. No. DSTO-TR-1665. Victoria : Defence Science and Technology Organisation Victoria (Australia) Air Operation Division.

Saleem, J. J., Russ, A. L., Justice, C. F., Hagg, H., Ebright, P. R., Woodbridge, P. A., & Doebbeling, B. N. (2009). Exploring the persistence of paper with the electronic health record. *International journal of medical informatics*, *78*(9), 618-628.

Singh, H., Esquivel, A., Sittig, D. F., Murphy, D., Kadiyala, H., Schiesser, R., ... & Petersen, L. A. (2011). Follow-up actions on electronic referral communication in a multispecialty outpatient setting. *Journal of general internal medicine*, *26*(1), 64-69.

Decision processes in action at sea, a methodological challenge for real world research

Sylvain MONDON[a] and Anne MARCHAIS-ROUBELAT[a]
[a]*CNAM, Laboratoire Interdisciplinaire de Recherches en Sciences de l'Action, Paris.*
Sylvain.Mondon.auditeur@lecnam.net, *Anne.Roubelat@lecnam.net*.

ABSTRACT
Relationships between decision process and action in real world conditions emerge as a key issue in elite sport as well as in organization studies. When pressure is high and environment is evolving rapidly, acting despite a limited situation awareness is common. In such stressed situations, the observation of phenomenon raises methodological difficulties such as the access to information and its validation with regard both to the decision-making process of the decision-maker and concurrently to the course of action. Studying in real time decision processes of a professional skipper at work during an ocean yacht race from the viewpoint of the router embedded in the skipper's team provides an opportunity to go forward in understanding relationships between decision processes and action in real world when phase shifts occur. The methodological choices are discussed and research directions for further real world researches are proposed.

KEYWORDS
Decision Making, Earth and Atmospheric Sciences, Uncertainty Management, Team and Organizational Factors in Complex Cognitive Work, Elite Sport.

INTRODUCTION

Relationship between decision processes and action in real world conditions emerge as a key issue in elite sport (Macquet & Kragba, 2015) as well as in organization studies (Marchais-Roubelat, 2012; McAndrew & Gore, 2015). During past decades laboratory approaches showed a variety of human behaviours and preferences in controlled experiments (Kahneman, 2003). Even if laboratory experiments can be very sophisticated as the one described by Brandt, Lachter, Battiste and Johnson for pilots (2015), some dimensions can't be replicated accurately enough to stand for real world. The full complexity of real world action involving multiple goals, time stress, uncertainty management, complex task is still not fully achievable in laboratory (Kahneman & Klein, 2009). In order to better understand relationships between decision processes and action, it is essential from our viewpoint, not to postulate anything about the relationship itself. Thus, we should be careful to incorporate no tacit premise in our approach. Especially, the framing of the research should not influence the way either decision processes or action are considered. A promising way to gain significant insights about relationships between decision processes and action is to work on real world by means of real time observation. Our guess is that stressed situations are likely to emphasise the relationships between decision processes and action because high time stress prevents from delaying acts. In such a stressed situation, it is common to act with limited situation awareness as shown by Klein, Orasanu, Calderwood and Zsambok (1993). The focus on real world practitioners adopted here meets the one promoted by naturalistic decision making (NDM) community for nearly 30 years (Gore, Flin, Stanton & Wong, 2015). We have chosen a field relevant for both elite sport and organizational science: the action of skipper during ocean yacht race. It has been chosen because it consists of an organisation (skipper and his or her team) involved in an action where decisions must be made and implemented to handle the boat and manage its trajectory. So we can assume that during the race few decisions should be made in a given period of time. Our intention is to develop a fitted methodology to perform an inquiry in the field of yacht race in real time in order to better understand relationship between decision processes and action at sea with a view to improve training to get better performance. Gore *et al.* (2015) recalled that the area of elite sport is being investigated only for few years. So, our approach represents also an opportunity to explore a new field, highly compatible with NDM requirements, which could provide new "concrete examples from a specific profession" as recommended by Klein (2015). Even if "methodological advances in accessing expertise have gained respect and validity", NDM practitioners "recognize that the frameworks, models, and methods" used "have their limitations" (Gore *et al.*, 2015). So, as we faced similar limitations, the developments proposed hereafter may be relevant for NDM practitioners. The methodological difficulties enhanced by NDM researchers are first highlighted when exploring the field in the context of action. Then, we discuss the design of a collaborative partnership for such a research during an ocean race, in terms of research embedment and of research methodological issues. First results and preliminary analysis are exposed. Finally, expected further outcomes are discussed and leads for further real world researches are proposed.

METHODOLOGICAL CHALLENGE OF ACCESS TO INFORMATION: ISSUES FOR OCEAN YACHT RACE DECISION AND ACTION PROCESSES

Developing a method aiming at performing a kind of “cognitive task analysis” (Crandall, Klein & Hoffman, 2006) to elicit knowledge and analyse data in order to represent “real world” relationships between decision processes and action in the field of ocean yacht race raises several methodological difficulties. Those methodological difficulties form the challenge of accessing relevant information on a convenient way. This challenge is composed of several difficulties of different levels: duration, distance, acceptability, stakes and diversity of sources. Most of them have been dealt for other fields by NDM community in earliest studies for outdoors operations (Klein *et al.*, 1993; Zsambok & Klein, 1997), for elite sport (Macquet, 2010, Macquet, Ferrand & Stanton, 2015) and for mountain expeditions (Allard-Poesi & Giordano, 2015).

Duration and phases of the action

Ocean yacht races usually last several days. Even if we expect some decision to be made in this period of time, the exact time of occurring can't be presumed. So the study must cover at least the entire period of the race not to miss any important phenomenon. Except in the case of mountaineering (Allard-Poesi & Giordano, 2015), most of the elite sports already studied by NDM practitioners were focused only on a short period of time. Furthermore, the sailors are 24h a day at work. Time is a key issue in organization (Orlikowski & Yates, 2002) as well as for elite athletes in managing competing activities (Macquet & Skalej, 2015). Sailors must dedicate their time to competing tasks like tuning, steering, computing route, analysing situation, observing environment, communication to media, checking the boat, repairing broken pieces or sleeping. The researchers must be ready to capture phenomenon which can occur at any time during any activity (Klein *et al.,* 1993). Because of its duration, ocean race as an action process includes many decisions, which can be connected with different phases of the race. As a result, the connection between these decisions and phase shifts in the action appear key issues for research in the field.

Distance to the action

In the case of sailing across an ocean, the size of natural field is only limited by continents and islands. Keeping a continuous access to action wherever the boat is, is a difficulty as soon as the boat leaves the dock. The idea to shorten distance between action and researcher is a leading idea in the earliest studies developed by NDM community about fireground commanders (Klein *et al.*, 1993). The issue is to be as close as possible to the action where decision are supposed to be made in order to get relevant information for subsequent analysis and interviews by performing direct observation in real time if possible. When the distance becomes a problem, means of communication may compensate it as exposed by Orasanu and Fisher (1997) in the field of spaceship crew and by Allard-Poesi and Giordano (2015) in the field of mountaineering. Those studies both show that means of communication (emails and phone calls) are essential to keep researchers in touch with practitioners when they can't be at the same place. During an ocean race, although information technology shortens the distance between the sailor and his or her team, the skipper has a full autonomy of decision to influence the course of the action.

Awareness of what is at stake

The professional skippers of the biggest modern yachts crossing ocean and rounding the world are very experienced sailors. They are considered as experts by others in the sense given by Klein (Klein, 1998). In such a technical complex field, being aware of what is at stake for the expert is a difficulty for the researcher. This is all the more critical when action is going on because direct interview can't be dedicated to long explanation. The researcher faces also the difficulty of knowing what are the meaningful cues to be noticed and how to seize them. The issue of understanding what is at stake is an essential point to get and validate information and to avoid later misinterpretation (Klein *et al.,* 1993). For ocean race, four main domains of interest can be identified to define what is at stake: boat handling, weather conditions, information and telecommunication systems, competition evolution.

Distributed situation awareness

During yacht race, information can be collected from a lot of different sources. Each source provides data from specific viewpoint useful to document at least one side of the action. Describing the side of the environment can be done by quantitative data as well as by qualitative data. Quantitative raw data can be provided by measurements (onboard sensors, remote sensing, buoy, weather stations) or by model analysis (national weather services). Qualitative data can be provided by weather bulletin, comprehension of the router, perception of skippers. Describing what is happening in term of action at race level can be informed by general trajectories of the whole fleet provided by race organizer, by detailed trajectory of a given boat provided by one team and by high frequency onboard monitoring of the route. These quantitative data sets can be completed by qualitative statements and comments from individuals or organizations (ex. press release). Taking into consideration “both human and technical agents as well as the way they interact” is the basis of distributed situation awareness (DSA) described by Stanton (Stanton, 2016). DSA is a critical issue for the athletes and coaches (Macquet & Stanton, 2014) and also for researcher analysis (Macquet, 2013; Macquet *et al.*, 2015). In ocean race, data provide a double perspective. On the one hand, data serve as a decision support system to help the sailor to make decisions to create new phases in the action processes. On the other hand, data are benchmarks for researchers to assess the decision process and action.

Acceptability of the research

Another crucial point for researcher interested in "real world" is to make the research acceptable for the field he or she would like to study. To that respect, the first point to deal with is to get an agreement in principle from a skipper to be studied at work in the course of action during a yacht race. As previously recalled time management is a critical point for skipper, the researcher position can't be too much intrusive. The difficulty for the researcher is to propose a research design light enough to be agreeable to the skipper and deep enough to enable fruitful inquiry: a balanced position must be negotiated. The purpose of the researcher is to have an access to relevant information in a transparent manner without significantly disturbing the way action is taking place.

DESIGNING A COLLABORATIVE PARTNERSHIP: PROCESS AND DISCUSSION

The research project intends to better understand the relationship between decision processes and action during ocean yacht race run by experienced sailors skipping multihull boats. Thanks to an 11 years experience in advising skippers between 2000 and 2011 in weather forecasting and routing, one of the two researchers involved in the project has a privileged access to the field. An agreement in principle has been given by one racing team to open the doors to a research project over several years. Since this point of departure in 2012, we have been progressively working out a convenient methodology to keep cautiously mutual trust and information exchange. We are presenting important issues encountered between 2012 and 2016 when the approach has been developed along with preparation, training and races ("Route du Rhum 2014" and "Transat Jacques Vabre 2015"). Then the focus is put on the next race "Transat Jacques Vabre 2017" for which we intend to implement an intensive observing period to highlight specific settings.

Embedded research

Even necessary, agreement in principle is not sufficient to perform an in depth inquiry. To our comprehension from preliminary discussions, the key issue is that the research has to be not only acceptable, but also promising enough for the skipper to sound useful in one way or another for short and long term. The time spent for the research from the viewpoint of the skipper is seen as a long term investment with regard to the expected outcome of the research. The second step is then to assess different possible positions for the researchers to do proper in depth research in a transparent manner and a convenient way for the team. Considering short term, the researchers and skipper acknowledged that the main challenge is to compensate in one way or another the fact that the sailors or team could spend time by discussing with researchers instead of paying attention only to they work in the course of action. Thus the deal concluded to that respect is that the researcher who is skilled in weather forecasting and routing can have a full access of the work of the team in real time in the condition that his or her skills can benefit to the team if he or she detects an occasion to do so. It is a kind of short term investment. The skipper expects that the skills of this researcher can save time to handle situation and complement distributed situation awareness. By doing this, the researcher is accepted as a special team member. Let's notice these short term as well as long term investments have no guaranty of success, they both are risky gambles for the skipper and for the researcher. Thanks to such a unique opportunity to be part of team, the research project is designed with one embedded researcher only whereas the second one keeps a distant outside viewpoint of the field and focuses on methodology and analysis, from common theoretical bases. The insider position is then balanced with the outsider one to reduce as much as possible biases in data collection as well as in subsequent analysis.

Time and ethics

Being inside a team is likely to have an influence on the action management. So there is a significant risk of disturbance and change in the focus of the sailors that has to be taken into account by the researcher. To that respect, our intention is to cover different stages of the preparation of the race by a compatible research activity so as to benefit each opportunity to learn more about the field and the skipper as well as to build trust all along the research project. It seems to us that the skipper should start the race (and we should start our intensive observation as well) with a mind free to focus entirely and deeply to his work (and we should do it as well). Furthermore, we would not accept a situation where our research is responsible for even the smallest trouble in the mind of the skipper. All discussions prior to the race must avoid to induce discomfort neither for the skipper nor for his or her team. This point is essential to our approach in order to neither influence the way of thinking nor the way of doing things of practitioners. This revealed to be crucial to create condition of acceptability of the research. To avoid any trouble four commitments have been made. The first commitment is to assure that the researchers behaviour will be as low profile as possible. The second commitment is to share the results with the sailors, as soon as they are robust enough. The third commitment is not to communicate personal information publicly and not to convey any information to other teams. The forth commitment is to mention any information in case it could be relevant for the race to the skipper, the router or the technical team.

Capacity and awareness

We have been implementing different tools and methods since 2013, bearing in mind the advice of Crandall *et al.* (2006, p. 143) by "understanding the way people think and reason in natural contexts, cognitive task analysis practitioners are more likely to recognise important aspects of cognition when they encounter them". Our method provided us a training of about four years in the field by observing action, interviewing practitioners, data collecting and analysis of events. We progressively developed a deeper understanding of stakes, multiple goals by discussion and interview, direct observation on board and in the shipyard. By doing this we validated the usefulness of the method and tools for the field and built continuously a trustworthy climate for collaboration. We now significantly better understand what's happening, what is at stake and the research is accepted by the skipper, the team and the sponsor. The researcher in reflexive position inside a team shares shemata and common knowledge of the team (Lipshitz & Ben Shaul, 1997) and is sensitized to collective mind (Weick & Roberts, 1993).

Intensive observing period

We intend to implement an intensive observing period for the next east to west double handed transatlantic involving "Classe Ultime" yachts (Trimarans longer than 60 feet) due to start from Le Havre (Normandie, France) in October 2017 to Salvador de Bahia (Brazil). As routing is allowed in race rules, co-skippers are going to work with a router to define and adjust race strategy. To be as close as possible to the practitioners without interfering in the race process, one researcher is shadowing the router. In this position the researcher will have real time access to all technical information. Furthermore, he or she will be able to hear and record discussions between skipper and router during the race and days before start. He or she also will be able to interview router as often as needed in the course of action and at times he may have an access to direct discussion with skipper to clarify point as well as with sponsor, technical team and communication agency. As the route of the race crosses Atlantic Ocean starting at mid latitudes and finishing in the tropics of southern hemisphere (Brazil) the sailors are facing typical navigation problems. Indeed, they must sail at first in the westerly flow of mid latitudes then find a way to catch the northeasterly trade winds of northern hemisphere. Then they must cross the doldrums and catch the trade the southeasterly trade winds of southeast hemisphere, which creates phase changes. As several weather systems are encountered, a proper management of the transition from the influence area of one to the influence area of the next one is a crucial issue. So not only decision are to be made, but important decisions are to be made at least at those major transitions.

Dynamic and complex field

Modern multihull yachts of "Classe Ultime" are the fastest sailing boats (speed are faster than 40 knots) having a race across ocean. In that case, those boats can move faster than most of usual weather systems on the one hand and can sail twice faster as the wind speed in particular conditions. Taking this into account, the sailing area is very large because an extra distance can be more than compensated by a small increase in wind speed or a better angle in wind direction. As biggest and fastest boat, "Classe Ultime" yachts are the one which are concentrating interest of media and investment of sponsors. The teams are composed of highly skilled workers, the crew are composed of experienced professional skippers and sailors. Skippers have to manage budget, boat, team, crew, innovation and are used to make decisions in stressed and uncertain situation at sea. Situation is dynamic because of three reasons: natural conditions (both atmosphere and ocean) are evolving in a continuous way, the boat is moving along its trajectory on its own pace, and rivals are sailing their way too. These yachts require high skill to be skipped properly. All the skippers of "Classe Ultime" trimaran have been sailing multihul for at least 10 years and up to 20 years for most of them. As an example, in particular conditions, the difference between high performance and danger for life (capsize) is less than 20 centimetres along a rope or less than 5 degrees of the true wind angle. The racing team is managed by one of the most experienced skipper (Lionel Lemonchois). He has been sailing for more than 35 years. During the last 15 years the sailor broke more than 10 world records in trans-oceanic route or round the world non-stop on big catamarans and trimarans. He is considered as one of the most experienced multihull skippers by others.

RESULTS AND PRELIMINARY ANALYSIS

Along the development of methodological framework, different embedded observations have been performed during last 4 years. Action based observation and preliminary analysis showed that phase shifts are important issues in the relationship between decision processes and action. Preliminary analysis is based on 3 cases involving the same skipper: crew training (2 days onboard and workplace observations), single handed transatlantic race (10 days observation inside the shoreteam in a routerlike position), double handed transatlantic race (10 days observation inside the shoreteam shadowing the router). Decision processes in those cases are performed in stressed situations especially high for the two race cases. From the analysis at the level of the action focused on the boat considered as an organization, we can assume that the way action is going on is governed by rules. Those rules depend either on the environment only (ex. weather, overarching action, side action), or on the organization only (ex. boat capacity, crew skills, team behaviour, expected goal), or on an interaction between environment and the organization (ex. the way low pressure is pulling a sail to create aerodynamic power, interaction between skipper and race organizer). A shift in phase occur when at least one rule is modified for a significant period of time according to length of action. If it seems quite easy to identify and describe one phase or another, the shift between them is much more complicated than it could be inferred at first guess. Embedded observations showed that a phase shift can spend quite a long time. Influence of a given rule can decrease gradually whereas another one is increasing. The influence of a given rule can fluctuate in a general trend of decrease or increase. It has been noticed that all along a transition period between two well settled phases, there is a gap in analysis of relation between action and decision if the influence of changing rules (emerging or disappearing) is not taken into account. Our guess is that such an approach focused on transition periods may benefit, not only to NDM practitioners, but also to organizations for performance improvement.

CONCLUSION AND RESEARCH DIRECTIONS

As skippers are convinced that human factor can make a difference in yacht race there is an opportunity to design a research project to study this field to better understand relationship between decision processes and action in “real world” to improve performance. Thanks to a specific skill held by one researcher in an important domain for the performance of the action, it has been possible to enter the field with mutual trust and respect. Considering longer term than a single race, skippers rely on researchers also to help them to improve their performance for next stages of their career. In terms of acceptability of the research as well as in terms of capacity building of the researcher, it appears to be all the more important for the researcher to be involved in the team as early as possible. Discussion showed that to be accepted by the team the researcher activity must not interfere the course of action and not modify the way of thinking of skippers during the race. The researchers are expected not to disturb the organisation even if the only fact that a research is performed can have an impact on the team behaviour at work. The design of intensive observing period lasted four years because of cautious step by step working out to find the deepest acceptable way to do research in embedded but not intrusive position.

The first outcome we were expecting was to show that the original position of the researcher can provide valuable information unreachable otherwise. Second, we expect to identify pieces of information that can lead to better understand relationships between decision processes and the changes occurring in the action process. We are also expecting to show how rich a qualitative empirical analysis of real world action can be when conducted in such dynamic context documented by very different sources of information.

Finally, we are expecting to identify tracks to follow to better understand how expert sailors are thinking at sea and how they can improve their ability to make a difference in real time. As our approach is relevant to macrocognition (Klein, Hoffman & Militello, 2016) expected outcomes may lead to refine NDM frameworks and applications in a multidisciplinary effort according to the views expressed by Mosier and Militello (2016). Complementary research involving embedded position in other fields could be developed. The proposal is to do more researches as an insider to identify small scale processes in real time as close as possible of the action in the real world. The way we conduct our research project could be implemented in other organizational contexts in order to look for similarities or differences in the relationship between decision processes and action. A comparison of knowledge elicitation in real world studies by the method proposed here and by other methods like ethnomethodology, action research or process studies could identify methodological advantages of each for a better understanding of the relationship between decision processes and action in different uncertain contexts where phase shifts occur. Finally, for each step forward it would be interesting to confront the fields results to existing decision making models.

ACKNOWLEDGMENTS

We gracefully thank sailors, teams and sponsors for their collaboration and the time they dedicated to our work.

REFERENCES

Allard-Poesi, F. & Giordano, Y. (2015). Performing Leadership “In-Between” Earth and Sky. *M@n@gement*, 18 (2), 103-131.

Brandt, S. L., Lachter, J., Battiste, V., & Johnson, W. (2015). Pilot Situation Awareness and its Implications for Single Pilot Operations: Analysis of a Human-in-the-Loop Study. *Procedia Manufacturing*, 3, 3017-3024.

Crandall, B., Klein, G., & Hoffman, R. (2006). *Working minds: A practitioner’s guide to cognitive task analysis*. Cambridge, MA: MIT Press.

Gore, J., Flin, R., Stanton, N., & Wong, B.L.W. (2015). Applications for naturalistic decision making. *Journal of Occupational and Organizational Psychology*, 88, 223-230.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, 93 (5), 1449-1475.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64, 515-526.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. (2015). Reflections on applications of naturalistic decision making. *Journal of Occupational and Organizational Psychology*, 88, 382-386.

Klein, G., Hoffman, R., Militello, L. (2016). Naturalistic Decision Making + Macrocognition. *Cognitia, HFES CEDM TG Newsletter*, 22 (2), 3-5.

Klein, G., Orasanu, J., Calderwood, R., & Zsambok, C.E. (1993). *Decision making in action. Models and methods*. Norwood, NJ: Ablex.

Lipshitz R, Ben Shaul O. (1997). Schemata and mental models in recognition-primed decision processes. In C.E. Zsambok C. & G. Klein (Eds.), *Naturalistic decision making* (pp. 293-304). Mahwah, NJ: Lawrence Erlbaum & Associates.

McAndrew, C., & Gore, J. (2013). Understanding Preferences in Experience-Based Choice: A Study of Cognition in the « Wild ». *Journal of Cognitive Engineering and Decision Making*, 7(2), 179-197.

Macquet, A.-C. (2010). Time management in the context of elite sport training. *The Sport Psychologist*, 24, 194-210.

Macquet, A.-C. (2013). Getting them on the same page: A method to study the consistency of coaches' and athletes' situation understanding during training sessions and competitions. *The Sport Psychologist*, 27, 292-295.

Macquet, A.-C., Ferrand, C., & Stanton, N. A. (2015). Divide and rule: A qualitative analysis of the debriefing process in elite team sports. *Applied Ergonomics*, 51, 30-38.

Macquet, A.-C., & Kragba, K. (2015). What makes basketball players continue with the planned play or change it? A case study of the relationships between sense-making and decision-making. *Cognition, Technology & Work*, 17(3), 345-353.

Macquet, A.-C., & Skalej, V. (2015). Time management in elite sports: How do elite athletes manage time under fatigue and stress conditions? *Journal of Occupational and Organizational Psychology*, 88(2), 341-363.

Macquet, A., & Stanton, N.A. (2014). Do the coach and athlete have the same "picture" of the situation? Distributed situation awareness in an elite sport context. *Applied Ergonomics*, 45, 724-733.

Marchais-Roubelat, A. (2012). Contracts to frame sustainable futures: the rational and symbolic sides of contract functions and dysfunctions , *Society and Business Review*, 7 (1), 50-64.

Mosier, K. & Militello, L.G. (2016). Preface to "Extending naturalistic decision making: Reaching Across Domains, Disciplines, and Applications". *Journal of Cognitive Engineering and decision processes*, 10 (3), 227-228.

Orasanu, J.M. & Fischer U. (1997). Finding decisions in natural environments: The view from the cockpit. In C.E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 343-357). Hillsdale, NJ: Lawrence Erlbaum Associates.

Orlikowski, W.J. & Yates, J. (2002). It's about time: Temporal structuring in organizations, *Organization Science*, 13 (6), 684-700.

Stanton, N.A. (2016). Distributed Situation Awareness. *Theoretical Issues in Ergonomics Science*, 17 (1), 1-7.

Weick, K.E., & Roberts, K.H. (1993). Collective Mind in Organizations: Heedful Interrelating on Flight Decks. *Administrative Science Quarterly*, 38, 357-381.

Designing for Self-Organisation with the Diagram of Work Organisation Possibilities

Neelam NAIKAR[3], Ben ELIX, Claire DÂGGE, and Thomas CALDWELL
Defence Science and Technology Group, Melbourne, Australia

ABSTRACT
The concept of self-organisation may provide a useful perspective for designing complex sociotechnical systems so they have greater capacity to deal with instability or uncertainty. Existing approaches concerned with promoting adaptation, such as resilience engineering and cognitive work analysis, are limited in their capacity to support self-organisation. However, a recently proposed extension to cognitive work analysis, the diagram of work organisation possibilities, may provide a means for designing for self-organisation. This tool models the structural possibilities of multiple actors and the behavioural opportunities of individual actors within a single, integrated representation. Consequently, designs may be developed that enable new spatial, temporal, and functional structures to emerge from the spontaneous actions of individual, interacting actors. The features of this new design tool in relation to promoting self-organisation are illustrated with examples from a military system.

KEYWORDS
Uncertainty management; self-organisation; cognitive work analysis; resilience engineering; design.

INTRODUCTION

Complex sociotechnical systems (Vicente, 1999) pose special challenges for design. Workers in hospitals, petrochemical refineries, emergency management centres, and many other contemporary workplaces must operate under continually shifting or evolving conditions. Consequently, design methods for such systems, specifically for such features as its interfaces, teams, training, and automation, must be able to accommodate the unstable and sometimes unpredictable conditions under which workers must perform.

This paper examines the concept of self-organisation as a basis for designing systems that can be effective in the face of instability or uncertainty. It considers the relationship between adaptation and self-organisation in complex sociotechnical systems, and then assesses the capacity of some existing approaches concerned with adaptation, namely cognitive work analysis and resilience engineering, to support the design of self-organising systems. Following that, the paper discusses the potential of a new design tool, the diagram of work organisation possibilities, to facilitate the development of systems with greater capacity for self-organisation than can be achieved with standard techniques. This tool is illustrated with examples from a military system.

ADAPTATION AND SELF-ORGANISATION

The idea that the design of complex sociotechnical systems should be focused on the goal of promoting adaptation (Rasmussen, Pejtersen, & Goodstein, 1994; Vicente, 1999) is now widely accepted. Underpinning this view of system design is the observation that workers in these systems are required to deal with events that *cannot be* anticipated, or fully specified a priori, by analysts or designers. Given that workers cannot be provided with pre-defined procedures for handling events we cannot foresee, the aim of design *must be* to support workers in adapting or improvising their actions to deal with such circumstances when they occur.

A number of frameworks or paradigms from a variety of fields, such as cognitive engineering, human factors, organisational psychology, management science, and safety science, recognise the importance of adaptation in the workplace. However, as the literature in these areas refers to the concept of self-organisation as well, it raises the question of the relationship between adaptation and self-organisation in these contexts.

For example, in providing an account of cognitive work analysis (CWA), Vicente (1999) emphasises the need for promoting adaptation in discussing the first three (work domain analysis, control task analysis, and strategies analysis) and fifth (worker competencies analysis) dimensions of this framework. However, in describing the fourth dimension (social organisation and cooperation analysis), he discusses at length the importance of supporting self-organisation. Similarly, of 60 journal papers on resilience engineering published between 2006 and 2017, almost all adopted a definition of resilience that encompassed the concept of adaptation. However, two

3 neelam.naikar@dsto.defence.gov.au

of these papers, which were published in the last four years, referred to self-organisation as well (Lundberg & Johansson, 2015; Lundberg & Rankin, 2013).

The relationship between adaptation and self-organisation is not discussed explicitly in these publications. However, it appears that the term self-organisation, and its variants, is used when the discussions are concerned primarily with changes in organisational structure, whereas the term adaptation, and its variants, is used more generally to encompass changes in actors' structures or behaviours. For instance, the fourth dimension of CWA focuses on the analysis of the work organisation in a system, specifically the allocation, distribution, and coordination of work among multiple actors, whereas the other dimensions of CWA need not focus on structure particularly. In the same way, Lundberg & Rankin (2013) refer to self-organisation in discussing role improvisations by workers, specifically when workers adopt roles on their own initiative, whereas they refer to adaptation in more general discussions, suggesting that this term encompasses improvisations in roles or tasks, whether by individuals, teams, or organisations.

While this interpretation of the usage of the two terms might not have been intended by the authors, it is important to recognise explicitly that, in self-organising systems, structure and behaviour are closely interconnected (e.g., Haken, 1988; Heylighen, 2001). Structural properties are seen as emerging from the spontaneous behaviours of individual, interacting elements, as well as constraining the behaviours of those elements. Figure 1 expresses this relationship with respect to social systems (Fuchs, 2004). The figure conveys that structures constrain, and enable, actors' behaviours or actions and emerge from their actions. Hence, in a social context, the concept of self-organisation recognises the interplay between the structures of multiple actors and the behaviours of individual actors, which is integral to the process of a system adapting to its environment.

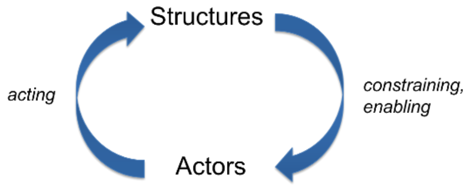


Figure 1. Social self-organisation. Adapted from Fuchs (2004).

It would be fair to ask, however, whether complex sociotechnical systems are in fact self-organising systems. The answer to this question depends on precisely what it means to be self-organising, and there are many conceptualisations in the literature associated with different types of system (e.g., chemical, biological, economical) and research disciplines, suggesting that particular definitions are useful in particular contexts (Gershenson & Heylighen, 2003). Nevertheless, certain properties of self-organising systems are mentioned recurrently, across a variety of contexts (e.g., Haken, 1988; Heylighen, 2001).

In relation to complex sociotechnical systems, Vicente (1999) draws attention to the fact that self-organising systems are generally regarded as being distributed systems with emergent features that arise as an adaptive response to local circumstances, such that these features cannot be completely planned or predicted a priori. In addition, although these systems are open to environmental influences, the external factors act on the system in a non-specific fashion, shaping rather than controlling or imposing the changes exhibited.

On the basis of these characteristics, it may be argued that complex sociotechnical systems are self-organising, at least some of the time (e.g., Bigley & Roberts, 2001; Bogdanovic, Perry, Guggenheim, & Manser, 2015; Rochlin, La Porte, & Roberts, 1987). For example, in Rochlin et al.'s field study of naval aircraft carriers, it was found that decision making during complex operations is distributed rather than centralised. Furthermore, as Vicente (1999) observes, the work organisation on these ships is emergent on two different time scales. The first relates to the gradual evolution of a work organisation that shifts flexibly between formal and informal structures, and the second relates to the informal organisation that is enacted spontaneously on any occasion. Moreover, this work organisation is not a result of prior scheme or intention, nor is it imposed by external factors.

Such empirical studies provide reason to believe not only that self-organisation occurs in complex sociotechnical systems but that it is necessary. Significant events occur too quickly and conditions are too unstable—and unpredictable often enough—for a priori planning to be a feasible strategy consistently. Instead, flexibility is required to achieve the "proper, immediate balance" between the system's safety and productivity goals, given the local conditions (Rochlin et al., 1987, p. 83).

Irrespective of the precise definition of a self-organising system, then, it appears that the concept of self-organisation, as it is described here in the context of complex sociotechnical systems, may provide a useful focus, or perspective (Gershenson & Heylighen, 2003), for designing systems so they have greater capacity to deal with instability and uncertainty. In particular, the concept emphasises that structure and behaviour are closely interconnected. Structural properties at the system level vary with the actions of interacting, individual elements and vice versa. Therefore, designs focused on the behaviours of individual actors will have implications for the structural relationships between these actors, and the converse is also true. Thus, to preserve or promote the capacity of sociotechnical systems for self-organisation, and consequently for dealing with instability or uncertainty, designs must support the structural possibilities of multiple actors and the behavioural opportunities of individual actors in an integrated fashion.

EXISTING FRAMEWORKS FOR DESIGN

As mentioned above, several frameworks or paradigms in various fields recognise the importance of adaptation in the workplace and refer sometimes to self-organisation. In this section, the capacity of two approaches, resilience engineering and CWA, to support self-organisation in complex sociotechnical systems is examined.

Resilience Engineering

Resilience engineering is viewed as a paradigm for safety management (e.g., Hollnagel, Woods, & Leveson, 2006) that is based on Jens Rasmussen's ideas about the importance of adaptation in managing uncertainty in work settings (e.g., Moorkamp, Kramer, van Gulijk, & Ale, 2014) and on the work of a number of organisational theorists, such as Karl Weick and Gene Rochlin, on high-reliability organisations (e.g., Hopkins, 2014).

Recently, Righi, Saurin, & Wachs (2015) undertook a systematic review of the resilience engineering literature, with one of their goals being to identify the categories of research conducted in this area. A key finding of this review, which encompassed 237 studies published between 2006 and 2014, was the relatively limited emphasis of these studies on design. As such, design was not identified as a research category in its own right.

The review also showed that the majority of research focused on theory or on understanding and describing resilience, with some of the studies in this category based on empirical data. By promoting a rich understanding of how effective performance occurs in a variety of complex settings, empirical studies can be useful for both theory and method development. For example, empirical studies of high-reliability organisations informed the development of CWA, a framework for design, as is evident from the citations in the standard texts in this area (Rasmussen et al., 1994; Vicente, 1999).

Vicente (1999), however, has cautioned against the use of descriptive methods as the fundamental basis for design in individual cases. A key part of his argument is that, in the case of any particular system, implications for design derived from such methods are based on descriptions of specific work practices observed under particular conditions. Generally, these are conditions that have been experienced—even if they were novel at the time—or conditions that can be anticipated. Consequently, the capacity of resulting designs to promote effective performance in unforeseen situations, or situations that cannot be predicted a priori, is limited by the extent to which the observed practices are relevant to the novel conditions. Arguably, then, insights from descriptive studies should be complemented with formative methods for design, such as that offered by CWA. A promising start in combining resilience engineering objectives with CWA has been made by Hassall, Sanderson, and Cameron (2014).

Finally, Moorkamp et al. (2014) provide a critique of the capacity of resilience engineering to support safety management or design in self-organising systems. A large part of their assessment concerns the functional resonance analysis method or FRAM (Hollnagel, 2012). Moorkamp et al. draw attention to the fact that FRAM focuses on describing a system's functions or activities without reference to its structure. The rationale for this approach is that a representation of a system's structure would necessarily portray a normative organisation of functions, which is unrealistic in many cases (Hollnagel & Goteman, 2004). However, Moorkamp et al. argue that an approach that examines functions without considering the structures required for conducting those functions ignores the relevance of interactions between actors, or social interactions, for improving safety, particularly through organisational design. Moreover, in actual applications of FRAM, it appears that the analysis of functions is not actually independent of structure. Rather, it seems that a structure must be assumed, which in Hollnagel and Goteman's (2004) study is the existing organisational structure, as it was considered to be "acceptable as it is" (p.160). For these and other reasons, Moorkamp et al. argue that FRAM is unable to account for systems with dynamic work organisation or self-organising systems.

Cognitive Work Analysis

CWA is a framework for the analysis, design, and evaluation of work in complex sociotechnical systems, which has the intent of fostering safety, productivity, and workers' health (Rasmussen et al., 1994; Vicente, 1999). Consistent with a formative orientation, this framework defines the constraints that must be respected by actors in an array of situations, including those that cannot be predicted a priori, if a system is to perform effectively. Within the constraints on their behaviours, actors still have many degrees of freedom for action. Therefore, by basing designs on these constraints, workers can be given the flexibility to adopt a wide variety of work practices, even those that have not been observed previously, without violating the boundaries of effective performance.

The five dimensions of CWA are concerned with different types of constraint, specifically those relating to the work domain, activity, strategies, work organisation, and workers.

Considerable evidence exists for the value of CWA for design, including a large body of experimental investigations of ecological interface design (Vicente, 2002) and detailed industrial case studies of team and training-system design (Naikar, 2013). However, the phenomenon of self-organisation draws attention to the fact that the standard framework (Rasmussen et al., 1994; Vicente, 1999) may result in solutions that limit not only the structures of multiple actors but also the behaviours of individual actors, as the two are tightly interconnected. Specifically, in the standard framework, the analysis of the work organisation is restricted to recurring classes of situation, so that unforeseen events may not be accommodated, and even further to organisational structures that have been observed or are judged to be reasonable in these situations, so that other possible structures may be excluded (Naikar & Elix, 2016b). Furthermore, the standard framework implies an approach to system design whereby the set of possibilities for work organisation may be limited to those that can be supported by the team design, which may be smaller than the set that might have been accommodated if the team design had not been established first but considered jointly with the design of other elements (Naikar & Elix, 2016a).

A clear consequence of the standard framework, then, is that it limits unnecessarily the structural possibilities of multiple actors in a system. However, the phenomenon of self-organisation emphasises that it may also limit the behavioural opportunities of individual actors without good reason. As discussed above, in self-organising systems, structures constrain and enable the behaviours of individual elements as well as emerging from their behaviours (e.g., Fuchs, 2004; Haken, 1988; Heylighen, 2001; Figure 1). Therefore, by limiting the structural possibilities of actors in a system, the standard framework may restrict actors' opportunities for behaviour as well, and ultimately a system's capacity for self-organisation.

DIAGRAM OF WORK ORGANISATION POSSIBILITIES

In this paper, we propose that a recently suggested extension to CWA, the diagram of work organisation possibilities, or WOP diagram (Naikar & Elix, 2016b), provides a basis for designing for self-organisation in complex sociotechnical systems, an argument which has not been formulated previously. To this end, this paper examines the relationship between the WOP diagram and the phenomenon of self-organisation and highlights that by modelling the structural possibilities of multiple actors and the behavioural opportunities of individual actors within a single, integrated representation, this tool provides a means for preserving a system's inherent capacity for self-organisation.

Description of WOP Diagram

The WOP diagram, which is constructed within the social organisation and cooperation dimension of CWA, encapsulates the set of possibilities for work organisation in a system, irrespective of specific events or circumstances (Naikar & Elix, 2016b; Figure 2). This diagram is created by defining the constraints on the possibilities for work organisation, or organisational constraints, rather than by specifying each of the possibilities. These constraints signify limits on the distribution of work demands across actors (i.e., human workers and automation), which cannot be broken under any circumstances, even those that cannot be foreseen or anticipated at present. Nevertheless, within these constraints, there is usually a very large number of work organisation possibilities. For example, in the case of a maritime surveillance aircraft, the number of possibilities was found to be in the order of 10^{27} (Naikar & Elix, 2016b).

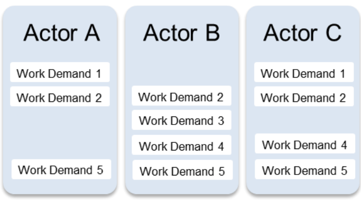


Figure 2. Generic diagram of work organisation possibilities. Reproduced from Naikar and Elix (2016b).

Relationship between WOP Diagram and Self-Organisation

In considering the connection of the WOP diagram to the phenomenon of self-organisation, it is important to recognise that the organisational constraints represented in the diagram establish relationships between the action possibilities afforded by the work context, or environment in which actors are situated, and the actors themselves. Thus this diagram simultaneously reveals the behavioural opportunities of individual actors and the structural possibilities of multiple actors (Figure 3) in a manner consistent with the phenomenon of self-organisation (cf. Figure 1). Consequently, designs based on this diagram may facilitate the emergence of novel temporal, spatial, or functional structures from the opportunistic behaviours of individual, interacting actors, so that the system has greater capacity for dealing with instability and uncertainty. The following sections describe these features of the WOP diagram in more detail, highlighting their relationship to the phenomenon of self-organisation.

Behavioural Opportunities of Individual Actors

Within the WOP diagram, the action possibilities of the work context are revealed at the level of individual actors in the form of their spaces of behavioural opportunities. Each actor's space of behavioural opportunities is designated by a set of work demands inherent to the work context, or environmental constraints, which are specified in the earlier dimensions of CWA. For each actor, these work demands or constraints collectively define a bounded space that differentiates behaviours that are possible or acceptable from those that are impossible or unacceptable for safe and productive performance. Thus the constraints on effective performance must be respected by each actor, irrespective of specific circumstances or events, including those that are novel or unforeseen. Yet, the constraints do not uniquely specify the behaviours of each actor, and within the constraints, each actor has many degrees of freedom for behaviour. Therefore, consistent with the phenomenon of self-organisation, the WOP diagram recognises that individual actors in sociotechnical systems may safely and productively engage in spontaneous action in response to the challenges posed by a dynamic and unpredictable work environment.

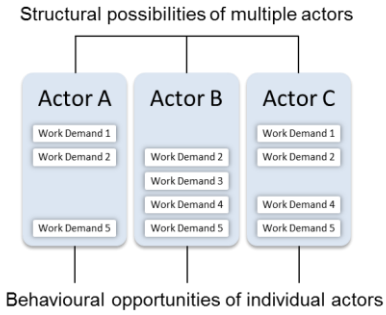


Figure 3. Relationship of WOP diagram to self-organisation (cf. Figure 1).

In the case of the maritime surveillance aircraft mentioned above (Naikar & Elix, 2016b), the behavioural spaces of actors at each of two flight deck stations encompass the work demands of flight; navigation; weapons authorisation; weapons release; tactical communications; air traffic control communications; observations of targets; detection, tracking, and identification of targets; and mission planning; among others. These work demands collectively necessitate certain kinds of behaviours, but within these constraints, there is considerable freedom for spontaneous action. For instance, the work demands of navigating an aircraft safely through a physical space may be achieved by utilising automated assistance, cross-checking electronic maps, visually sighting terrain and other entities through a window, listening to radio communications, or various other behaviours, which may not all be conceived a priori. Similarly, the ways in which these and other behaviours are interleaved or sequenced may vary opportunistically as a function of the circumstances, and there may be considerable variation in other aspects of their execution as well, for example, their durations. Furthermore, the behavioural spaces of actors at two observer stations and six workstations in the aircraft's cabin also encompass the work demand of navigation, so that any of these actors may also be engaged in related behaviours at any moment, given the circumstances.

Structural Possibilities of Multiple Actors

Within the WOP diagram, the action possibilities of the work context are revealed at the level of multiple actors in the form of their structural, or relational, possibilities. Moreover, the structural possibilities are interconnected with individual actors' behavioural opportunities in such a way that the resulting representation accommodates the observation that, in self-organising systems, the structural forms of multiple actors emerge from the opportunistic actions of individual, interacting actors, while at the same time constraining or enabling their specific behaviours (Figure 1).

As a simplified example, Figure 3 shows that the behavioural opportunities of both actors A and C, though not B, encompass work demand 1. Thus, some of the structural forms that may emerge from this set of actors are as follows: actor A is engaged in behaviours relating to work demand 1 while actors C and B are occupied in other behaviours; actor C is involved in behaviours relating to work demand 1 while actors A and B are engaged in other behaviours; or actors A and C are both occupied in behaviours relating to work demand 1 while actor B is involved in other behaviours.

These structural forms also constrain or enable the behaviours of individual actors, such that when the first structural form is in place, only actor A may be observed to be engaging in behaviours relating to work demand 1, whereas when the third structural form is active both actors A and C may be observed to be involved in such behaviours. However, both structural forms are possible. Therefore, depending on local contingencies, the structural forms may vary in accordance with the spontaneous behaviours of individual actors. For instance, the first structural form may be in place routinely, constraining or enabling the behaviours of individual actors in accordance with that structure, such that only actor A may be seen to be engaging in behaviours relating to work demand 1. However, if the circumstances change sufficiently, such that it is necessary or beneficial for actor C to assist actor A with work demand 1, a different structural form emerges from this opportunistic behaviour, which may remain in place until the situation changes further.

In the case of the maritime surveillance aircraft, the structural possibilities encompass actors at either one or both flight deck stations engaging in behaviours associated with the work demands of flying and navigating the aircraft safely. Therefore, how these behaviours are distributed across these actors may vary opportunistically as a function of the circumstances, so that the emergent structural forms may involve actors at either one or both flight deck stations at any point in time. Furthermore, as highlighted in the preceding section, the structural forms in relation to the work demand of navigation may also accommodate any one or more of the actors at the two observer stations and six workstations in the aircraft's cabin.

Similarly, the structural possibilities of actors in relation to the work demand of observing targets out of a window encompass actors at the two flight deck stations and two observer stations. However, if there is an electrical failure, such that none of the sensor systems available to actors at the six workstations can be used for detecting, tracking, or identifying targets, any one or more of these actors might relocate to the stations with a window to increase the chances of sighting the target. Thus, the emergent structural forms may exhibit new spatial properties, as well as temporal or functional characteristics.

Designing with the WOP Diagram

To support self-organisation in complex sociotechnical systems, the WOP diagram may be used as a basis for design. The preceding discussion demonstrates that this diagram models the structural possibilities of multiple actors and the behavioural opportunities of individual actors within a single, integrated representation in a manner compatible with the phenomenon of self-organisation. Therefore, it is plausible that designs based on the work organisation possibilities specified in this diagram may accommodate spontaneous actions from individual, interacting actors, such that new or different structural forms may emerge from their collective behaviours, as a function of the challenges posed by a changeable and ambiguous work environment.

Recently, the WOP diagram was used as a basis for developing designs for the maritime surveillance aircraft already mentioned. Although the relationship between the WOP diagram and the phenomenon of self-organisation had not been established at the time of design or initial reporting (Elix & Naikar, 2016), the resulting solutions appear to have the potential for supporting self-organisation. Specifically, the study concerned a future concept for the aircraft, which involves the use of an uninhabited aerial system (UAS) to fly to lower altitudes than the aircraft with the intent of detecting targets submerged underwater. A WOP diagram was constructed, which identified the individual actors on the aircraft who could be involved in operating the UAS, namely the two dry sensor operators, the two wet sensor operators, the junior tactical commander, and the senior tactical commander. Following that, designs for supporting the resulting possibilities for work organisation, including the behavioural opportunities of individual actors and structural possibilities of multiple actors, were considered. In particular, it was recognised that as the dry sensor operators progress to the role of wet sensor operators, and as the junior tactical commander progresses to the role of senior tactical commander, providing both the dry sensor operators and junior tactical commander with the required training for operating the UAS would be a means for giving all of these crew members some capacity for handling these work demands, thereby maximising the possibilities for work organisation in the system. Arguably, such an approach has greater potential for promoting self-organisation than more conventional techniques, which would have sought to identify which one of the crew members is best suited to controlling the UAS. However, the best crew member for the job is likely to be dependent on the situation.

CONCLUSION

In conclusion, the concept of self-organisation may provide a useful focus for designing complex sociotechnical systems, so they have greater capacity for dealing with instability and uncertainty. In particular, this concept emphasises that actors' structures and behaviours are closely interconnected. Therefore, designs must support the structural possibilities of multiple actors and the behavioural opportunities of individual actors in an integrated fashion to promote self-organisation.

In this paper, the relationship between the WOP diagram and the phenomenon of self-organisation was examined. This diagram encompasses the structural possibilities of multiple actors and the behavioural opportunities of individual actors within a single, integrated representation in a way that is consistent with many of the observations of self-organising systems. Therefore, it is proposed that designs based on this diagram have the potential to preserve a system's inherent capacity for self-organisation.

The ideas in this paper justify extensions to CWA so that it does not limit needlessly the structural possibilities and behavioural opportunities of actors, and ultimately a system's capacity for self-organisation. Furthermore, the assessments of CWA and resilience engineering highlight that the two approaches may be used complementarily, with resilience engineering contributing descriptive studies that enrich our understanding of how effective performance occurs in complex sociotechnical systems, and CWA providing a formative approach for design, which is informed by knowledge of how work can be achieved in complex settings.

Finally, empirical validation of the ideas presented in this paper is desirable. Studies should seek to address a range of questions including those relating to the specific nature or characteristics of self-organisation in complex sociotechnical systems, the validity of the WOP diagram for modelling the possibilities for work organisation in a range of systems, and the value of this tool for creating designs that promote self-organisation. Admittedly, this is a challenging research agenda, but even a thorough consideration of how some of these questions could be addressed would be a valuable contribution.

REFERENCES

Bigley, G.A., & Roberts, K.H. (2001). The incident command system: high-reliability organizing for complex and volatile task environments. Academy of Management Journal, 44(6), 1281-1299.

Bogdanovic, J., Perry, J., Guggenheim, M., & Manser, T. (2015). Adaptive coordination in surgical teams: an interview study. BMC Health Services Research, 15(1), 128.

Elix, B., & Naikar, N. (2016). Promoting worker adaptation in complex sociotechnical systems with integrated system design: Case study of team design [Abstract]. Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics. Louisville, KY: AFHE.

Fuchs, C. (2004). Knowledge management in self-organizing social systems. Journal of Knowledge Management Practice, 5. Retrieved from http://www.tlainc.com/articl61.htm (accessed 9 December, 2017).

Gershenson, C., & Heylighen, F. (2003). When can we call a system self-organizing? In W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, & J.T. Kim (Eds.), Advances in Artificial Life (pp. 606-614). Dortman, Germany: Springer. doi:10.1007/978-3-540-39432-7_65.

Haken, H. (1988). Information and Self-Organization: A Macroscopic Approach to Complex Systems. Springer: Berlin, Heidelberg, New York.

Hassall, M.E., Sanderson, P.M., & Cameron, I.T. (2014). The development and testing of SAfER: A resilience-based human factors method. Journal of Cognitive Engineering and Decision Making, 8(2), 162–186. doi: 10.1177/1555343414527287.

Heylighen, F. (2001). The science of self-organization and adaptivity. The Encyclopedia of Life Support Systems, 5(3), 253-280.

Hollnagel, E. (2012). FRAM: The Functional Resonance Analysis Method. Farnham, UK: Ashgate.

Hollnagel, E., & Goteman, Ö. (2004). The functional resonance accident model. Proceedings of the International Conference on Cognitive Systems Engineering in Process Control (pp. 155-161). Sendai, Japan: Tohoku University.

Hollnagel, E., Woods, D.D., and Leveson, N.G. (2006). Resilience Engineering: Concepts and Precepts. Aldershot, UK: Ashgate.

Hopkins, A. (2014). Issues in Safety Science. Safety Science, 67, 6-14.

Lundberg, J., & Johansson, B. J. E. (2015). Systemic resilience model. Reliability Engineering & System Safety, 141, 22-32. doi: http://dx.doi.org/10.1016/j.ress.2015.03.013.

Lundberg, J., & Rankin, A. (2013). Resilience and vulnerability of small flexible crisis response teams: Implications for training and preparation. Cognition, Technology and Work, 16, 143-155.

Moorkamp, M., Kramer, E-H., van Gulijk, C., & Ale, B. (2014). Safety management theory and the expeditionary organization: A critical theoretical reflection. Safety Science, 69, 71-81.

Naikar, N. (2013). Work Domain Analysis: Concepts, Guidelines, and Cases. Boca Raton, FL: CRC Press.

Naikar, N., & Elix, B. (2016a). A consideration of design approaches based on cognitive work analysis: System design and integrated system design. Proceedings of the 34th European Conference on Cognitive Ergonomics (pp. 1-7). New York, NY: Association of Computing Machinery.

Naikar, N., & Elix, B. (2016b). Integrated system design: Promoting the capacity of sociotechnical systems for adaptation through extensions of cognitive work analysis. Frontiers in Psychology, 7:962. doi: 10.3389/fpsyg.2016.00962.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). Cognitive Systems Engineering. New York, NY: John Wiley & Sons.

Righi, A.W., Saurin, T.A., & Wachs, P. (2015). A systematic literature review of resilience engineering: Research areas and a research agenda proposal. Reliability Engineering and System Safety, 141, 142-152.

Rochlin, G.I., La Porte, T.R., & Roberts, K.H. (1987). The self-designing high-reliability organization: Aircraft carrier flight operations at sea. Naval War College Review, 40(4), 76-90.

Vicente, K.J. (1999). Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Mahwah, NJ: Lawrence Erlbaum Associates.

Vicente, K.J. (2002). Ecological interface design: Progress and challenges. Human Factors, 44(1), 62-78. doi: 10.1518/0018720024494829.

Cognitive prompts for the elicitation of perceptual-cycle data

Katherine L PLANTa and Neville A STANTONa

[a] Transportation Research Group, Faculty of Engineering and Environment, Engineering Centre for Excellence, Boldrewood Campus, University of Southampton, Southampton, SO16 7QF, UK

ABSTRACT

The work presented here forms part of a larger body of research that explores the decision making process from the perspective of the Perceptual Cycle Model (PCM). The PCM provides a distributed approach, acknowledging the role of both internally held mental schemata and external environmental information in shaping decision making. However, there are no associated methods with this model and therefore researchers have had to be creative with other methods in order to infer perceptual cycle processing. This research presents cognitive prompts derived from data with twenty helicopter pilots. The prompts are designed to elicit information relating to the three elements of the PCM: schemata, actions and world information. It is intended that the data can be used to understand decision making processes from the perspective of the PCM. Future research endeavours are acknowledged in which studies are intended in order to establish the validity and reliability of the method.

KEYWORDS

Decision making; Transportation; Perceptual Cycle Model; Cognitive Task Analysis

INTRODUCTION

Neisser's (1976) Perceptual Cycle Model (PCM) provides a process oriented approach to understand decision making and we have previously applied it as a framework for explaining the mechanisms involved in aeronautical decision making (Plant and Stanton, 2012; 2013a). Aviation epitomises the Naturalistic Decision Making (NDM; Klein et al., 1986) environment in which experts make decisions with little information, in time critical situations and that can have great consequence. Decision and judgement errors are often a significant contributory factor in incidents and accidents; with some arguing that they are the primary factor in over 50% of general aviation accidents (Simpson, 2001). The decision maker acts according to their understanding of the situation at the time and therefore errors arise because of deficiencies in the decision makers' knowledge base or in the process of reaching a decision.

As depicted in Figure 1, the PCM emphasises the role that both schemata (internally held mental templates) and world information play on shaping actions, including decisions. Schemata are triggered by contextual conditions and direct perception and behaviour, thus interaction in the world. The world information that is experienced can have a modifying and updating effect on cognitive schemata, thus influencing further interaction. Behaviours are grounded within the context of the environment in which they occurred and therefore this model extends beyond the individual, offering a distributed cognition perspective. This framework can therefore be used to help understand why the actions and assessments undertaken by a decision maker made sense to them at the time.

A fundamental component of the PCM are schemata, these are akin to internal mental templates and therefore cannot be directly measured but only inferred through the manifestation of observable behaviour or recalled information (Plant and Stanton, 2013b). Contemporary approaches for eliciting schema-based data and inferring perceptual cycle processes generally pair data collection methods such as interviews with qualitative data analysis including network modelling or thematic analysis. This is a form of cognitive task analysis (CTA) as they are approaches that determine the cognitive elements (i.e. mental processes and skills) required for task performance and the changes that occur as skills develop (Militello and Hutton, 1998). In CTA the majority of data collection usually occurs through interviews (Militello and Hutton, 1998). One of the most popular in Ergonomics research is the Critical Decision Method (CDM; Klein et al., 1989). The CDM uses semi-structured cognitive probes to understand decision-making during non-routine situations resulting in a transcript that can then be analysed. We have previously used the CDM to infer perceptual cycle processes by thematically analysing CDM data in relation to the three elements of the PCM; schemata, actions and world information (Plant and Stanton, 2013a; 2014a, 2014b).

Specifically, the CDM focuses on eliciting knowledge for behaviours Klein et al. (1986) classed as recognition-primed decisions (RPD), i.e. decisions for which alternative actions are derived from recognition of critical information and prior knowledge. The RPD model is the most popular and enduring model in the NDM domain. However, we have previously argued that it does not go far enough at acknowledging the interaction between schemata and environmental information and the modifying effect each can have on the other (Plant and Stanton, 2014a). Rather, the RPD model tends to focus on the decision making processes that occur in the mind of the decision maker, consequently the CDM focuses on eliciting this content knowledge of operators. The role of the environment is acknowledged, however, this is only at a very high level with questions such as ‘what information cues were attended to’ and ‘what information was considered’. The aim of this research is to develop a data-driven interview probe method that is more applicable to extracting the three elements of the PCM in order that perceptual cycle processes can be better understood.

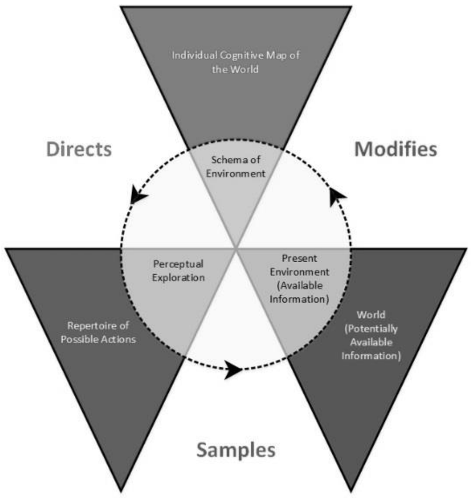


Figure 1. The Perceptual Cycle Model

METHOD

Participants

Twenty helicopter pilots were interviewed using the CDM. The pilots were recruited through an advert placed on the British Helicopter Associate website and via word-of-mouth. The sample consisted of 19 males and 1 female. Twenty five percent of the sample were aged between 31-40 years, 40% were aged between 41-50 years and 35% were aged between 51-60 years. The pilots were all relatively experienced; flying hours ranged from 1150-13000 (mean = 5942, SD = 3304, median = 5000) The pilots were employed in a variety of occupations including Search and Rescue, military, personal passenger transport, North Sea transport and test pilots. This study was granted ethical permission by the University of Southampton Research Ethics Committee.

Data collection

The CDM procedure involves participants describing a critical incident they were involved with, defining a timeline of events and answering the deepening probes. The deepening probes cover factors such as key goals, the role of experience, decisions that were made, options available and key sources of information. For a more detailed description of the procedure and an evaluation of the method readers are directed to: Crandall et al. (2006), Klein and Armstrong (2005), Stanton et al. (2013).

Each pilot was interviewed at their place of work and was asked to think of a critical incident they had been involved with, which was defined as being ‘a non-routine or un-expected event that was highly challenging and involved a high workload in which you were the primary decision maker’ (Klein & Armstrong, 2005). Each participant provided a high-level overview of the incident and structured a timeline of events. After the incident description/timeline construction phase, the cognitive probes were asked in relation to the decision making during the incident. The interviews were audio recorded and later transcribed.

Data analysis

In accordance with the guidelines on qualitative data analysis the text was chunked into meaningful segments of approximately one sentence or less in length. This resulted in 904 text segments across the twenty interviews. Both deductive and inductive thematic analysis was used to analyse the data. Firstly, deductive thematic analysis structured the data into the three components of the PCM; schemata, actions and world information because this is a process where themes are generated from existing theory, i.e. the PCM (Boyatzis, 1998). The classification scheme was as follows: schema (statements relating to the use of prior knowledge and knowing things because of experience or expectations), action (statements of doing an action or discussing potential actions that could be taken) and world (statements relating to potential or actually available information in the world including physical things, conditions or states of being). More detail is provided in Plant and Stanton (2013a).

Inductive thematic analysis was undertaken on the data in each of three high level categories of schema, action and world in order to uncover more detailed themes within this data. Inductive thematic analysis is the process by which the data are used to generate themes (Boyatzis, 1998). Here, the constant comparison technique was employed whereby each text segment was compared with previous items to see whether the same or a different phenomenon was described. This PCM taxonomy was developed through an iterative process of review and refinement using the opinions and expertise of colleagues in the research group. The process of inductive analysis resulted in the identification of six schema-subtypes, 11 action-subtypes and 11 world-subtypes. These are presented Table 1.

RESULTS

The categories from the PCM taxonomy were used to devise the cognitive prompts that are intended as a data elicitation method to help understand the perceptual cycle decision making process. The probes for each category are presented in Figures 2-4. The prompts are designed to enable operators to reflect on the three elements of the PCM during decision making. As with the CDM, the prompts can be utilised in the context of critical decision making, but also have just as much utility in non-critical decision making contexts such as normal operational scenarios, hence the focus on events rather than incidents in the prompts. These prompts are also intended to be domain specific which is discussed further in the next section.

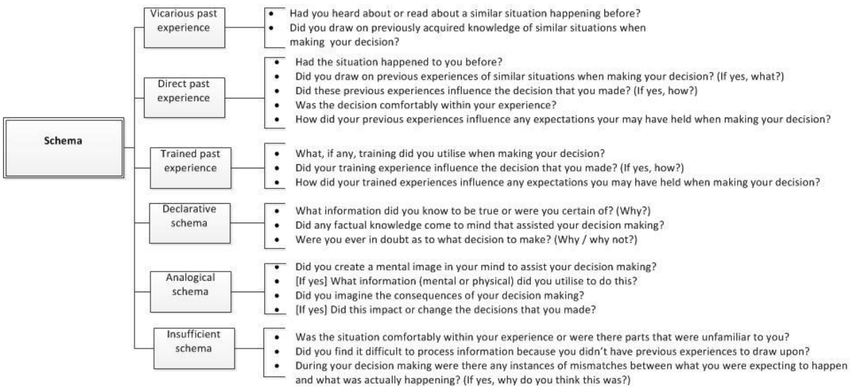


Figure 2. Prompts for the schema sub-types

Figure 3. Prompts for the action sub-types

Table 1. The PCM taxonomy

PCM element	Detailed code	Description
Schema	Vicarious past experience	Statements relating to experiencing something in the imagination through the description by another person or documentation.
	Direct past experience	Statements relating to direct personal experience of similar events or situations in the past (liver operational scenario)
	Trained past experience	Statements relating to knowledge developed by experience of a specific task, event or situation within the confines of a training scenario
	Declarative schema	Statements relating to a schema that manifests as a descriptive knowledge of facts, usually as a product of the world information available
	Analogical schema	Statements relating to comparisons between things for the purpose of explanation and clarification
	Insufficient schema	Statements relating to inadequate or lacking knowledge, i.e. a schema is not developed for a certain situation
Action	Aviate	Statements relating to direct manipulation (handling) of flight controls in order that the aircraft can be flown and safety is maintained
	Navigate	Statements relating to the process of accurately ascertaining position and planning and following a route or desired course
	Communicate	Statements relating to the sharing or exchange of information
	System management	Statements relating to the processes of making an input into technological systems in order that the interaction or manipulation has an explicit output
	System monitoring	Statements relating to looking at (observing, checking) displays to gain an understanding of the situation
	Environment monitoring	Statements relating to observing or checking the internal or external physical environment in order to establish the current state-of-affairs
	Concurrent diagnostic action	Statements relating to the process of determining, or attempting to determine, the cause or nature of a problem by examining the available information at the time the incident is occurring
	Decision action	Statements relating to a conclusion or resolution that is reached after considering the available information
	Situation assessment	Statements relating to actions that relate to the evaluation and interpretation of available information
	Non-action	Statements relating to actions that were not performed, either because the situation didn't warrant a particular action or because equipment faults did not allow a particular action to be performed or because the pilot made an error or omission.
World	Natural environmental conditions	Statements about natural environmental conditions (e.g. weather, light, temperature, noise)
	Technological conditions	Statements relating to the state of technological artefacts (e.g. with regards to appearance and working order)
	Communicated information	Statements relating to information available to the pilot from other people (e.g. other crew members, ATC, coastguard etc.)
	Location	Statements relating to particular places or positions
	Artefacts	Statements discussing physical objects, including written information, symbols, diagrams or equipment
	Display indications	Statements relating to the information elicited from the physical artefacts
	Operational context	Statements relating to the routine functions or activities of the organisation (e.g. Search and Rescue, Police search, military training etc
	Aircraft status	Statements relating to the current status of the aircraft's integrity or performance (e.g. how good or bad it is flying, the current configuration of the aircraft, autopilot activation etc.)
	Severity of problem	Statements relating to how bad (or otherwise) the critical incident is
	Physical cues	Statements relating to external cues that provide information of conditions or states of being (e.g. noises, sounds, vibration, smells)
	Absent information	Statements relating to information that was missing, not present or lacking. Reasons for this may include technical faults with equipment or non-existent information
	Natural environmental conditions	Statements about natural environmental conditions (e.g. weather, light, temperature, noise)

Figure 4. Prompts for the world sub-types

DISCUSSION

Summary

Decision making can be the deciding factor as to whether normal situations turn into incidents and incidents turn into accidents. As argued by Salmon et al. (2012), accidents represent a complex systems-phenomenon because they are the result of the interaction between causal factors that reside at all levels of a system. As such, it is appropriate to view decision making through the lens of distributed cognition, rather than as a process that occurs in the head of an isolated individual. Neisser's (1976) PCM provides this distributed cognition perspective by acknowledging the interaction between internal cognitive schemata held by the decision makers and the external environment in which decisions are made. The aim of this research is to develop an approach that is suitable for eliciting data relating to the perceptual cycle process.

The development of the PCM prompts was data-driven as they evolved from the more detailed taxonomy of the PCM components. Stanton (2005) argued that method development comes from processes based upon enhancements, modifications, and combinations of existing methods. We have previously advocated the use of the CDM to uncover perceptual cycle processes because the CDM elicits knowledge for decisions that are devised through the recognition of critical information from the environment (i.e. world) based on past experience and expertise (i.e. schemata) (Plant and Stanton, 2013a). However, as stated, the CDM is designed to elicit RPD making information and focuses more on the internal decision making process. The PCM prompts are clearly a descendent of the CDM but the intention is that they provide more insights into the nature of perceptual cycle processing by being specifically designed around components of this model.

Intended applications

It is envisaged that the PCM prompts will be utilised in a variety of research settings. Its most common application is likely to be as a semi-structured interview schedule for decision making situations, with the intention that it will aid operators in articulating knowledge that is difficult to verbalise. Unlike the CDM, these prompts are intended for use in both critical and non-critical decision making situations and the questions are written to reflect this. Non-critical decision making presents an important area of study because incidents and accidents are the by-product of normal functioning with people acting in a way that made sense to them at the time (Dekker, 2006). The only category of the prompts that explicitly applies to dealing with a critical incident is 'problem severity' and this can be omitted when the relevant prompts are selected as it is intended that the prompts can be selected to meet the requirements of the research question(s) under investigation. This is in line with Militello and Hutton's (1998) assertion that when conducting CTA using probes it is not expected that all probes will be equally relevant for all situations or domains. For example, if the research was predominately interested in the role of schemata in decision making then only the schema-based prompts could be asked in a post-scenario interview.

The prompts could also be utilised in a freeze-probe setting or at the end of task. For example, Schutte and Trujillo (1996) analysed task management priorities in aviation and used ‘natural conversational probes’ (e.g. “what happened back there”) during task execution to capture information about thought processes and decisions. In this study the probes were inserted into the scenarios via air traffic controllers or conversations with dispatchers. The PCM probes could have application in a study of this nature to provide insights into various aspects of decision making.

Evaluation and next steps

The PCM prompts were developed using a structured and exhaustive approach which produced a comprehensive list of prompts. They are based on the theoretical perspective of the PCM, O’Hare et al. (1998) argued that the importance of theoretically-driven CTA should not be underestimated as this has a huge potential to impact on subsequent system design and training improvements. We have previously studied decision making processes from the perspective of the PCM and therefore it seems appropriate to begin to develop a method based on this theoretical approach. Previously, to understand the perceptual cycle process in data collected via the CDM (or other means), the analysis required deductive thematic coding based on the high-level PCM categories. Data elicited from these prompts will already be structured around these high-level categories and therefore value lies in the time saved by removing a step in the analysis process. The prompts were written from the context of aeronautical decision making, however only two of the sub-types, aviate and aircraft status, make explicit reference to aviation terminology. Therefore with only minimal modifications it is envisaged that the prompts can become a domain independent research tool.

It is however, only through correlating evidence with other sources that confidence in the validity and reliability of the method can be established and this is the intention of future research. Stanton and Young (1999: 197) stated that “*the objective way to see whether ergonomic methods work is to assess their reliability and validity*”. Stanton et al. (2013) questioned the amount of verifiable evidence that many Human Factors methods and models are based on and argued that the results of validity and realibility assessments should be one of the key considerations when selecting methods. Validity concerns the accuracy or truth of the method, i.e. how accurately a method measures what it sets out to measure and corresponds to the real world. There are various facets of validity including constructct validty, face validity and criterion-rated validity. Broadly speaking, the PCM prompts can be described as valid if they demonstrate that they are capable of eliciting data that describes perceptual cycle information processing. Reliabilty concerns the consistency of measurement, i.e. that a method measures the same thing on two different occasions. This can be assessed in terms of inter-rater (different people using the method) reliability and intra-rater reliabilty (the same person using the method at different times). Future research endeavours will put the PCM prompts into practice in order to establish the validity and reliability of the approach. Militello and Hutton (1998) provided a rigorous assessment of a new CTA method and lessons of best-practice will be taken from their approach to guide our future efforts. It is hoped that disseminating these prompts to like-minded researchers and practitioners will allow them to be utilised and applied in order to begin this process.

CONCLUSION

The PCM views decision making through the lens of distributed cognition, providing a process-orientated approach for understanding how internally held mental schemata interact with information perceived in the external environment to produce actions and behaviours. To date, however, there are no specific methods associated with this theoretical perspective hence the motivation behind this work. We have produced PCM-based cognitive prompts designed to elicit data about decision making from the perspective of the three elements of the PCM. The methodological development is in its infancy, the next step is to apply the prompts in order to establish the validity and reliability of the approach.

ACKNOWLEDGMENTS

The authors thank the pilots who gave their time to participant in our interviews.

REFERENCES

Boyatzis, R. E. 1998. Transforming Qualitative Information: Thematic Analysis and code development, Thousand Oaks, CA: Sage.

Crandall, B. Klein, G.A. & Hoffman, R.R. 2006. *Working minds: A practitioner’s guide to cognitive task analysis*. Cambridge, MA: MIT Press.

Dekker, S. W. A. 2006. *The field guide to understanding human error,* Aldershot: Ashgate.

Klein, G. A., and A. Armstrong. 2005. “Critical Decision Method.” In Handbook of Human Factors and Ergonomics Methods, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, 35.1–35.8. Boca Raton, FL: CRC Press.

Klein, G. A., Calderwood, R. & Clinton-Cirocco, A. 1986. Rapid Decision Making on the fire ground. *30th Annual Human Factors Society Meeting.* Dayton, OH.

Klein, G. A., Calderwood, R. & Macgregor, D. 1989. Critical Decision Method for Eliciting Knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, 19 (3), 462-472.

Militello, L.G. & Hutton, R.J.B. 1998. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, 41(11), 1618-1641.

Neisser, U. 1976. *Cognition and Reality*. San Francisco: W.H.Freemand and Co

O’Hare, D., Wiggins, M., Williams, A. & Wong, W. 1998. Cognitive task analysis for design centred design and training. *Ergonomics*, 41(11), 1698-1718.

Plant, K.L. & Stanton, N.A. 2012. Why did the pilots shut down the wrong engine? Explaining errors in context using Schema Theory and the Perceptual Cycle Model. *Safety Science*, 50, 300-315.

Plant, K.L. & Stanton, N.A. 2013a. What’s on your mind? Using the Perceptual Cycle Model and Critical Decision Method to understand the decision-making process in the cockpit. *Ergonomics,* 56(8), 1232-1250.

Plant, K.L. & Stanton, N.A. 2013b. The explanatory power of Schema Theory: Theoretical foundations and future applications in Ergonomics. *Ergonomics*, 56(1), 1-15.

Plant, K.L. & Stanton, N.A. 2014a. The process of processing: exploring the validity of Neisser’s perceptual cycle with accounts from critical decision-making in the cockpit. *Ergonomics*, DOI: 10.1080/00140139.2014.991765

Plant, K.L. & Stanton, N.A. 2014b. All for one and one for all: Representing teams as a collection of individuals and an individual collective using a network perceptual cycle approach. *International Journal of Industrial Ergonomics*, 44, 777-792.

Salmon, P.M., Cornelissen, M. & Trotter, M.J. 2012. Systems-based accident analysis methods: A comparison of Accimap, HFACS, and STAMP. Safety Science, 50, 1158-1170

Schutte, P.C. & Trujillo, A.C. 1996. Flight crew task management in non-normal situations. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 244-248. Santa Monica, CA: HFES.

Simpson, P.A. 2001. *Naturalistic Decision Making in Aviation Environments*. Report No. DSTO-GD-0279. DSTO Aeronautical and Maritime Research Laboratory, Australia.

Stanton, N.A. 2005. Behaviour and Cognitive Methods. *In:* Stanton, N. A., Hedge, A., Brookhuis, K., Salas, E. & Hendrick, H., eds. *Handbook of Human Factors and Ergonomics Methods.* Boca Raton, FL: CRC Press, 27.1-27.8

Stanton, N. A., Salmon, P. M., Rafferty, L.A., Walker, G. H., Baber, C. & Jenkins, D. 2013. Human Factor Methods. A Practical Guide for Engineering and Design, 2nd ed. Aldershot: Ashgate.

Naturalistic decision-making perspective on uncertainty reduction by civil engineers about the location of underground utilities

Paulina RACZ, Marc VAN BUITEN, and André DOREE
University of Twente
Drienerlolaan 5
7522NB Enschede
The Netherlands
p.racz@utwente.nl; m.vanbuiten@utwente.nl; a.g.doree@utwente.nl

ABSTRACT

Modern engineers must perform their work carefully to avoid damaging buried underground utilities. Before starting ground works the exact location of pipes and cables must be confirmed. Current detection equipment still cannot provide complete certainty and requires extensive training in order to obtain the correct data. Digging test trenches remains an important practical tool to interpret subsurface conditions, but deciding on the number and location of test trenches is problematic. Decisions seem to be taken randomly and are based mostly on intuitive judgments. We conducted interviews and workshops in order to uncover the strategies used to select test trench location. Our results show that choices are influenced significantly by decision-makers' experience. We describe our findings using Rasmussen's Skills-Rules-Knowledge model. We propose that any future decision support system should combine elements from both analytical and naturalistic models.

KEYWORDS

Decision-Making; Civil Engineering; Test Trenches; Naturalistic Decision Models

INTRODUCTION

Excavation work demands that modern engineers take many precautions. They must schedule their works in such a way that existing underground utilities will not be disturbed. This is a challenge, given that underground spaces have become increasingly busy. Many types of cables and pipes, underground infrastructures, precious fauna and flora, as well as archaeological findings, can make excavation work difficult. Underground utilities, particularly, and depending on their content, can cause many problems. Damages to high-risk pipes and cables (e.g. gas pipelines, the sewage system, industry pipelines and/or high-voltage electricity cables) can have profound impacts on the environment and the health of workers and people living and working in the area neighbouring the excavation area.

Ground Penetrating Radars (GPRs) and Augmented Reality (AR) technologies are used by engineers to improve the safety of excavation procedures. However, these technologies are costly and still do not provide complete certainty. Furthermore, even if maps and plans of underground utilities are available, the information contained by these sources might be incorrect. Therefore, test trenches are often dug to confirm the exact locations. These trenches are variously described as test holes (Canada), trial holes (UK and US), potholing (Australia) and proefsleuven (the Netherlands). They all have the same goal: to establish precisely what is underground before excavation starts.

There is always doubt as to where to locate test trenches. Utilities' maps, GPRs and AR support those decisions. Nevertheless, it is the responsible decision-makers (designers, contractors) who make the final choice. The process of selecting the test trench location appears to be random and mostly based on the personal judgment of an individual. This exerts considerable pressure on these decision-makers faced with meeting deadlines and faced with the nagging uncertainty about the actual location of subsurface utilities and the risk of excavation damage to the utilities. The problem is well known to practitioners. However, there is a dearth of information about locating test trenches in the scientific literature which is surprising given the importance of digging test trenches that is readily apparent from various national excavation damages programmes, such as "Reduction of Damage to Utilities and Careful Excavation" (ReDUCE) in the Netherlands, PAS 128:2014 standards in United Kingdom (UK) or "Call before you dig" in United States (US).

The analysis in this article examines the Dutch excavation context and where clarity of test trenches location strategy is one major goal of the Dutch Government. The Netherlands has a national digital mapping system (referred to as "klic"). This system was developed by the Cables and Pipes Information Agency (Cadastre) and provides the excavator with maps of utilities in the area of interest, together with necessary documentation. In spite of this, damage to underground utilities persists. According to the data provided by Cadastre, almost 33,000 cases of damage were recorded involving Dutch utilities in 2015 and the cost of their repairs mounted to €30 million per year (Kabel en leiding overleg KLO, 2016) . The Underground Networks Information Exchange Act (WION), together with klic, provides a solid foundation for making decisions. The WION describes the steps that excavators must follow before breaking the surface, and the klic system supports them with the relevant maps. Before excavation starts it is mandatory for a request to be sent to Cadastre. The agency will provide the excavator with necessary maps, together with important documentation (e.g. about precautions).
In Figure 18, the klic system is presented.

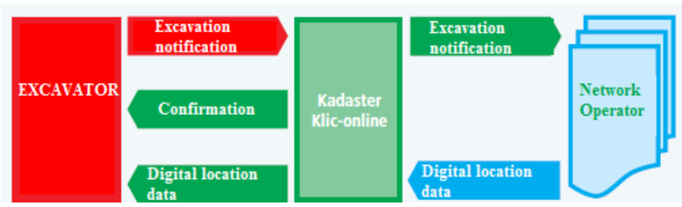


Figure 18. The klic system in the Netherlands (Groot, 2008)

New European regulations, such as Infrastructure for Spatial Information in the European Community (INSPIRE), as well as feedback from klic users, allow Cadastre to receive feedback to improve the system. As a result, the klic-win system was developed that is slowly being implemented to provide users with location data in vector format to create data sets with added value, such as 3D visualizations (Kadaster, 2016).
The authors intend to develop a Decision Support System (DSS)[4] for selecting test trenches to help decision-makers minimize risk and uncertainty during excavation and, as a result, reduce the number of cases where excavation damages underground utilities.

To achieve this goal, it is necessary to optimize the current decision-making strategy. This, in turn, requires careful analysis of current strategies, logic and decision-makers' behaviour. To that end, interviews and workshops were conducted. The purpose of those meetings was to check how people make decisions when factors such as risk, uncertainty and time play an important role. The research showed that on many occasions mental simulations conducted by decision-makers leading to their subsequent choices, were aligned with the degree of their expert level (i.e. expert vs novice). In this article, we investigate how decision-makers use their experience to decide on where to locate test trenches and whether it is possible to use Naturalistic Decision Models (NDM) as a decision support tool for Civil Engineering.

The structure of this article starts by outlining the background information of Naturalistic Decision Making models. Then, the methodology used to collect the data is described. Subsequently, the results of analysis are presented. The paper finishes with conclusions.

BACKGROUND ON NATURALISTIC DECISION MAKING (NDM) MODELS

To make a decision entails using information to choose between options from amongst the available possibilities (Ishibushi & Nii, 2000). On many occasions, researchers develop Decision Support Systems (DSS) from an analysis of several decision models as we did also (Racz, Van Buiten, & Dorée, 2016). However, we observed that this approach directed us more towards the development of the system itself and de-emphasised a focus on its users. The system would then represent an idealized example of reality by excluding actual users from the decision process (Blanchard & Fabrycky, 2011). NDM researchers consider the DSS development process differently (Klein, 2008). First, current strategies are analysed to examine how people behave during extreme situations where there is time pressure, uncertainty, unexpected conditions and/or risk (Klein & Klinger, 1991), rather than just simply checking whether people were effective or not. Civil engineering projects seem natural candidates for such an approach. Engineers and workers always perform their job under time pressure. The rush commonly results in mistakes with potentially dangerous consequences. In addition, construction work puts considerable responsibility

4 University of Twente, Reggefiber and other partners created the ReDUCE program (Reduction of Damage to Utilities and Careful Excavation). This programme is realised to promote careful approaches in construction and maintenance of the underground infrastructure, by investing in the development of new methods and technologies that prevent excavation damages (Reggefiber & University of Twente, 2015). The analysis are part of a PDEng (Professional Doctorate in Engineering) project entitled "Improved strategies, logic and decision support for selecting test trench location".

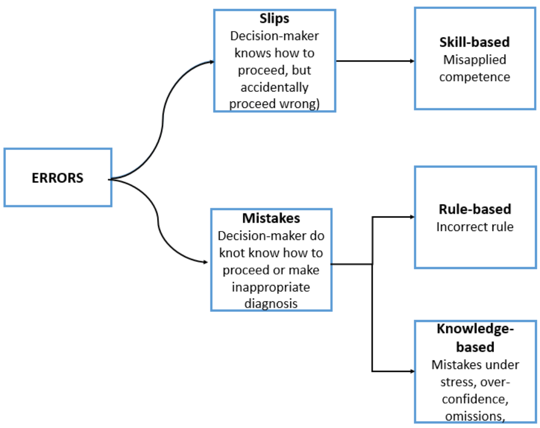
onto the persons involved when the work must also be done within environment and health and safety regulations. Typically, problems do naturally arise during civil engineering projects. If these problems were correctly identified, an adequate response can be formulated.

Researchers have distinguished six types of problems (Ishizaka & Nemery, 2013; Roy, 1981): (1) The choice problem (select best option or reduce the set of options), (2) The sorting problem (classification into categories), (3) The ranking problem (from the best to the worst), (4) The description problem (option and their consequences), (5) The elimination problem (related to sorting problem), and (6) The design problem (identification of the goal or creating a new actions). Civil engineers often face these problems while making decisions about work performance.

Even though modern technology helps to perform the work much more safely than in the past, there is always a risk involved. Some civil engineering works, such as excavation, are also very uncertain and require care. According to the literature (Lipshitz, Klein, Orasanu, & Salas, 2001), there are three forms of uncertainty: (1) inadequate understanding (insufficient situations awareness), (2) lack of information (incomplete or unreliable information) and (3) conflicted alternatives (difference between them is insufficient). Organisational units and people involved in the construction process have different experience and, even if they have the same data sources, their decisions may differ. Lipshitz (1993) identified nine NDM models. We would like to focus on three that we believe are applicable to the design project: (1) Rasmussen's model of cognitive control SRK, (2) Klein's Recognition Primed Decision model (RPD) and (3) Recognition/Metacognition model (R/M).

Rasmussen differentiated three kinds of behaviour (Rasmussen, 1983): (1) Skill-based, (2) Rule-based and (3) Knowledge-based. *Skill-based* behaviour arrives from experience. The decision is made based on past information stored in the subconscious. In case of r*ule-based* behaviour in decision-making, the unit will behave according to the rules set because these had generated good results in the past. This kind of behaviour is typical for decision-makers who do not have enough experience, but know the rules needed to accomplish a task. *Knowledge-based* behaviour is assigned to new situations when there is no set of rules available. The units must first identify the problem, generate the alternatives to solve it and, finally, choose the best solution.

Psychologists Gary Klein, Roberta Calderwood and Anne Clinton-Cirocco created the RPD model to describe how people make decisions. Decision-makers based their decisions on experience by connecting the current situation with the solutions performed in the past (pattern-matching). There are three variations of RPD models that differ according to their complexity (Klein & Klinger, 1991): (1) Simple Match RPS, (2) Developing a Course of Action RPD and (3) Complex RPD Strategy. The RPD model cannot be used in the case of knowledge-based behaviour, multi-criteria analysis cases and Search-for-Dominance Structure (SDS) strategy (Klein & Crandall, 1996). *Variation 1* describes a situation well known to decision-makers. Thus, they can easily simulate what the further scenarios are and notice the typical signs that indicate the use of particular solution. Conversely, in *variation 2,* the situation is not so easy to diagnose. Thus, decision-makers need to collect more data and/or assess the situation by looking into their past experience. Sometimes, decision-makers do not know which solution to select for the case (*variation 3*). However, in case of NDM methods, the choice of the best solution is not done by Multi-criteria Decision Analysis, but by generating a single action using mental simulation of possible scenarios. Nevertheless, even the most experienced decision-maker can make mistakes. The possibility of failure should always be considered. Human failures are distinguished by researchers (Embrey & Lane, 2005) from errors (which are related to the skill, rules and knowledge-based behaviour) and violations (which are exceptional or caused by routine). The types of errors are presented in Figure **19**.



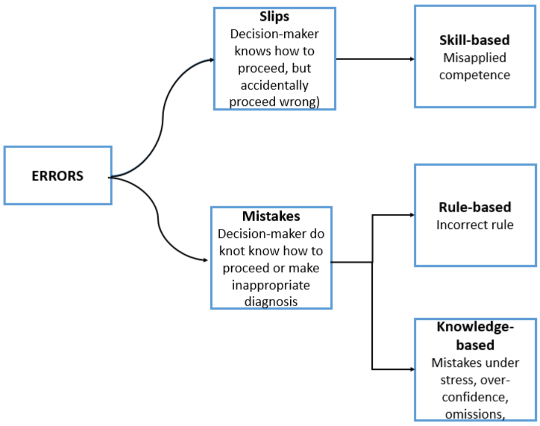


Figure 19.Types of errors (adopted from Embrey and Lane (2005))

The Recognition/Metacognition (R/M) model compares to the RPD focuses on the situations when decision-makers fail, while recognizing the case (Lipshitz et al., 2001). The incorrect situation recognition leads to the renewed data evaluation. Decision-makers needs to elaborate situation (recognition), add missed data and modify the strategy when there is: insufficient situations awareness, incomplete information or conflicted differences between alternatives (metacognition). The R/M model reconciles both pattern-matching recognition and problem solving strategies (Azuma, Daily, & Furmanski, 2006).

METHODOLOGY

We conducted eleven interviews and two workshops with decision-makers (designers and contractors), as well as with operators, data specialists, network operators and developers of applications for construction engineering industry.

Interviews for the first stage of project were held between March and June 2016. We prepared a set of questions and asked interviewees to prepare example(s) of projects that they found interesting regarding locating test trenches. To extract information about the user experience, practice rules, decision results and strategies we questioned each specialist about the number and location of test trenches, the reasons for decisions, opinions, support aids, results, communication and suggestions about changes. Each interview session took approximately two hours.

Workshops were divided into two sessions; one with six designers and (afterwards) one with six contractors. This division resulted from interviews when we noticed that the decisions made by designers and contractors were different (due to different stages of project). The first session took place in July 2016 and the second in October 2016. Both were based on serious gaming techniques (Schell, 2015) to check the behaviour of decision-makers when taking decisions under uncertainty. This approach was challenging as the participants had to face situations they were not prepared for and to explain their choices. We provided them with a map of underground utilities, extracted from Cadastre's klic system on which we superimposed the following projects: (1) build a new parking place, (2) place new fibre cables and connect them to houses and (3) replace an old sewage system. For each project, they had to decide how to locate test trenches and place them on map using transparent foils. We included obstacles, such as trees and waste containers, to make the assignment more difficult and influence the utilities path. The game board, together with examples of the results, are shown in Figure 20.

Figure 20. Workshop session with decision-makers

To analyse the collected data and information we used analytical and cognitive approaches to provide us with interesting results to generate further ideas about DSS design. In addition to the NDM model described earlier we also used the following methods: (1) The Design Process Unit (DPU) and Design for eXcellence approach (DfX) (Becker Jaruregui & Wessel, 2013), (2) System Engineering (Blanchard & Fabrycky, 2011), (3) Simon's Rational Decision-Making Model (Simon, 1977), and (4) Multi-criteria Decision Analysis (Ishizaka & Nemery, 2013). We discuss our results and adopt a NDM perspective in the following section.

RESULTS

During the interviews and workshops we observed a number of good and bad practices. The best practices showed that, in order to decide about the number and location of necessary test trenches, the following techniques needed to be combined:

- Data analysis (i.e. maps of underground utilities);
- Site analysis (i.e. ground conditions and obstacles);
- Detection tools (i.e. ground scanning tools);
- Use of best practices (i.e. provided by guidelines);
- Computer support (i.e. visualization tools and data bases to store data).

Only a few companies we interviewed used the combination of these methods, with others using just one or two methods. Clearly, using judgments is not incorrect. The experience, skills and knowledge of decision-makers are important to the decision process. Nevertheless, all best practices (experience, computer support, ground scanners, and utilities maps) can and should be joined together and used, not only by experts, but also by novices who might otherwise assess the excavation case wrongly.

It was interesting for us to check why damages still occurred, despite the fact that test trenches were dug. We analysed the most frequent cases of damage to assess whether the errors were due to poor risk assessment, a lack of data or due to rushed activities. The damage to utility services that occurred most often were: (1) mechanical damages, (2) damages to house connection, (3) damages while fibre cables were connected to houses, (4) damages to data-transport cables, (5) damages due to lack of information on utilities maps, (6) damages to disconnected utilities, (7) damages when tree was pulled together with utilities, (8) damages due to absence of Network Operator supervisor, (9) damages due to ground conditions (soil type, reduce visibility), and (10) damages while digging test trenches.

Those results directed us towards an analysis of decision-makers' behaviour. The interviews and workshops uncovered that decisions about the number and location of test trenches depended on the following factors: (1) stage of the project (orientation test trenches during design, test trenches dug before starting the excavation or test trenches dug when the reparation to the utility is necessary due to system failure), (2) type of work (i.e. building a new road or replacing the sewage system), (3) actions taken (i.e. drilling, open excavation or trenchless methods), and (4) decision-makers' and their experience.

Decision-makers used a set of good practices, such us digging test trenches, when there are doubts about the accuracy of utility maps, or when high risk utilities, such as gas pipelines, are present in the excavation area or many cables and pipes cross over one another. Designers dig test trenches to check if there is enough space to perform the project and to avoid unnecessary cost and changes. In contrast, contractors dig test trenches because they want to keep deadlines and to perform their work safely. To achieve this, they have to confirm, not only the location and depth of the utilities, but also check their material (e.g. resistance for vibration) and type of connections (e.g. can have impact on utilities depth).

During the workshop, we noticed how fast some decision-makers took decisions. This was clearly related to the knowledge and experience they had. In some cases, they could directly decide if the test trench was necessary or not, because they had faced similar situations before. For example, if on the excavation area a water pump station was present, the experienced decision-maker knew the depth of associated utilities can differ. Some decisions, such as checking the area up to 1,5 m from the excavation place, were influenced by (legal) rules (Crow, 2013) established according to best practices. Nevertheless, designers and contractors can face situations new to them. Furthermore, the decision as to whether an excavation should happen should not be made in a hurry, but proceed only after careful case-by-case analysis. If information was insufficient, then the information gaps need to be addressed before taking a decision so as to reduce uncertainty (R/M model). Thus, the situation should be well identified, the tasks made clear and planned properly using the best available techniques (i.e. maps, ground radars and best practices).

The current behaviour of decision-makers is presented using the SRK model in Figure 21.

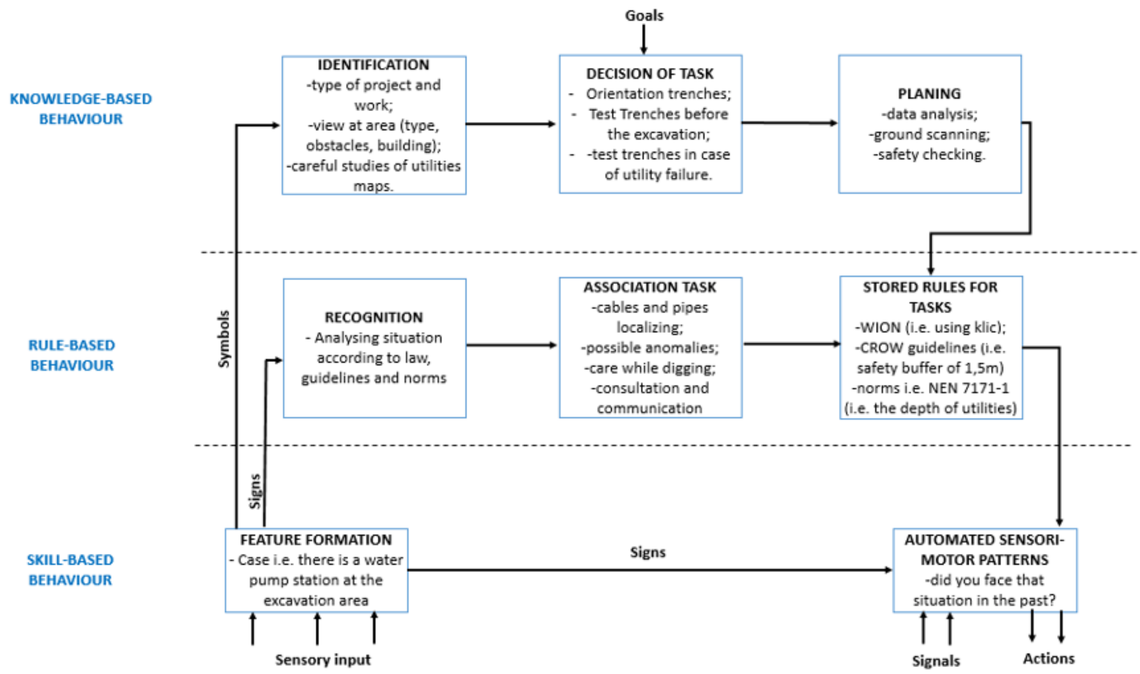


Figure 21. Rasmussen's model of cognitive control SRK for test trench location process (adopted from Rasmussen (1983))

Figure 21 is based on the example of the water pump station, which was discussed during workshop sessions. Only four of the twelve decision-makers decided that it was necessary to dig test trenches near the pump because the depth of the utilities can differ relative to the other parts of the excavation area. They supported their choice by using an example from their past experience (*skill-based behaviour*). Other participants agreed with their suggestion. The decision was also influenced by the set of rules (*rule-based behaviour*). To reduce the risk of failure, decision makers used, for example, the 1,5 m buffer. This ensures that no utility will be damaged while executing the work. Nevertheless, new situations could be identified using available technologies *(knowledge-based behaviour*). During the interviews, we observed cases where damage occurred because of a lack of a correct situation assessment. For instance, in one case, the age of buildings influenced the material and depth of the utilities and this was not taken into account by decision-makers.

Many strikes are attributable to the lack of the information on the utilities maps. However, the relation between SRK behaviours and the failures can be noted as well. The damages to the utilities caused by:

- removing the trees ;
- absence of the network operator supervisor,
- excavation machines (if utility marked on klic), and
- misunderstandings;

could be avoided by correct assessment of the situation.

Experience sharing could be an interesting component part of a future DSS. The excavation case could be analysed using, not only data from radars, maps, documentation and observations, but also by matching the situation to ones that had occurred in the past. The RPD and R/M models related to pattern-matching could be used by creating an experience database within the DSS. The new situations faced by decision-makers and best practices could be collected and updated, for example: how obstacles can influence the location and depth of utilities, the resistance of the material for vibrations, and so on. Based on the input information, the tool could look for patterns (using the RPD model variations) using algorithms and to present this to decision-makers by (e.g. warning signals). The information about previously dug test trenches can be also collected. So, if it is necessary to dig new test trenches, the decision-makers can have an overview of the past decisions and findings. Together with other techniques, this would help generate a more reliable risk assessment. A combination of analytical and naturalistic decision-models clearly could support better situation awareness, help to collect enough reliable information and, last, but not least, reduce uncertainty related to conflicts between alternatives approaches (i.e. dig one, two or three test trenches).

CONCLUSIONS

Digging test trenches is one method used to confirm the exact location of the cables and pipes. Often, the decision about their number and location appears random and driven by the personal judgments of the individual decision makers. The interviews and workshops conducted uncovered strategies used by the decision-makers (designers and contractors). They used good practices, such as digging test trenches when there is uncertainty about the accuracy of utilities maps, when high-risk utilities are present, or when many cables and pipes cross over one another. The regulatory rules and other guidelines can also help to reduce uncertainty (e.g. the rule of adding the 1,5 m checking buffer to ensure that the excavation project will not damage utilities).

The knowledge and skills of decision-makers both play an important role in the decision-process. It influences how many test trenches will be dug and where they will be located. The use of different techniques, such as detection tools, utility maps, visualisation programmes and data management tools, can support decision process. However, decision-makers ultimately assess the risk of damage. Their experience levels differ. They often face new situations. These can involve poor decisions or incorrect assumptions if they are having to analyse large amounts of data. Many people involved in the excavation process are not sufficiently familiar with the use of ground scanning methods. This can result in further incorrect assumptions about the situation. Full automation of decision-making regarding how and where to locate test trenches appears to be untenable. Neither is it desirable, as it prevents leveraging expertise from those directly involved.

A combination of analytical and naturalistic decision-making elements within one DSS seems to be an interesting and viable option. Pattern-matching could give additional input for decisions, especially when decision-makers lack experience or when new situations have to be faced. An experience database could be created by the users and updated by adding information about the best practices, findings, failures, obstacles that influence the location of utilities and digging test trenches. The DSS could search the database to find similar cases to the project under consideration. Moreover, based on all collected data and conducted analyse, the risk of damage (including the case where test trenches are not used) can be calculated.

We argue that DSS which collects all the necessary data, analyses them, visualises them and, in addition, compares them with best practices and previous experience, can reduce all of the aforementioned three forms of uncertainty mentioned in the literature (Lipshitz et al., 2001) (inadequate understanding, lack of information and conflicted alternatives). The system of pattern matching within the tool can be based on the RPD model by comparing the current situation with others in the database (variation 1), requiring additional input information (variation 2), or informing about the necessity of “evaluation of course of actions” by, for instance, risk calculation of the generated scenario (variation 3). The RPD model can be supported by using the R/M model and inform the users about any deficiency in information which is needed to generate a viable solution. Of course, these ideas await further elaboration and testing. Nevertheless, we are confident that they will help shape the thinking about tomorrow’s DSS for selecting the nest locations for test trenches.

ACKNOWLEDGMENTS

We would like to thank all the companies and their representatives who participated in the interviews and workshops and who supported us with their knowledge regarding the excavation process of selecting test trench locations.

REFERENCES

Azuma, R., Daily, M., & Furmanski, C. (2006). *A review of time critical decision making models and human cognitive processes*. Paper presented at the *IEE Aerospace Coference*, Big Sky, MT, USA.

Becker Jaruregui, M. J., & Wessel, W. W. (2013). *An information model for product development: a case study at PHILIPS Shavers*. Paper presented at the *2nd CIRP Global Web Conference*.

Blanchard, S. B., & Fabrycky, J. W. (2011). *System Engineering and analysis* (5th ed.). Essex: Pearson.

Crow. (2013). *Instructiekaart zorgvuldig graven-graafschade voorkomen aan kabels en leidingen*. The Netherlands.

Embrey, D., & Lane, H. (2005). Understanding Human Behaviour and Error. *Human Reliability Associates*, 1-10.

Groot, R. D. (2008). *Graafschade voorkomen aan kabels en leidingen: Richtlijn zorgvuldig graafproces*.

Ishibushi, H., & Nii, M. (2000). Neural networks for soft decision making. *Fuzzy Sers and Systems, 115*, 121-140.

Ishizaka, A., & Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. West Sussex, United Kingdom: John Wiley & Sons Ltd.

Kabel en leiding overleg KLO. (2016). Aantal graafschades opnieuw gedaald. Retrieved from http://www.kabelenleidingoverleg.nl/nieuws/2016-04-19-aantal-graafschades-opnieuw-gedaald

Kadaster. (2016). klic-win. Retrieved from https://www.kadaster.nl/klic-win

Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456-460.

Klein, G., & Crandall, B. (1996). *Recognition-Promed Decision Strategy*. Retrieved from Alexandria:

Klein, G., & Klinger, D. (1991). Naturalistic Decision Making. *Human System IAC GATEWAY, 11*(3), 16-19.

Lipshitz, R. (1993). *Converging themes in the study of decision making in realistic settings*. Norwood.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Focus Article: Taking Stock of Naturalistic Decision Making. *Journal of Behavioral Decision Making, 14*, 331-352.

Racz, P., Van Buiten, M., & Dorée, A. G. (2016). *To dig or not to dig: How to determine the number and location of test trenches*. Paper presented at the *13th International Conferencce on Design & Decision Support System in Architecture and Urban Planning*, Eindhoven, The Netherlands.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinction in Human Performance Models. *IEE Transactions on Systems, Man, and Cybernetics, 13*(3), 257-266.

Reggefiber, & University of Twente. (2015). *The ZOARG program*. Retrieved from http://www.zoarg.com/

Roy, B. (1981). The optimisation problem formulation: Criticism and overstepping. *Journal of the Operational Research Society, 36*(6), 427-436.

Schell, J. (2015). *The Art of Game Design. A book of Lenses*. Pittsburgh: CRC Press Taylor & Francis Group.

Simon, H. (1977). *The New Science of Management Decision*. Englewood Cliffs: Prentice Hall.

Beyond Macrocognition: The Transaction Level

Jan Maarten SCHRAAGEN[a],
[a]TNO/University of Twente, Soesterberg/Enschede, The Netherlands

ABSTRACT

The purpose of this paper is to respond to a number of developments in the past 15 years in the areas of NDM and macrocognition. In my view, these developments necessitate the definition of a new systems level, extending and modifying Newell's (1990) "bands of cognition". I will argue that what has been called 'microcognition' may best be viewed as a field compatible with the 'cognitive band', whereas 'macrocognition' is compatible with Newell's 'rational band'. Recent developments in various fields necessitate the introduction of a next systems level, the 'transaction level'. I will discuss the interconnection amongst the various levels. Several studies using social network analysis are discussed that illustrate the utility of adding this new level to our current repertoire of methods.

KEYWORDS

Theory and modelling; Macrocognition; Cognitive systems engineering; resilience; general.

INTRODUCTION: STATE OF THE ART

As a movement, NDM consists of applied researchers who are interested in how professionals make decisions in real-world situations, with the goal of supporting these professionals through decision aiding and training. Although NDM has broadened its scope under the flag of macrocognition to include other processes than 'decision making', its focus is still 'cognitive'. From the beginning, the focus of the movement has been on experts rather than novices, and on real-world situations rather than the controlled world of the laboratory. Although this has led to numerous insights, such as the Recognition-Primed Decision Model (Klein, 1993), theoretical progress seems to have halted over the past years. The reason for this state of affairs, it seems to me, can be traced back to a single a priori assumption that governs the practice of NDM, namely that expertise is a cognitive and knowledge-rich phenomenon. This assumption is shared by many others in the cognitive science and AI community as well, and goes back at least to the early 1970s where psychological studies of expertise demonstrated the importance of knowledge over general cognitive processes (e.g., Simon & Chase, 1973) and the early 1980s which saw the rise of the « Knowledge-is-Power » hypothesis in AI (Lenat & Feigenbaum, 1991). It is in line with Simon's (1981; 1991) views on the adaptability of artificial systems, which states that systems that are completely adapted to their environments (and experts are a prime example of those systems) will not show any of their architectural constraints. It is not possible here to flesh out the complete history of this view nor all of its theoretical ramifications or supporting empirical evidence. This view, however, leads to a subtle problem, already noted by Simon (1980), in that adaptive or artificial systems will change as their environment changes. Therefore, it is difficult to discover and verify empirical invariants in these systems, as any laws that govern their behavior must contain reference to their relativity to environmental features. As noted by Simon (1980, p.36) : « It is common experience in experimental psychology, for example, to discover that we are studying sociology—the effects of the past histories of our subjects—when we think we are studying physiology—the effects of properties of the human nervous system ». In other words, when NDM studies expert behavior, it will mostly describe the environments the experts adapt to, rather than any invariants in expert behavior itself. This may easily lead to a rather anecdotal, domain-specific, and atheoretical state of affairs within NDM, where the primary breakthroughs come from the study of a novel domain that has not been studied so far.

Simon goes on to state, however, that we should not despair of finding invariants. Rather than looking for absolute invariants, as the physicists do, cognitive science should search for relative invariants that hold over considerable stretches of time and ranges of systems : « What is invariant in adaptive systems will depend on the time intervals during which we observe them » (Simon, 1980, p. 36). Simon then mentions three time scales on which adaptation takes place : the shortest time scale in which heuristic search takes place ; a longer time scale in which learning (storing and retrieving knowledge) takes place ; and the longest time scale during which social and biological evolution takes place, in the form of discovery of new knowledge and strategies and the transmission of one system to another.

Newell (1982, 1990) has later expanded on this view by taking the time scales Simon (1980) described as his starting point. Newell (1990) distinguished a biological, cognitive, rational, and social 'stratum', each defined by a particular time band during which processes take place. For instance, the typical cognitive processes take place between 100 msec and 10 sec and the typical rational processes between minutes and a few hours. However, Newell (1982) went one step beyond Simon and equated these time scales with system levels that occur in nature

(in his 1982 article, Newell confined these system levels to computer system levels, but it became apparent in his 1990 book that he meant these system levels to apply to all biological and artificial systems).
In the next paragraph, I will expand upon the notion of 'system levels', as I believe that the introduction of a new system level could bring significant advances to the study of artificial systems, going beyond macrocognition and current views of NDM.

SYSTEM LEVELS

There are many definitions on what a 'system level' is. According to Newell (1982), a true system level is a reflection of the nature of the physical world, not simply a level of abstraction. It should be a specialization of the class of systems capable of being described at the next (higher) level. Aggregation occurs within each level and does not take us to the next level on its own; meaning should be added and some things therefore become invisible at the next level (this is another way of saying that phenomena at higher levels have emergent properties). Therefore, although the levels are ontologically irreducible to each other, each level may still be implemented at the next lower level.

Note that this view of system levels is quite different from what we normally take to be 'units of analysis'. For instance, the distinction between micro, meso and macro levels (individual, group, organization) is not a true system level description, in Newell's definition. For instance, Karsh, Waterson, and Holden (2014) proposed 'mesoergonomics' as a way to specify macro- and microergonomic integration. Their aim was to reveal cross-level interactions and to describe relationships between and among levels rather than describing phenomena that emerge from their components but that cannot be explained by them (Hackman, 2003). For instance, Carley and Newell (1984) described what a Model Social Agent would look like, in an attempt to make an individual agent behave like a social agent. They concluded that a Model Social Agent could be obtained by minimal requirements on information processing capabilities but with major extensions in the types and amount of knowledge required. Note this is a matter of aggregation and does not take us to a next system level on its own. We obtain social behavior by basically adding more knowledge to an existing system level (which Newell called the 'knowledge level' and later the 'rational band'). Newell (1990) even doubted whether a separate social band existed, due to the lack of invariants found up until then. In the paragraph below, I will note some general developments that justify the introduction of a new level, situated right above the knowledge level.

DEVELOPMENTS

In the past 15 years, some notable developments have taken place in the fields related to Cognitive Systems Engineering (CSE) and Naturalistic Decision Making (NDM), as well as in some other, unrelated, fields. Without being exhaustive, the following developments may be noted:

- The extension of NDM to macrocognition. Some particular models developed under the guise of macrocognition are the Data-Frame model of Sensemaking (Klein, Phillips, Rall, & Peluso, 2007) and the Flexecution model of Replanning (Klein, 2007a,b). These models depart from the linear Input-Output models and stress the dynamics of (re)framing, elaboration and preservation in cycles of macrocognitive activity.
- The extension of the situation awareness concept from being strictly 'in the head' to being a distributed, emergent process, arising from the dynamic interplay of humans and technological artefacts (e.g., Stanton, Salmon, Walker, & Jenkins, 2010).
- The Interactive Team Cognition approach (Cooke et al., 2013) provides a theoretical alternative to existing approaches that exclusively focus on individual team members' cognitive processes. Interactive Team Cognition views team cognition as context-dependent team interaction rather than a monolithic entity that a team can either have or not have. Rather than viewing team cognition as something that is shared among team members and then aggregated, it is viewed as an interdependent network that should be studied at the team level. Team cognition arises as an emergent property as team members communicate with each other (Walker et al., 2006).
- The move from a dyadic semiotic model of 'meaning' that views the mind as a symbol-processing system to a triadic semiotic model that suggests that a sign's meaning is grounded in the functional dynamics of the problem domain. This particular development can be traced back to ecological psychology (Gibson, 1979), and from this lineage to Rasmussen, Pejtersen, & Goodstein (1994) and Flach (2015). The implication of this philosophical view is that the focus of applied work has shifted from human-computer interaction to interactions between humans and work domains, mediated by interfaces and automation (Flach, 2015). The same shift may be noted when comparing (cognitive) task analysis with work analysis (Vicente, 1999).
- The shift from CSE to Resilience Engineering may be viewed as a shift from the engineering of cognitive systems to engineering with a particular purpose in mind, i.e. achieving sustained adaptability (Woods, 2015). The aim is still to 'outmaneuver complexity' by 'graceful extensibility', but the focus has shifted from cognitive systems to 'tangled layered networks'. There is evidence that the particular network structure that has evolved determines the long-term success of organizations to deal with disturbances (Rivkin & Siggelkow, 2007; Stanton, Walker, Sorensen, 2012). In a variety of domains, scale-free network structures

have been shown to be resilient architectures, in the sense of being robust to the (random) removal of nodes (Schraagen, 2015). Specifically, in such resilient architectures the degree distribution of nodes follows a power law.
- The introduction of the network and relational perspective in management science (Hollenbeck & Jamieson, 2015) and the social sciences in general (Borgatti, Mehra, Brass, & Labianca, 2009) and the concomitant rise of social network analysis as a methodology (Wasserman & Faust, 1994).

Summarizing these developments, we may note that there is a move away from viewing cognitive agents as independent units to viewing them as part of larger networks of interconnected units of adaptive behaviour. These networks are constrained by factors external to them, including time, physical constraints, task, technology, social structure, and culture. The networks are cross-cutting, embedded in each other, and heavily intertwined. As units of adaptive behaviour interact over time, regularities in behavior emerge. These regularities may be visible in the form of 'patterned interactions' and may be studied by methods such as social network analysis. The implications of these developments for NDM have not yet been discerned. In the next paragraph, I will connect the micro-macro cognition distinction to Newell's system level constructs and then build up to the new "Transaction level".

MACROCOGNITION AND MICROCOGNITION AS SYSTEM LEVEL CONSTRUCTS

Macrocognition as a field has defined itself as an extension to the Naturalistic Decision Making field, in the sense that its focus is on a broader set of macrocognitive functions than merely decision making. For instance, it also distinguishes sensemaking, planning, adaptation, problem detection and coordination as important macrocognitive functions. Macrocognitive functions need to be performed by individuals, by teams, by organizations, and by joint cognitive systems that coordinate people with technology. It is the study of cognitive adaptations to complexity (Schraagen, Klein, & Hoffman, 2008).

Macrocognition is distinguished from microcognition primarily by its time scale of analysis. Whereas microcognition focuses on cognitive processes in the time band of 100 msec up to 10 sec., macrocognition focuses on cognitive processes from minutes to hours. The preferred means by which these processes are studied then also varies, with microcognition frequently opting for constrained tasks in confined environments with high experimental control, while macrocognition opting for real-life tasks under actual working conditions with less experimental control (Cacciabue & Hollnagel, 1995; see Hoffman & McNeese, 2009 for a historical overview).
Interestingly, although macrocognition in principle extends to the organizational level, in practice there are very few studies that adopt the methods employed by organizational scientists, nor are the time scales adopted in macrocognitive studies weeks or months or years, but rather hours at the upper level. Methodologically, then, an important shortcoming in current macrocognitive studies is the lack of longitudinal data collection, which prohibits the discovery of emergent behaviors at longer time scales.

If microcognition is equated with Newell's cognitive band and macrocognition with Newell's rational band, then it follows that macrocognition is a 'knowledge level' construct (Newell, 1982). At the 'knowledge level', the principle of rationality applies: if an agent (e.g., an expert) has a goal and knows that knowledge A will bring him or her closer to that goal, then the agent will choose knowledge A. From the outside, the behavior of the expert is highly predictable, once we know the expert's goals and the knowledge that is required to attain those goals. For the NDM community, the principle of rationality should look familiar: it is functionally, if not logically, equivalent to the RPD model. The expert has a representation that can be accessed without any computational costs, or, in Newell's (1982) 'slogan equation': Representation = Knowledge + Access. The representation consists of a system for providing access to a body of knowledge. Access is a computational process, hence has associated costs. Thus, a representation imposes a profile of computational costs on delivering different parts of the total knowledge encoded in the representation (Newell, 1982). But if we look closer, we notice that this equation generalizes beyond the RPD model. In fact, the RPD model is a special case that applies when computational costs are low, i.e., the knowledge can be accessed highly efficiently to make selections of actions in the service of goals. In other (unfamiliar, non-routine) cases, knowledge may not be accessed as efficiently, and it is here that the macrocognitive 'deliberative search' processes are put to work to access the knowledge.

A NEW LEVEL

I now propose that there does exist yet another system level, which I will call the *transaction level*. It is a true systems level in Newell's sense, that is, it is a reflection of the nature of the physical (and social!) world and not just a point of view that exists solely in the eye of the beholder. It is not a level of analysis that can be applied to any unit of analysis, as, for instance, a network perspective that can be applied to both brains and societies. It is not an aggregation of knowledge, so its behavior cannot be obtained by expanding agents at the knowledge level with simply more knowledge. Rather, no amount of knowledge added will yield the transaction level properties that are characteristic of this level.

A quick overview of the transaction level, in terms of the system under consideration, its components, its laws of composition, its behavior laws and its medium, are in order before entering into details.

The system at the transaction level, the entity to be described, is the *network*. The system's primitive elements, its components, are *nodes* and *links*. Thus, a network is composed of a set of agents and a set of links. The components are assembled into systems by laws of composition that yield *strength* and *reciprocity*. The medium at the transaction level is the *transaction* (as might be suspected). Thus, the network generates transactions by connecting nodes through links. The transactional content may differ widely, from affect and influence to goods and services, and information. Finally, the behavior law, how the system depends upon its components and composition, is the *principle of relationality* : links are selected to attain transactions. As links are characterized by strength and reciprocity, the generation of transactions is dependent upon link strength and reciprocity.

In contrast to the knowledge level, the concept of 'goal' does not play a role at the transaction level. However, just as with the knowledge level, the transaction level is a radical approximation : entire ranges of behavior may not be describable at the transaction level, but only in terms of systems at a lower level. For instance, the transaction level is poor for predicting how team members that have never met before will interact. It is also poor for predicting the effectiveness of the introduction of a new technology on a person's behavior. However, it is good for predicting the impact of losing someone central to an organization's informal network. It is also good for predicting that a well-established team will exchange relevant information in a timely fashion.

The physical structure of a transaction is filled indirectly and approximately by knowledge systems at the next lower level. This is depicted in figure 1 :

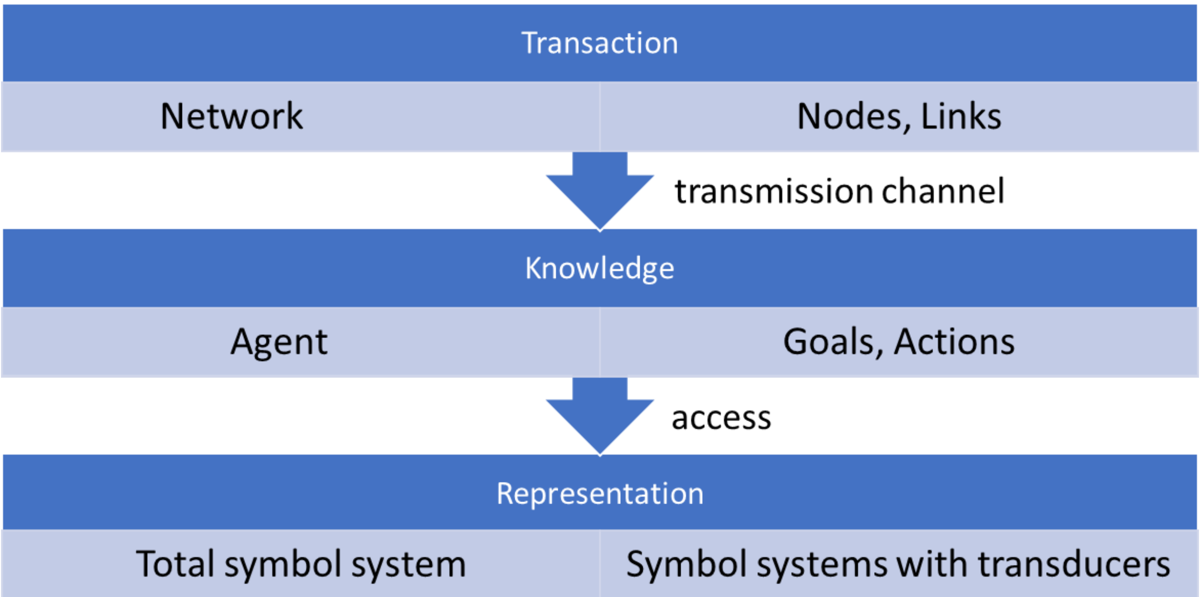


Figure 1. Reduction of the transaction level to the knowledge level and the symbol level.

The slogan equation for connecting the transaction level to the knowledge level is: Knowledge = Transaction + Transmission channel. What this means in psychological terms is that humans are social beings that have knowledge of all kinds of possible transactions, but only to a certain extent, due to limitations on communication bandwidth. This knowledge is a person's *social capital*, or the interpersonal relationships they have with others that enable successful functioning (Hollenbeck & Jamieson, 2015). Social capatial is conceptually distinct from the individual-level constructs and it has the potential to enhance prediction of individual and team performance incrementally over those constructs. Social capital information can be quantified using social network analysis. In the next paragraph, I will provide a few examples of recent research carried out at the transaction level.

SOCIAL NETWORK ANALYSIS APPLIED TO MILITARY, MEDICAL, AND SPORTS TEAMS

Social network analysis (SNA) (Wasserman & Faust, 1994) is a promising approach for studying real-time team interaction at the transaction level. It is not dependent on the availability of trained raters, as it only needs to take into account a network consisting of a set of units (or actors) and the relations that exist among them (Alba, 1982). I have applied SNA in two case studies: a medical team performing paediatric cardiac surgical procedures (Barth, Schraagen, & Schmettow, 2015), and two naval teams (Schraagen & Post, 2014).

In the medical teamwork study, degree centralisation, density, closeness centralisation, betweenness centralisation and reciprocity were chosen as the metrics of centralisation, in order to capture context-dependency of communicative behaviour within a medical team. Compared to classical behavioural rating schemes, these metrics of centralisation are not influenced by hindsight bias as they are based on real-time communication and not established immediately after the surgical procedure. Fourty surgical procedures were observed by trained human factors researchers, and communication processes amongst team members were recorded. Focusing on who talked to whom, team communication structures, in response to changing task demands, were characterised by various network measures. Results showed that in complex procedures, the communication patterns were more decentralised and flatter. Also, in critical transition phases of the procedure, such as going on or off cardio-pulmonary bypass, communication was characterised by higher information sharing and participation.

In particular, we found that in any given phase, there were always two team members linked to many others, thus scoring high on total degree centrality. Not surprisingly, the primary surgeon was always one of these, with the anaesthetist and perfusionist being the second actor, depending on the surgical phase. We also looked at complex versus non-complex procedures and found that during complex procedures the role of the assisting surgeon increased relative to the role of the primary anaesthetist, especially when going on or off cardio-pulmonary bypass. Although the primary surgeon still scored highest in total degree centrality in virtually all cases, the assisting surgeon filled in the role of communicator to the rest of the team whenever the workload of the primary surgeon prevented him from speaking to the rest of the team. This form of ‘heedful interrelating’ (Schraagen, 2011) shows that this team is at least adaptive, if not resilient.

The goal of the naval study was to apply SNA techniques to the study of naval teamwork, to be able to characterize naval team readiness. As part of a larger study to develop methods for socio-technical systems analysis applied to Oceangoing Patrol Vessels (OPV), we conducted ethnographic observations on board of two OPVs (see Post et al., 2013, for details). The ethnographic observations entailed making video recordings of two ‘internal battle’ exercises. To be able to distinguish between different levels of naval team readiness, we studied an ‘unpractised’ team and a ‘team in training’ (in more conventional terms, the first could be referred to as ‘beginners’, whereas the second team could be referred to as ‘intermediate’; however, we prefer to use our more neutral terms, as they do not imply any rank ordering on a non-existing scale). The descriptive, non-inferential, results showed that, at the transaction level, the more experienced team displayed higher levels of information sharing and team member participation compared to the less experienced team. At the actor level, the team coordinator played a much more central role in the more experienced team, whereas in the less experienced team this role was taken up by various other team members.

For both the medical and the more experienced military teams, the results clearly showed that the total degree centrality displayed a power-law distribution, indicating that the teams adopted a scale-free network structure (Schraagen, 2015). This means that one actor was highly connected to a few other actors, while the remaining actors were less well connected. We hypothesize that this particular structure is able to deal with disturbances that are not well modelled, unexpected events that an OR team or a naval team such as we have studied needs to deal with as the occasion arises. We believe therefore that, in fact, these are network architectures that can sustain the ability to adapt to future surprises as conditions evolve (Woods, 2015).

Finally, as an example of recent study in the domain of sports, Dalal, Nolan, and Gannon (2016) studied whether professional ice hockey players’ Olympic Tournament team performance could be predicted from their pre-assembly shared work experiences, that is, the extent to which pairs of teammates were familiar with each other’s work patterns, for instance because they had played together in the same division, league or position. In the Olympic hockey tournament case, team members first interacted as a team about two days before the tournament started. This question could be answered from a shared mental models perspective, which would be a knowledge level perspective. However, Dalal et al. (2016) used social network analysis measures to answer this question, taking a transaction level perspective. This perspective should be good at predicting how team members’ prior familiarity with each other’s work patterns will impact team performance. The results showed that as teams became more centralized around single members, performance decreased, basically confirming our results obtained with medical teams (Barth et al., 2015). However, a lot of variance was unaccounted for, confirming the notion that the transaction level is radically incomplete, and that the knowledge level (e.g., team mental models, team satisfaction) needs to be taken into account as well.

CONSEQUENCES

One may well wonder whether the introduction of a new system level, above the knowledge level, has any added value, in terms of surprising features, and will significantly advance the field of NDM. One surprising feature is that we may study human-machine teaming or ‘joint activity’ (Klein et al., 2004) not merely from the perspective of ‘shared goals’ and ‘joint actions’, which would be a knowledge level analysis, but also from the perspective of patterned interactions between ensembles of humans and machines. This brings with it the necessity of collecting longitudinal data of human-machine interaction. Another surprising feature is that, whereas the knowledge level has a complete absence of structure, the transaction level is characterized by high degrees of structure. Particular network structures are more robust to disturbances than other network structures (Schraagen, 2015). A third surprise is that the transaction level provides the context for when macrocognitive functions come into play, namely whenever the principle of relationality does not apply, i.e. when transactions do not occur even when nodes are connected through links. Put differently, the function of macrocognition is to transmit knowledge fluently throughout the network.

Whether the introduction of the transaction level will significantly advance the field of NDM remains to be seen, of course. The increasing use of social network analysis is merely one step in the right direction. Perhaps more significant will be the broadening of perspective that will result from embracing the transaction level. This level makes us aware that instead of focusing on single agents (commanders, officers, captains, pilots, etc.), we need to also take the broader network into account in which these agents are embedded, a network that increasingly also consists of intelligent machines. If anything, a conscious deliberation of where to draw the network boundaries would be extremely helpful in enriching our stories.

CONCLUSION

Finding relative invariants in adaptive systems takes place across different time scales. So far, the focus of NDM and macrocognition has been on the shorter to medium time scales, to the neglect of longer time scales. If nothing else, this paper is a call for more studies at the longer time scales. By introducing another systems level, the transaction level, I have attempted to describe the concepts that are required to carry out such studies. Although these concepts are closey related to social network analysis, they should not be equated with this particular method of analysis. Rather, taking a broader network-level perspective may enrich NDM and help find relative invariants across broader time scales.

REFERENCES

Alba, R. D. (1982). Taking stock of network analysis: A decade's results. In S. B. Bacharach (Ed.), *Research in the sociology of organizations* (pp. 39-74). Greenwich, CT: JAI Press.

Barth, S., Schraagen, J.M.C., & Schmettow, M. (2015). Network measures for characterizing team adaptation processes. *Ergonomics, 58*(8), 1287-1302.

Borgatti, S.P., Mehra, A., Brass, D.J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*(February 13), 892-896.

Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of cognition: Applications. In J. M. Hoc, P. C. Cacciabue, & E. Hollnagel (Eds.), *Expertise and technology: Cognition and human-computer cooperation* (pp. 55–73). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carley, K., & Newell, A. (1994). The nature of the social agent. *Journal of Mathematical Sociology, 19*(4), 221-262.

Cooke, N.J., Gorman, J.C., Myers, C.W., & Duran, J.L. (2013). Interactive team cognition. *Cognitive Science, 37*, 255-285.

Flach, J. (2015). Situation awareness: Context matters! A commentary on Endsley. *Journal of Cognitive Engineering and Decision Making, 9*(1), 59-72.

Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Hackman, J.R. (2003). Learning more by crossing levels: evidence from airplanes, hospitals, and orchestras. *Journal of Organizational Behavior, 24,* 905-922.

Hoffman, R.R., & McNeese, M.D. (2009). A history for macrocognition. *Journal of Cognitive Engineering and Decision Making, 3*(2), 97-110.

Hollenbeck, J.R., & Jamieson, B.B. (2015). Human capital, social capital, and social network analysis: Implications for strategic human resource management. *Academy of Management Perspectives*, *29*(3), 370-385.

Karsh, B.-T., Waterson, P., & Holden, R.J. (2014). Crossing levels in systems ergonomics: A framework to support 'mesoergonomic inquiry'. *Applied Ergonomics, 45*(1), 45-54.

Klein, G. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex.

Klein, G. (2007a). Flexecution as a paradigm for replanning, Part 1. *IEEE Intelligent Systems, 22*(5), 79-83.

Klein, G. (2007b). Flexecution, part 2: Understanding and supporting flexible execution. *IEEE Intelligent Systems*, *22*(6), 108-112.

Klein, G., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems,* May/June, 81-85.

Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R., &Feltovich, P.J. (2004). Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intelligent Systems*, November/December, 91-95.

Klein, G, Phillips,J.K., Rall, E.L., & Peluso, D.A. (2007). A data-frame theory of sensemaking. In R.R. Hoffman (Ed.), *Expertise out of context* (pp. 113-158) (Proceedings of the Sixth International Conference on Naturalistic Decision Making). New York: Lawrence Erlbaum Associates.

Lenat, D.B., & Feigenbaum, E.A. (1991). On the tresholds of knowledge. *Artificial Intelligence, 47,* 185-250.

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18,* 87-127.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Post, W.M., Grootjen, M., Kuiper, H., Rakhorst-Oudendijk, M.L.W., & Schraagen, J.M.C. (2013). Methods for sociotechnical system analysis applied to OPV Operational Maintenance (TNO Report 2013 R11327). Soesterberg, The Netherlands: TNO Behavioural and Societal Sciences.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York, NY: Wiley.

Rivkin, J.W., & Siggelkow,N. (2007). Patterned interactions in complex systems: Implications for exploration. *Management Science, 53(7),* 1068-1085.

Schraagen, J.M.C., Klein, G., & Hoffman, R.R. (2008). The macrocognition framework of naturalistic decision making. In J.M. Schraagen, L.G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp. 3-25). Aldershot, Hampshire: Ashgate Publishing Limited.

Schraagen, J.M.C. (2011). Dealing with unforeseen complexity in the OR: The role of heedful interrelating in medical teams. *Theoretical Issues in Ergonomics Science, 12*(3), 256-272.

Schraagen, J.M.C., & Post, W.M. (2014). Characterizing naval team readiness through social network analysis. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting* (pp. 325-329), Chicago, Il., October 27-31, 2014. Santa Monica, CA: Human Factors and Ergonomics Society.

Schraagen, J.M.C. (2015). Resilience and networks. *Proceedings of the Resilience Engineering Association 6th Symposium*, Lisbon, Portugal, June 22-25. Downloaded from: http://www.resilience-engineering-association.org/download/resources/symposium/symposium_2015/Schraagen_J.M.-Resilience-and-networks-Paper.pdf

Simon, H.A. (1980). Cognitive science : The newest science of the artificial. *Cognitive Science, 4*, 33-46.

Simon, H.A. (1981). *The sciences of the artificial* (2nd ed.). Cambridge, MA: The MIT Press.

Simon, H.A. (1991). Cognitive architectures and rational analysis: Comment. In K. VanLehn (ed.), *Architectures for Intelligence* (pp. 25-39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H.A., & Chase, W.G. (1973). Skill in chess. *American Scientist, 61*, 394-403.

Stanton, N.A., Salmon, P.M., Walker, G.H., Jenkins, D.P. (2010). Is situation awareness all in the mind? *Theoretical Issues in Ergonomics Science, 11*(1-2), 29-40.

Stanton, N. A., Walker, G.H., & Sorensen, L.J (2012). It’s a small world after all: Contrasting hierarchical and edge networks in a simulated Intelligence analysis task. *Ergonomics, 55*(3), 265–281.i

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum.

Walker, G.H., Gibson, H., Stanton, N.A., Baber, C., Salmon, P., & Green, D. (2006). Event analysis of systemic teamwork (EAST): A novel integration of ergonomics methods to analyse C4i activity. *Ergonomics, 49* (12-13), 1345-1369.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University.

Woods, D.D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety, 141*, 5-9.

Macrocognition in Flight Operations: Representing distributed teamworking in sociotechnical systems

Neville A STANTON, Katherine L PLANT, Kirsten M A REVELL, Thomas G C GRIFFIN, Scott MOFFAT and Maggie J STANTON
Human Factors Engineering, Transportation Research Group, Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton, SO16 7QF, UK.
n.stanton@soton.ac.uk
+44 (0) 1794 518856

ABSTRACT

The study reported in this paper used the Event Analysis of Systemic Teamwork (EAST) method to examine aviation operations from multiple perspectives (Dispatch, ATC/ATM, Maintenance, Loading, and the Cockpit). These networks were created for five key phases of flight: (i) crew briefing, (ii) preflight checks and engines start (iii) taxi and take-off, (iv) descent and landing, and (v) taxi, park and shutdown. The networks have been produced as an ‘information audit’ in order to understand the interactions and connections within the current system. .

KEYWORDS

Aviation, Networks, STS, EAST.

INTRODUCTION

The Aviation is a sociotechnical ‘system of systems’, encompassing technical, human, and organisational aspects (Harris and Stanton, 2010). Within these systems there are distinct operational independencies (aircraft operations; maintenance; air traffic management/control) and each of these aspects has managerial independence (they are run by independent companies or national providers); however, they are bound by a set of common operating principles and international regulations for design and operation. The inherent complexity of these operations is difficult to capture in their entirety, as they are distributed both in time and space. To overcome the challenges of modelling distributed cognition, Stanton and colleagues devised the EAST method (Stanton et al., 2008). EAST is underpinned by the notion that complex collaborative systems can be meaningfully understood through a network-of-networks approach. The EAST methodology is a contemporary approach for analyzing and modelling distributed cognition. EAST is underpinned by the notion that complex collaborative systems can be meaningfully understood through a network-of-network approach. The networks show multiple perspectives on the activities in the system that is a necessary requirement for socio-technical analysis. It has been argued that the multifaceted nature of the different networks (i.e., social, task and information networks) have revealed the aggregated behaviours that emerge in complex sociotechnical systems (Stanton et al, 2008). This representation was proposed as an alternative to the reductionistic approaches often used to understand systems, which presented systems in their constituent parts but failed to capture the system as a whole. Specifically, three networks are included as illustrated in Figure 1.

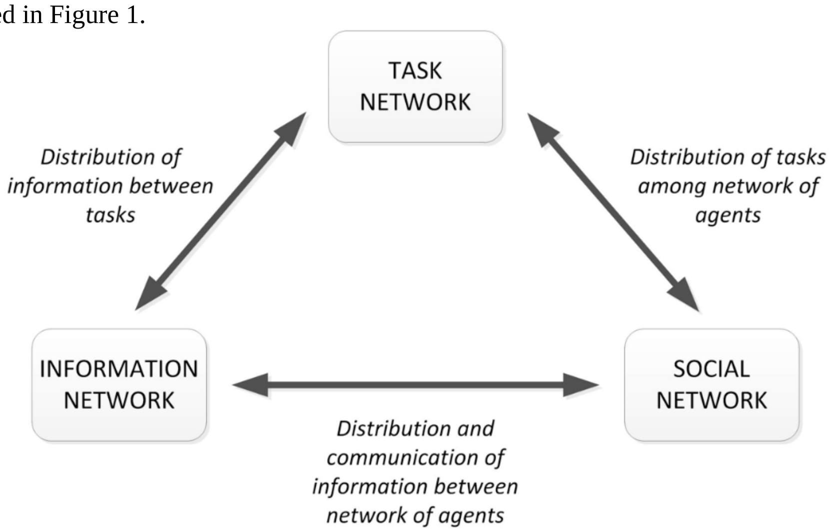


Figure 1. The network of networks approach in EAST

Since its development the method has been employed in a number of areas, including aviation accident analysis (Griffin et al., 2010), air traffic control (Walker et al., 2010), naval operations (Stanton et al., 2006; Stanton, 2014) and military command and control (Walker et al., 2006; Stewart et al., 2008). It has been argued that the multifaceted nature of the different networks (i.e., social, task and information networks) have revealed the aggregated behaviours that emerge in complex sociotechnical systems (Stanton et al, 2008). The original version of EAST required input from a number of data sources, including Hierarchal Task Analysis, Critical Decision Method, Coordination Demand Analysis, Communications Usage Diagram and Operation Sequence Diagrams (Stanton et al., 2008; Stanton et al., 2013). Stanton (2014) presented a shortened version of EAST in which networks are developed directly from raw data; the approach used in this paper has been inspired by the shortened version of the EAST methodology. The EAST framework lends itself to in-depth evaluations of complex system performance, examination of specific constructs within complex socio-technical systems (e.g. situation awareness, decision making, teamwork), and also system, training, procedure, and technology design. Whilst not providing direct recommendations, the analyses produced are often highly useful in identifying specific issues limiting performance or highlighting areas where system redesign could be beneficial. Walker et al (2010) suggested that the insights gained by network modelling were superior to the traditional ethnographic narrative that has previously been used to describe distributed cognition because they present graphical models of systems.

METHODS

Data for this study were collected over a week-long observational field trial at an international air cargo operator in the Middle East. Ethical permission to conduct the study was granted by the Research Ethics Committee at the University of Southampton (RGO number 16250). All employees were aware of the nature of the study and signed a consent form prior to the observations. During the study period, six researchers were assigned to key operational areas within the cargo company and simultaneously observed and recorded the associated activities within the areas of: Crew Briefing Room (Dispatcher, First Officer and Captain), Loading (Supervisor, Team leader, Heavy Load Operator and Loaders) Maintenance (Engineers and Bowserman), Flight Deck (Captain, First Officer and Engineer), Air Traffic Control (ATC) Tower (Ground Controller, Air Controller and Assistant) and Air Traffic Management (ATM) (Sector Controller).

Audio and video recordings captured the work being undertaken in each area. Each researcher also captured additional observations using pen and paper. An outbound flight was observed from the time the pilots entered the crew briefing room until it left the airspace. An inbound flight was observed from when the flight re-entered the airspace until it was shut down at the stand. Three outbound flights and two inbound flights were observed. Each flight resulted in approximately ten hours of recordings, which were transcribed, and supplemented with the additional observations, into excel files. The excel data transcripts were used to populate the three EAST networks for each phase of flight (i.e., (i) crew briefing, (ii) pre-taxi, taxi and take-off, (iii) descent, landing and (iv) taxi, park and shutdown), Task networks were generated from specific actions in the transcripts and key phases of work. These were arranged in chronological order to create a task network with colour coded nodes to represent the tasks and responsible actors. Social networks involved summing the frequency of from/to communications between actors and agents, i.e. noting which actors and agents the exchange went to and from. The social network details direction of information flow and strength of relationships between agents is represented by arrow thickness. Information networks were produced from key words within the transcripts to inform the nodes. These were typically were nouns as they represented the information content being transmitted by actors and agents. Key words that appeared together in the transcripts were linked to produce the network.

The network analyses are further enhanced through the application of network analysis metrics via the AgnaTM software (version 2.1.1). AgnaTM is a social network analysis tool but is becoming an increasingly popular method for general network analysis as a way to gain deeper, quantitative, insights on qualitatively derived networks (Houghton et al, 2006; Baber et al, 2013; Plant and Stanton, 2016). Stanton (2014) proposed that metrics should be selected based on the evaluation being performed and that not all metrics are relevant to all research questions. For this work, the metrics of density, cohesion, and diameter were calculated for the whole network and for individual nodes the metric of sociometric status was calculated.

The complexities of the data collection process are depicted in the ‘rich picture’ in Figure 2. The initials denote the member of the research team and the role they were observing.

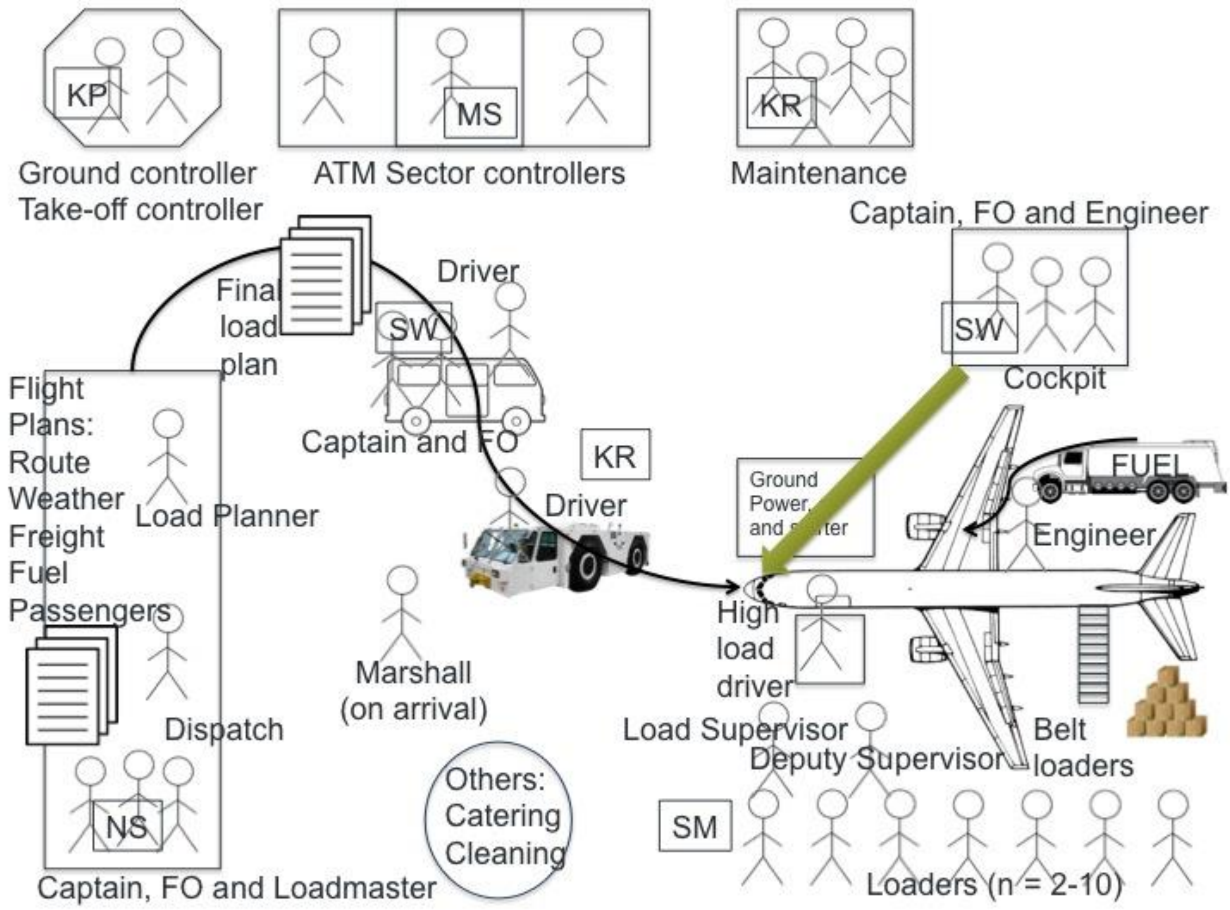


Figure 2. Rich picture of data collection from multiple perspectives in flight operations.

RESULTS

Space limitations prevent all of the networks for all of the phases of flight being presented within this paper (although all will be presented at the conference) so just one phase is focused on, that of Taxi and Take-Off Phase. The task network for the taxi and take-off phase is shown in Figure 3. The flight deck tasks are linear, following a standardised pre-take off procedure, where one task cannot begin until the completion of the previous one. The flight deck tasks are also dependent on ATC tasks, for example the aircraft cannot be taxied until ATC have provided permission to do so. Once the aircraft has taken off ATC pass the flight to radar control in the Air Traffic Management (ATM) sector. ATM provided the flight deck crew with instructions (e.g. height, heading, flight level, route). Data collection ended once the aircraft had left the controlled airspace of the country the study was conducted in.

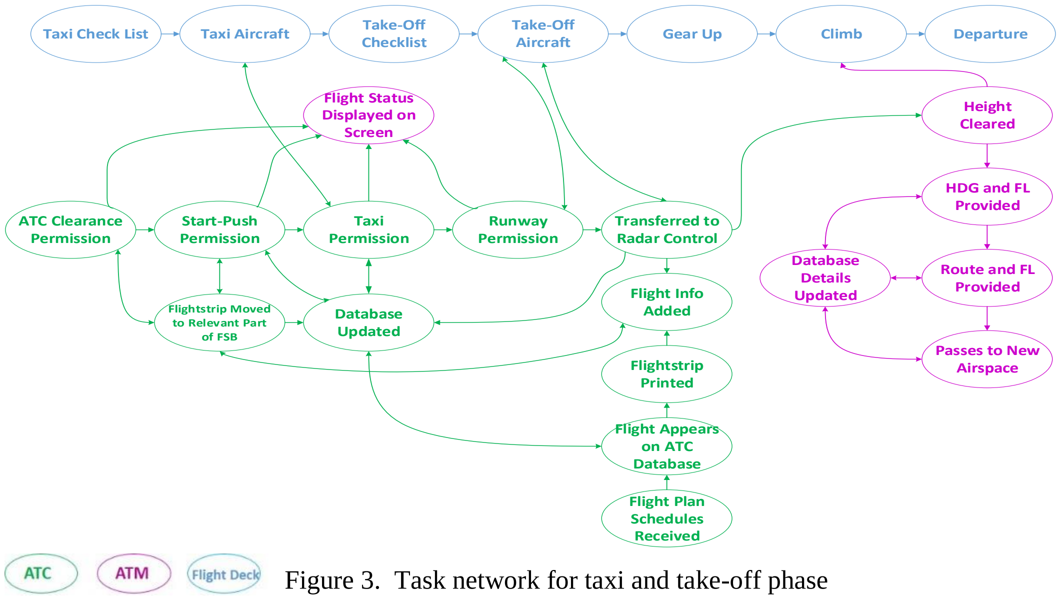


Figure 3. Task network for taxi and take-off phase

The social network for the taxi phase is shown in Figure 4, where the frequency of communications determine the strength of the connection between agents, i.e. darker lines represent stronger links based on high frequencies (and vice versa). This phase consisted of 45 human and technical agents but for clarity the top 30 agents are depicted. Social network metrics were computed from the raw data. The network density value is 0.03, indicative of a relatively unconnected network, although this network is slightly more connected than the pre-taxi network (0.01). Network cohesion is 0.007, suggesting few reciprocal links within the network. Network diameter is 6 (i.e. 6 'hops' from one side of the network to the other). In relation to the individual nodes, sociometric status was calculated to determine the relative importance of each node. The Captain and First Officer are the most important agents in this network both being defined as primary concepts. The other key concepts were ATC agents and the ATC Tower in general. This is unsurprising given that the task network demonstrated how this phase is dominated by flight deck and ATC tasks. Clearly, this has implications for a distributed crewing environment where it will need to be decided who has primary communication responsibility with air traffic services. Currently, the pilot monitoring communicates with ATC (although this may or may not be appropriate when this pilot is remotely located in a ground station). Similarly, the social network diagram depicts how many reciprocal links there are between the Captain and First Officer, thus the challenge for the distributed crewing environment is to maintain this level of communication across a distributed situation.

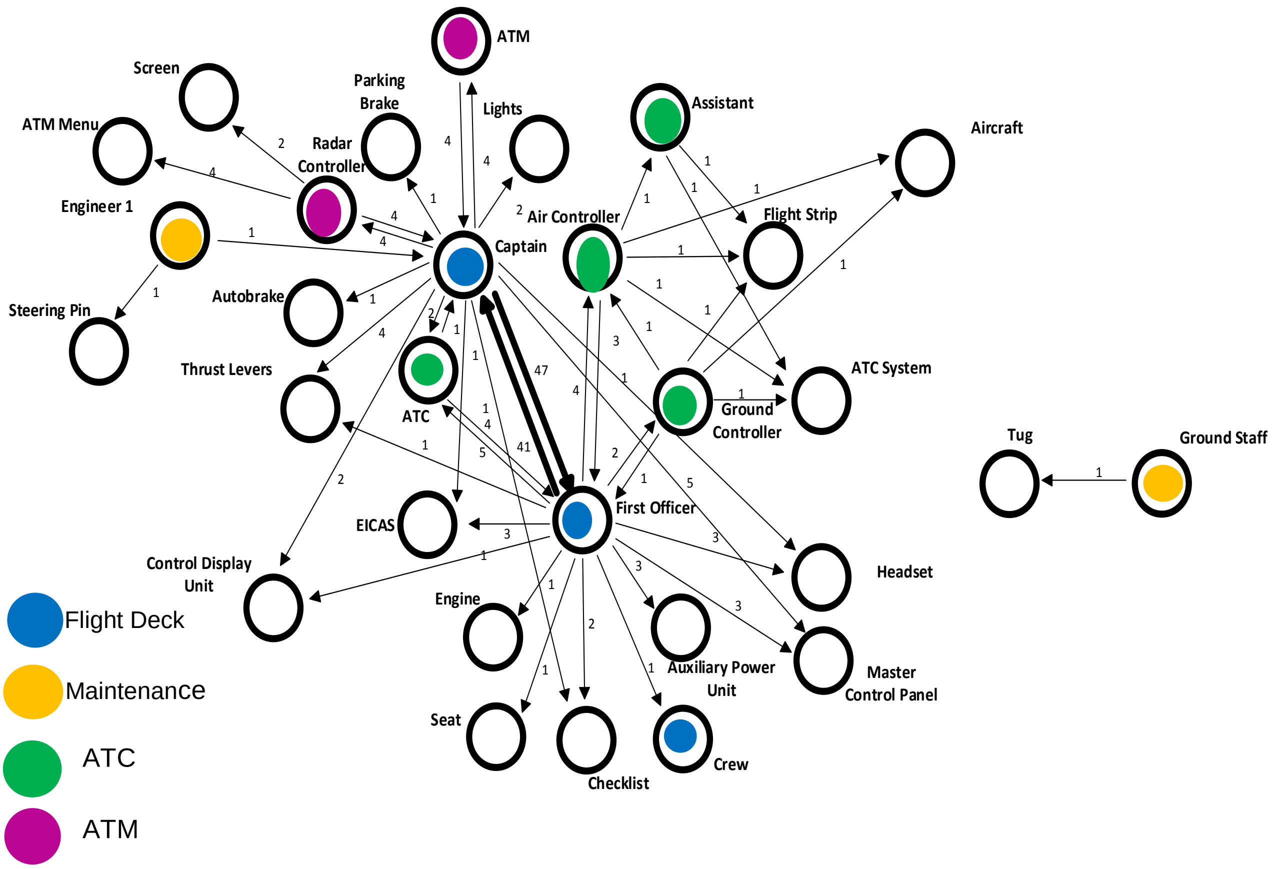


Figure 4. Social network for taxi and take-off phase

As part of this phase the emergency brief is conducted between the Captain and First Officer (the full information network would not fit into the space of this conference paper – but will be presented). The network for the emergency brief is shown in Figure 5. This is where the pilots discuss what they will do in an emergency and the network highlights the information elements the pilots work through to complete brief. For example, the 'problems fire/engine failure' node is connected to 'below 80 knots' which is connected to 'stop', i.e. if the fire occurs below 80 knots the crew will stop the aircraft. The role of the emergency brief is to provide the crew with an opportunity to cross-check safety critical information. It is also a planning activity to ensure that the initial checks are actually made in order to complete the brief. Figure 5 presents one example of a flight crew brief but many other briefs make up the pre-flight preparation phase. Careful consideration will have to be given to the nature of the crew brief in a distributed crewing environment.

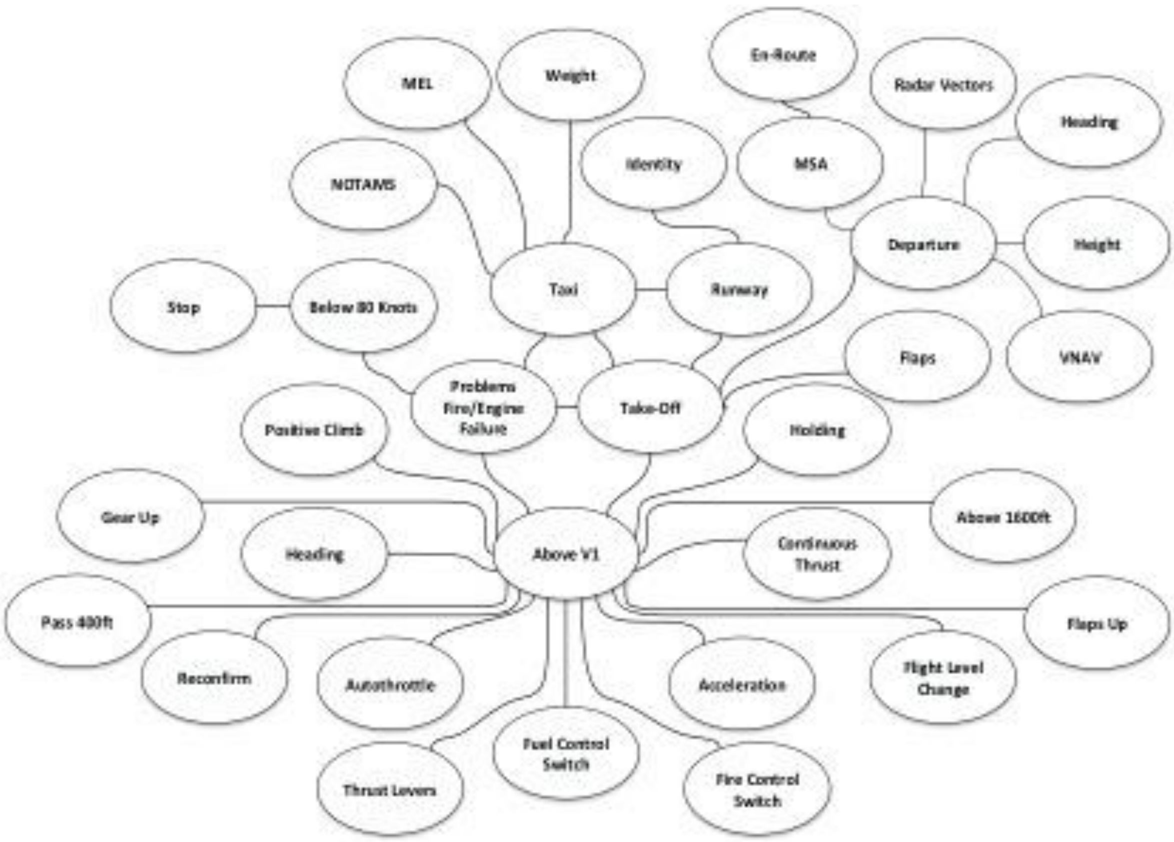


Figure 5. Information network for the emergency brief.

CONCLUSIONS

The analysis of current operations using the EAST method to produce task, social, and information networks. The context of this analysis was in short-haul cargo operations as this is likely to be the initial context of distributed crewing environments (Harris et al, 2015; Stanton et al, 2016). This analysis supports an alternative crewing model because it serves as an 'information audit' of the interactions and connections within the current system. It is not possible to consider the differences within a future operating environment without having first understood the current situation. In addition to generating the three networks for each phase of flight, key network metrics were calculated to quantify some of the qualitative insights gained from the networks. In this analysis of current operations, the individual node metric of sociometric status has been of most relevance to the discussions. However, it is anticipated that the whole network metrics will be used as a comparison between the networks created for the future operations analysis.

REFERENCES

Baber, C., Stanton, N.A.. Atkinson, J., McMaster, R. & Houghton, R.J. (2013) Using Social Network Analysis and Agent-Based Modelling to Explore Information Flow Using Common Operational Pictures for Maritime Search and Rescue Operations. Ergonomics, 56, 889–905.

Griffin, T.G.C., Young, M.S., Stanton, N.A., (2010) Investigating accident causation through information network modelling. Ergonomics, 53(2), 198–210.

Harris, D. & Stanton, N.A. (2010) Editorial: Aviation as a system of systems: Preface to the special issue of human factors in aviation. Ergonomics, 53(2), 145-148.

Harris, D., Stanton, N.A., & Starr, A (2015). Spot the difference: Operational event sequence diagrams as a formal method for work allocation in the development of single pilot operations for commercial aircraft. Ergonomics 58, 11, 1773-1791.

Houghton, R.T., Baber, C., McMaster, R., Stanton, N.A., Salmon, P.M., Stewart, R. & Walker, G. (2006) Command and control in emergency services operations: a social network analysis. Ergonomics, 49(12-13), 1204-1225.

Hutchins, E. (2005) Cognition in the Wild. MIT Press, Cambridge, Mass.

Plant, K.L. & Stanton, N.A. (2016) Distributed cognition in Search and Rescue: Loosely coupled tasks in tightly coupled roles. Ergonomics, DOI: 10.1080/00140139.2016.1143531

Revell, K., Allison, C., Stanton, N.A. & Sears, R. (2016) Modelling Distributed Crewing with STAMP. In Proceedings of HCI-Aero, 14-16th September 2016, Paris, France.

Stanton, N.A. (2014) Representing distributed cognition in complex systems: how a submarine returns to periscope depth. Ergonomics, 57(3), 403-148.

Stanton, N.A., Baber, C. & Harris, D. (2008) Modelling Command and Control. Aldershot: Ashgate. Stanton, N.A. Harris, D. & Starr, A. 2016. The Future Flight Deck: Modelling dual, single, and distributed crewing options. Applied Ergonomics, 53, 331-342.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C. and Jenkins, D. (2013) Human Factors Methods: A Practical Guide for Engineering and Design (second edition). Ashgate: Aldershot.

Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M., Linsell, M., Dymott, R. & Green, D. (2006) Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, 49(12), 1288-1311.

Stewart, R., Stanton, N.A., Harris, D., Baber, C., Salmon, P.M., Mock, M., Tatlock, K., Wells, L. & Kay, A. (2008) Distributed Situation awareness in an airborne warning and control system: application of novel ergonomics methodology. Cognition, Technology and Work, 10(3), 221-229.

Walker, G.H., Gibson, H., Stanton, N.A., Baber, C., Salmon, P. & Green, D. (2006) Event Analysis of Systemic Teamwork (EAST): a novel integration of ergonomics methods to analyse C4i activity. Ergonomics, 49(12-13), 1345-1369.

Walker, G.H., Stanton, N.A., Baber, C., Wells, L., Gibson, H., Salmon, P.M. & Jenkins, D. (2010) From ethnography to the EAST method: A tractable approach for representing distributed cognition in Air Traffic Control. Ergonomics, 53 (2), 184-197

Distributed Situation Awareness: Naturalistic decision making in complex sociotechnical systems

Neville A STANTON
Human Factors Engineering, Transportation Research Group, Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton, SO16 7QF, UK.
n.stanton@soton.ac.uk
+44 (0) 1794 518856

ABSTRACT

This paper presents contemporary thinking on Distributed Situation Awareness (DSA). This has developed over the past decade from a concept into a testable theory with associated methodology. Early forays into understanding the nature of DSA are presented together with examples of case applications. The tenets from the original research paper are presented, which have remained robust over the past decade. DSA is based on the original ideas from Distributed Cognition, which have been extended to show how systems might have awareness. The unit of analysis for DSA has been as declared as the whole Sociotechnical System.

KEYWORDS

Theory and modelling, Cognitive systems, Command and control, Transportation.

INTRODUCTION TO DSA

The argument put forward in this paper favours a sociotechnical systems theoretic approach to Situation Awareness (SA). This approach takes both human and technical agents as well as the way in which they interact into consideration. Sociotechnical systems theory also offers the key to mediation between the different positions taken on Situation Awareness. The systems theoretic approach is potentially useful in addressing the interaction between sub-components in systems. Systems theory proposes a hierarchical order of system components, i.e. all structures and functions are ordered by their relation to other structures and functions and any particular object or event comprises lesser objects and events. Thus, when examining a system, the level and boundaries need to be declared. The resolution that is proposed is that viewing SA 'in-mind', 'in-world' or 'in-interaction' is a declaration of the boundaries that are applied to the analysis. This is not to say that one position is necessarily right or wrong, rather those boundaries are declared openly in the analysis.

HOW DSA WORKS

In his seminal papers on distributed cognition, which have served as inspiration for Distributed Situation Awareness (DSA), Hutchins (1995a,b) described how sociotechnical systems work in practice. He proposed that sociotechnical systems have cognitive properties and these are not reducible to the properties of individuals. By way of an example, Hutchins chose to examine an aircraft cockpit, focusing on the division of work between the 'agents' in the cockpit on approach for landing. The term 'agents' has been chosen to represent both the aircrew and the cognitive artefacts. The landing tasks present an interesting case study because the speed of the aircraft and the flaps and slats in the wing require precise adjustments at set points on the descent. The changes in speed and the flaps/slats need to be undertaken in concert, in order to avoid undue stress being placed on the wings. These settings cannot simply be memorised by the aircrew, as they are highly dependent on the weight of the aircraft. In the example presented by Hutchins (1995a), four different speeds are required by different points on the approach and descent: starting at 245knots the airspeed has be reduced to 227knots, then 177knots, then 152knots and finally to 128 knots. Each reduction in speed is accompanied by a change in the wing's configuration, either by moving the flaps and/or slats. To assist in the task the pilot relies heavily upon external representation of the speed settings. Devices called 'speed bugs' (black pointers that can be moved around the airspeed indicator dial – each with its own flap and slat setting name) are set by the pilot before the approach and descent. The pilot gets the speed settings from a speed card in the aircraft speed card booklet after working out the weight of the aircraft. Then the speed bugs can be set ready for the approach, one bug assigned to each of the four speed settings. Clearly, the pilots are no longer required to remember the speed settings of the aircraft.

Depending on the stage of flight, different 'agents' in a sociotechnical system will have different awareness of a system. These agents are likely to comprise the artefacts in the cockpit (such as the fuel quantity indicator, speed cards, airspeed indicator, altimeter), the pilot flying and the pilot not flying, the air traffic controller, the radar and flight strips in air traffic control. Taking the entire sociotechnical system in the cockpit as the unit of analysis during the descent tasks, DSA focuses at the transactions between the pilots and the artefacts to understand how the aircraft undertakes the descent tasks and what each of the 'agents' is aware of at any given point in time. This approach would show that the pilots hold information about changes in flaps and slats settings with a given point on approach and descent, whereas the speed bugs hold information about the required speed associated with that flaps and slats setting. It is only when the two sub-systems interact (the social sub-system in terms of the pilots and the technical sub-system in terms of the air speed indicator, speed bugs flaps and slats controls), that one can begin to understand how DSA is maintained in the cockpit. Hutchins (1995a) points out that the cognitive processes are distributed amongst the agents in the system, some are human and others are not. The difference between this view and Endsley and colleagues is that the DSA view holds that the sociotechnical system is the unit of analysis, whereas the three-level SA view holds that the individual mind is the unit of analysis. DSA is concerned with the transactions between agents and the physical structure of the environment in sociotechnical systems.

To further understand how this system might work, imagine a network where nodes are activated and deactivated as time passes in response to changes in the task, environment, and interactions (both social and technological). In regard to the system as a whole, it does not matter if humans or technology own this information, just that the right information is activated and passed to the right agent at the right time. This idea is founded on the theory of "transactional memory," which involves the reliance that people have on other people (Wegner, 1986) and machines (Sparrow, Liu, & Wegner, 2011) to remember for them. It does not matter if the individual human agents do not know everything (indeed, it would be impossible for them to), provided that the system has the information, which enables the system to perform effectively (Hutchins, 1995b). We know that agents are able to compensate for each other, enabling the system to maintain safe operation. This dynamism is impossible to model using reductionist, linear approaches. The systems thinking paradigm provides the necessary theoretical foundations and tools to explore the nonlinearity experienced in complex sociotechnical systems (Walker et al., 2010).

DSA THEORY

The fundamental ideas of SA distributed in a system lead to a set of tenets that form the basis of the theory (Stanton et al., 2006). These propositions are as follows:

SA is held by human and non-human agents. Technological artefacts (as well as human operators) have some level of situation awareness (at least in the sense that they are holders of contextually relevant information). This is particularly true as technologies are able to sense their environment and become more animate.

Different agents have different views on the same scene. This draws on schema theory, suggesting that the role of past experience, memory, training and perspective. Animate technologies may be able to learn about their environment.

Whether or not one agent's SA overlaps with that of another depends on their respective goals. Different agents could actually representing different aspects of SA.

Communication between agents may be non-verbal behaviour, customs, and practice (but this may pose problems for non-native system users). Technologies may give off non-verbal cues through sounds, signs, symbols and other aspects relating to their state.

SA holds loosely coupled systems together. It is argued that without this coupling the systems performance may collapse.

One agent may compensate for degradation in SA in another agent. This represents an aspect of the emergent behaviour associated with complex systems.

In the original paper specifying the DSA theory and approach, Stanton et al. (2006) indicate how the system can be viewed as a whole, by consideration of the information held by the artifacts and people and the way in which they interact. The dynamic nature of SA phenomena means they change moment by moment, in light of changes in the task, environment, and interactions (both social and techno- logical). These changes need to be tracked in real time if the phenomena are to be understood (Patrick, James, Ahmed, & Halliday, 2006). DSA is considered to be activated knowledge for a specific task within a system at a specific time by specific agents, that is, the human and nonhuman actors in a system. Although this perspective can be challenging when viewed through a cognitive psychology lens, from a systems perspective it is not (Hollnagel, 1993; Wilson, 2012). Thus, one could imagine a network of information elements, linked by salience, being activated by a task and belonging to an agent—the "hive mind" of the system, if you will (Seeley et al., 2012). For a more complete explanation of DSA theory and measurement, the interested reader is referred to the book by Salmon, Stanton, Walker, and Jenkins (2009).

DSA APPLICATIONS

It The theory has led DSA research into many new domains, including road design (Walker, Stanton, & Chowdhury, 2013), evaluation of road systems and road user behavior (Salmon, Lenne, Walker, Stanton, & Filtness, 2014; Salmon, Stanton, & Young, 2012), advanced driver training (Walker, Stanton, Kazi, Salmon, & Jenkins, 2009), aviation accident investigation (Griffin, Young & Stanton, 2010), and submarine control rooms (Stanton, 2014). The DSA approach has been used in many other studies, including the four presented here.

The very first application of DSA was undertaken in at HMS Dryad Type 23 frigate operations control room simulator, by recording all radio exchanges. The anti-air warfare officer (AAWO) was observed during the air threat, and the principal warfare officer (PWO) was observed during the subsurface and surface threats. All forms of communication were recorded, including verbal exchanges not communicated via radio, hand gestures, and written communication (on paper). Within these scenarios there are four main agents: the officer of the watch (OOW), the PWO, the AAWO, and the captain. The OOW is an officer on the ship's bridge who maintains the visual lookout and controls the ship. The OOW can overrule the manoeuvring orders from the operations room if he/she considers them to be inappropriate.. The PWO is responsible for the tactical handling of the ship and the integrated use of its weapons systems and sensors. The PWO takes a tactical command role in multi-threat missions. The AAWO is responsible for the plan of defence in response to an air attack. The captain oversees the operations room. In addition to personnel, the ship has a computer-based command system, which can communicate and control weapons and sensor systems, allowing information to be passed independently of the command system itself. An information network was constructed for the surface, subsurface and air threat tasks. The total situation for the system under analysis was described by 64 information elements. The information network makes no reference to any particular job roles, and technology is only referred to in a general sense (e.g. weapons, satellite, radar, and sonar). While this is a general system-level representation, activation of any of the knowledge objects was been identified with particular tasks from the task. This activation of the information network illustrates the ideas governing DSA of the system in a very literal sense.

In another application, a field trial of a new £2.4 billion mission-planning and battlespace management system was undertaken (Stanton, Walker, et al., 2009). For our purposes, the DSA approach was considered by the research team to be most appropriate methodology that could be applied to assess SA in this complex naturalistic setting because of the dynamic nature of the activities. It would be impossible to "script" the system activities and metrics ahead of time, as the planning and operations teams had to adapt to the changing nature of the environment. Based on live observations, the DSA analysis identified design issues adversely affecting system performance (Salmon, Stanton, Walker, Jenkins, Ladva, et al., 2009). The outputs were used to generate explicit system redesign recommendations (Stanton, Walker, et al., 2009) that have been subsequently implemented. Consequential improvements in system performance were observed.

In the third application, DSA has also been studied in the energy distribution domain (see Salmon et al. 2008) to demonstrate the concepts of compatible and transactive aspects of SA. This study demonstrated how the activities and transactions that occurred within the energy distribution system can be mapped onto the perceptual cycle model. The first transaction to take place is the issue of instructions by the operator. This serves to update each schema of the system and of the work required, which in turn drives the activities that the system then undertakes. The outcome of these activities is then checked by the others in the field and the operator at the control centre (via circuit displays), which in turn modifies both the systems and the field and operator schema of the current status of the system. The study demonstrated how the cyclical perception-action notion can be applied to the entire system as well as the teams and individuals working within it.

A laboratory study has demonstrated way in which media could be designed to keep distributed teams involved in a collaborative task (Walker, Stanton, Salmon, & Jenkins, 2009b). Different media were investigated to support the collaboration. There were four conditions: voice only (a telephone link between participants), voice and video (a live video link between participants), voice and data (an electronic shared workspace), and voice, video, and data (all three media). The finding showed better in the media-rich condition (i.e., voice, video, and data). The explanation lies in that the greater the support from the environment, the less the person has to remember as the artifacts in the system hold the information (similar to the manner in which mobile phones hold contact numbers). In the same way that pilots use the speed bugs to remember for them (Hutchins, 1995a), the participants were using the video and shared electronic workspace to remember. Similar findings are being reported in the wider literature (Sparrow, Liu, & Wegner, 2011). The awareness of the system was distributed across the agents and media. The sociotechnical view of DSA led to a different, and considerably richer, conclusion for system design.

Finally, Salmon, Walker and Stanton (2016) consider a disaster from a systems perspective, in this case the infamous aviation accident of AF447. They argue that although the official report put much of the blame on the pilots of the aircraft, this is at odds with contemporary thinking in sociotehnical systems. From a DSA perspective, the unit of analysis should be the whole sociotechnical system rather than individuals. Salmon et al draw on DSA theory to put forward the idea of the 'hive mind' of the system and propose the investigation should have focused on the 'transactions' between system agents to get a more accurate understanding of events. Moreover, it is the 'system' that loses awareness rather than individuals. Salmon et al identify three main phases in the incident: entrance into the ITCZ until the autopilot disconnects, the aircrew's response to the disconnection and the return of the captain to the cockpit. Their analyses show significant decrements to DSA in all three phases. In the analysis, Salmon et al identify four classes of transaction failure that directly contributed to the accident: absent transaction, inappropriate transactions, incomplete transactions and misunderstood transactions. All incidents are likely to have one or more of these classes of system failures present.

The relationship between SA and task performance has remained resolutely difficult to prove, with some research both proving and falsifying the link, even within the same study (Endsley, 1995), which begs the question, why bother with SA if it is not revealing anything about how teams actually perform on tasks? The systems view of SA is not as equivocal. Research into the conversations teams have when performing tasks and found a very strong positive relationship between DSA and the teams' performance on the task (Sorensen & Stanton, 2013). The research has also shown the same effect in high-fidelity, pre-deployment, training environments (Rafferty et al., 2013). DSA, therefore, does tell us how teams actually perform, making SA as a concept more, rather than less, useful. This is a key insight that has been supported by the research of others.

CONCLUSIONS

Other researchers are beginning to catch on and find the DSA concept, theory and methodology useful in their work, including, but not restricted to: (Bourbousson, Poizat, Saury, & Seve, 2011; Fioratou, Flin, Glavin, & Patey, 2010; Golightly et al., 2013; Golightly, Wilson, Lowe, & Sharples, 2010; Macquet & Stanton, 2014; Patrick & Morgan, 2010; Haavik, 2011; Schulz, Endsley, Kochs, Gelb, & Wagner, 2013). The groundswell of evidence suggests that DSA has much to offer our understanding of some of the complexities of sociotechnical systems, and this is just the beginning.

In science and engineering, there is a natural rise and fall of paradigms as progress is made. In this way, a new paradigm becomes more established until it gives way to new developments. It is contended that DSA presents a new paradigm for analyzing and explaining SA in systems, and there is a groundswell of studies that are tipping the balance of evidence in that direction. The debate is not expected to end here, but readers are encouraged to approach all of the ideas with an open mind, try out the approaches, and decide for themselves.

REFERENCES

Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2011) Description of dynamic shared knowledge: An exploratory study during a competitive team sports interaction. Ergonomics, 54, 120–138.

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems. Human Factors, 37, 32–64.

Fioratou, E., Flin, R., Glavin, R., & Patey, R. (2010). Beyond mon- itoring: Distributed situation awareness in anaesthesia. British Journal Anaesthesia, 105, 83–90.

Golightly, D., Ryan, B., Dadashi, N., Pickup, L., & Wilson, J. R. (2013). Use of scenarios and function analyses to understand the impact of situation awareness on safe and effective work on rail tracks. Safety Science, 56, 52–62.

Golightly, D., Wilson, J. R., Lowe, E., & Sharples, S. (2010). The role of situation awareness for understanding signaling and control in rail operations. Theoretical Issues in Ergonomics Science, 11, 84–98.

Griffin, T. G. C., Young, M. S., & Stanton, N. A. (2010). Investigating accident causation through information network modeling. Ergonomics, 53, 198–210.

Haavik, T. K. (2011). Chasing shared understanding in drilling operations. Cognition, Technology and Work, 13, 281–294.

Hollnagel, E. (1993). Human reliability analysis: Context and control. London, UK: Academic Press.

Hutchins, E. (1995a). How a cockpit remembers its speeds. Cognitive Science, 19, 265–288.

Hutchins, E. (1995b). Cognition in the wild. Cambridge, MA: MIT Press.

Macquet, A., & Stanton, N. A. (2014). Do the coach and athlete have the same "picture" of the situation? Distributed situation awareness in an elite sport context. Applied Ergonomics, 45, 724–733.

Patrick, J., James, N., Ahmed, A., & Halliday, P. J. (2006). Obser- vational assessment of situation awareness, team differences and training implications. Ergonomics, 49, 393–417.

Patrick, J., & Morgan, P. L. (2010). Approaches to understanding, analyzing and developing situation awareness. Theoretical Issues in Ergonomics Science, 11, 41–57.

Rafferty, L. A., Stanton, N. A., & Walker, G. H. (2013). Great expectations: A thematic analysis of situation awareness in fratricide. Safety Science, 56, 63–71.

Salmon, P. M., Lenne, M. G., Walker, G. H., Stanton, N. A., & Filtness, A. (2014). Exploring schema-driven differences in situation awareness across road users: An on-road study of driver, cyclist and motorcyclist situation awareness. Ergonomics, 57, 191–209.

Salmon, P. M., Stanton, N. A. & Walker, G. H. (2016). Pilot error versus sociotechnical systems failure: a distributed situation awareness analysis of Air France 447. Theoretical Issues in Ergonomics Science, 17 (1), 64-79.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Jenkins, D. P. (2009). Distributed situation awareness: Theory, measurement and application to teamwork. Ashgate, UK: Aldershot.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., Baber, C., & McMaster, R. (2008). Representing situation awareness in collaborative systems: A case study in the energy distribution domain. Ergonomics, 51, 367–384.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., Ladva, D., Rafferty, L., & Young, M. S. (2009). Measuring situation awareness in complex systems: Comparison of measures study. International Journal of Industrial Ergonomics, 39, 490–500.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., & Rafferty, L. (2010). Is it really better to share? Distributed situation awareness and its implications for collaborative system design. Theoretical Issues in Ergonomics Science, 11, 58–83.

Schulz, C. M., Endsley, M. R., Kochs, E. F., Gelb, A. W., & Wagner, A. J. (2013). Situation awareness in anaesthesia: Concept and research. Anaesthesiology, 118, 729–742.

Seeley, T. D., Visscher, P. K., Schlegel, T., Hogan, P. M., Franks, N. R., & Marshall, J. A.(2012). Stop signals provide cross inhi- bition in collective decision-making by honeybee swarms. Science, 335, 108–111.

Sorensen, L. J., & Stanton, N. A. (2013). Y is best: How distributed situational awareness is mediated by organisational structure and correlated with task success. Safety Science, 56, 72–79.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science, 333, 776–778.

Stanton, N. A. (2014). Representing distributed cognition in complex systems: How a submarine returns to periscope depth. Ergonomics, 57, 403–418.

Stanton, N. A., Walker, G. H., Jenkins, D. P., Salmon, P. M., Revell, K., & Rafferty, L. (2009). Digitising command and control: Human factors and ergonomics analysis of mission planning and battlespace management. Aldershot, UK: Ashgate.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., . . . Green, D. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. Ergonomics, 49, 1288–1311.

Walker, G. H., Stanton, N. A., & Chowdhury, I. (2013). Situational awareness and self explaining roads. Safety Science, 56, 18–28.

Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., & Jenkins, D. P. (2009). Does advanced driver training improve situation awareness? Applied Ergonomics, 40, 678–687.

Walker, G. H., Stanton, N. A., Salmon, P. M., & Jenkins, D. P. (2009). How can we support the commander's involvement in the planning process? An exploratory study into remote and co-located command planning. International Journal of Industrial Ergonomics, 39, 456–464.

Walker, G. H., Stanton, N. A., Salmon, P. M., Jenkins, D. P., & Rafferty, L. (2010). Translating concepts of complexity to the field of ergonomics. Ergonomics, 53, 1175–1186.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), Theories of group behavior (pp. 185–208). New York, NY: Springer-Verlag.

Wilson, J. R. (2012) Fundamentals of systems ergonomics. Work, 41, 3861–3868.

Operative Team Function – Impact of Procedural Complexity, Number of Handoffs, and Nursing Expertise

Lygia STEWART MD [a,b]

[a]Professor of Surgery, University of California, San Francisco
513 Parnassus Avenue, Room S-321, San Francisco, CA 94143

[b]Chief General Surgery, San Francisco VA Medical Center
Surgery (112), 4150 Clement Street, San Francisco, CA 94121

ABSTRACT

Teamwork is important in the operating room (OR). Efforts to bring the team together (Medical Team Training) have shown promise. But, expertise among OR teams is diverse. Some specialties mandate specific nursing/scrub expertise (Cardiac, Transplant), but this is not universally available. Additionally, team composition frequently changes intra-operatively due to handoffs mandated for lunch/change of shift breaks. Factors potentially influencing team performance include: individual team member expertise, changes in team composition (handoffs), procedural complexity, and complex instrumentation. We examined team performance utilizing Medical Team Training data from 6,136 cases. We examined the influence of nursing expertise overall; and drilled down on General Surgery (1,881 cases) to examine the influence of handoffs, procedural complexity, and complex instrumentation. We found that number of handoffs, procedural complexity, and complex instrumentation (laparoscopic cases) challenged team performance; while the presence of nursing expertise (Section charge nurse, experienced scrub technician) improved team performance even when these factors were present.

KEYWORDS

Surgical teams, Medical Team Training, Expertise, Handoffs, Patient Safety.

INTRODUCTION

The operating room (OR) is a complex high-stakes environment in which effective communication and the coordination of multiple team members is crucial for safe and efficient functioning. An interprofessional team must complete a series of complex, interrelated tasks, and rely on one another's expertise for completing these tasks successfully. They must share information rapidly when responding to expected and unexpected events.
Studies of high-functioning teams have described a degree of shared awareness and coordination termed "team cognition," or a "shared mental model" that allows them to work together efficiently. Compared to ad hoc informal teams, permanent teams are better able to use implicit, unspoken communication in addition to explicit communication, since they are familiar with each other's roles, preferences, and the details of their procedures. A number of studies have found that intact teams—teams that worked together consistently—perform better than ad hoc teams when carrying out difficult procedures (De Leval et. al., 2000). However, permanent teams are not the rule in today's operating rooms, and achieving a seamless execution of coordinated behaviors can be challenging.

Additionally, teams are continually changing during the operative procedure due to mandated lunch/change of shift breaks (handoffs). We previously examined the impact of OR handoffs on teamwork, stress, and work among the members of the operative team (surgeons, anesthesia providers, circulator nurses, scrub technicians) using a 360 degree evaluation (Anderson & Stewart, 2015). We found that surgical attendings reported decreased teamwork, increased stress, and increased work due to handoffs in about 30-50% of cases (Anderson & Stewart, 2015). Additionally, there has been a significant increase in the complexity of instrumentation due to innovations in minimally invasive and endovascular procedures. Thus potential obstacles to coordinated teamwork are team member turnover (handoffs), inexperience, complex procedures, complex instrumentation requirements, and lack of familiarity with a procedure or other team members.

Efforts to maintain team continuity and to build expertise in the operating room may include formation of semi-permanent or partially permanent teams and utilization of Medical Team Training programs (Awad et al., 2005, Dunn et al, 2007; Haynes et al., 2009, Halverson et al,. 2009; Lingard et al. 2008, Makary et al, 2006, Neily et al, 2010, Pizzi 2001; Salas et al., 2008, Salas et al., 2009, Weiser et al., 2010, Young-Xu et al, 2011). We implemented a robust OR Medical Team Training program at the SF VA Medical Center (since 2006). We have reported on this program's effectiveness in improving team performance and decreasing delays in the OR (Wolf, Way & Stewart, 2010). At our medical center, each of our eleven surgical specialties has its own dedicated specialty nurse, the section charge nurse, who is permanently assigned to a particular surgical specialty. The section charge nurse manages the specialty's needs, including equipment procurement, training new staff, maintaining surgeons' preference lists, overseeing room set-up, and facilitating overall smooth functioning of the operating room; they also act as the circulating nurse for that specialty whenever possible. In addition, certain scrub technicians/nurses (SuperScrub), have more experience with a specialty, anticipate the surgeon's needs during the case, and greatly facilitate the case.

We sought to understand the obstacles to team performance, opportunities for improvement, and whether participation of a dedicated specialty charge nurse or SuperScrub had an impact on OR functioning. We examined potential factors including: number of handoffs, procedural complexity, and complex instrumentation. We hypothesized that handoffs, procedural complexity, and complex instrumentation would challenge team performance; while the presence of the nursing expertise (section charge nurse, SuperScrub) in the operating room would result in improved team coordination.

METHODS

We instituted a Medical Team Training program at the San Francisco VA Medical Center starting in September 2006. The results of our endeavour over the first two years, and our pre-op briefing and post-op debriefing form has been published (Wolf, Way & Stewart, 2010). The post-op debriefing included key elements including: identification of case delays, issues requiring follow-up, recognition of good teamwork, and assigning an overall case score to the team's performance. The 5-point case score, had each point specifically defined, to facilitate accurate data capture. Data from the briefing/debriefing forms were entered into a secure database for quality improvement review. We analysed Medical Team Training data obtained, for all surgical specialties, from 6,136 operations, over a 4 year period (Sept 2006 – Sept 2010).

Over the entire cohort we examined how the presence of section charge nurse during the procedure influenced the case. We specifically examined the impact on overall case delays, and delays attributable to various team members (Anesthesia, Surgery, Nursing, patient, etc).

We further examined 1,881 General Surgery cases (Sept 2006 – July 2011) and examined the influence of procedural complexity, use of minimally invasive (laparoscopic) approach, number of handoffs, presence of a superior expert scrub technician (SuperScrub), and/or the section charge nurse, had on team performance. Surgical procedures were categorized (3 groups) based on procedural complexity (Low, Moderate, Complex)

We also performed a multivariate analysis of the factors independently influencing team performance (as defined by case score). We used SPSS version 21 for the analysis. We used the following factors in our analysis: procedural complexity, laparoscopic approach, number of handoffs, presence the section charge nurse, and presence of the SuperScrub.

RESULTS

All Surgical Specialties (6,136 cases)

The section charge nurse was present in 4,177 cases (68%). Delay rates (cases with any delay) were 7.7% when the charge nurse was present vs. 20.9% when the charge nurse was not present ($P=0.0001$, χ^2). The average number of delays was lower with the charge nurse present (0.09 vs. 0.13, $P<0.0001$, ANOVA). Among types of delays, anesthesia-related, surgeon-related, and scheduling-related delays decreased significantly (Table 1), while nursing, medication, testing, patient, equipment and other delays were unchanged. The presence of the charge nurse was associated with more frequent acknowledgement of good teamwork (8.3% vs. 6.5%, $P=0.006$)

Table 1. Presence of Section Charge Nurse and Decreased Delays

Type of Delay	Section Charge Nurse N = 4177, N (%)	Other Nurse N=1959, N (%)	P Value
Any Delay	320 (7.7%)	411 (20.9%)	0.0001
Anesthesia Delay	53 (1.3%)	49 (2.6%)	0.001
Surgery Delay	60 (1.5%)	49 (2.6%)	0.007
Scheduling Delay	20 (0.5%)	20 (1%)	0.013

General Surgery Analysis (1,881 cases)

Among the 1,881 General Surgery cases, the Section Charge Nurse was present in 967 (52%) cases; a scrub technician, with high General Surgery expertise, was present in 615 (33%) cases. The presence of these team members was associated with more frequent acknowledgement of good teamwork (Charge Nurse vs Other: 20.3% vs 9.7%, P=0.0001, χ^2; Super Scrub vs Other: 24.2% vs 10.8%, P=0.0001, χ^2).

One or more handoffs were done in 1493 (79%) cases, and they were associated with lower case scores (4.88 vs 4.7, none vs handoff, P=0.003). The number of handoffs inversely correlated with case score. Average case scores by number of handoffs were: no handoffs 4.89, 1 handoff 4.88, 2 handoffs 4.81, 3 handoffs 4.72, 4 handoffs 4.52, and 5 or more handoffs 4.42 (Pearson correlation P<0.0001)

Procedural complexity also influenced case score and number of handoffs. The number of cases in each area of procedural complexity was: Low 794(42%), Moderate: 696 (37%), Complex: 391 (21%). Increasing procedural complexity was associated with lower case score: Low: 4.93, Moderate 4.8, Complex 4.46, P<0.0001, ANOVA). Cases with a higher procedural complexity were associated with a higher number of handoffs: Low: 1.08, Moderate 1.74, Complex 3.65, P<0.0001, ANOVA). Both procedural complexity and number of handoffs correlated with case score, increasing handoff number and increasing procedural complexity was associated with a lower case score (Figure 1).

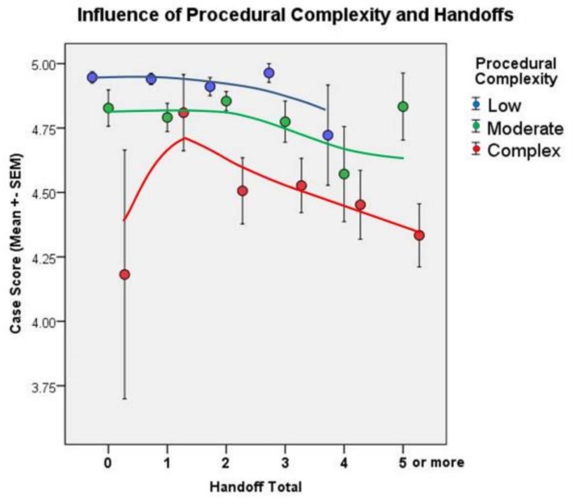


Figure 1

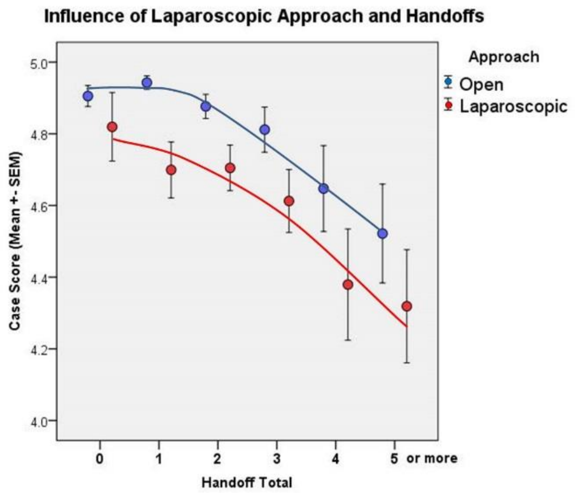


Figure 2

The use of a laparoscopic approach also influenced case score. Laparoscopic cases had a lower case score than open cases: Open (1262 cases): 4.87 vs Laparoscopic (619 cases): 4.63, (P<0.0001 ANOVA). In addition, the use of a laparoscopic approach and number of handoffs correlated with case score; increasing handoffs and a laparoscopic approach was associated with lower case score. (Figure 2).

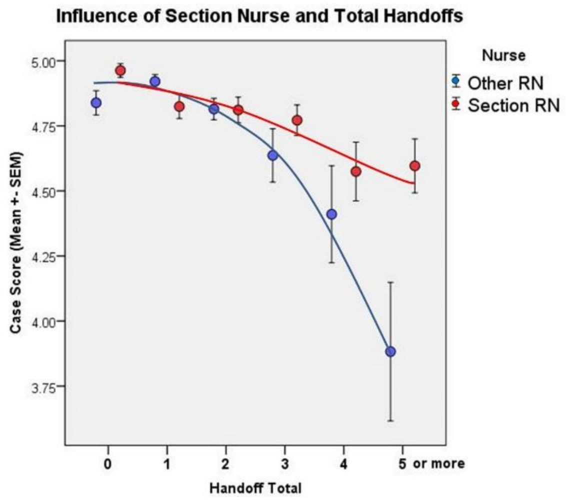


Figure 3

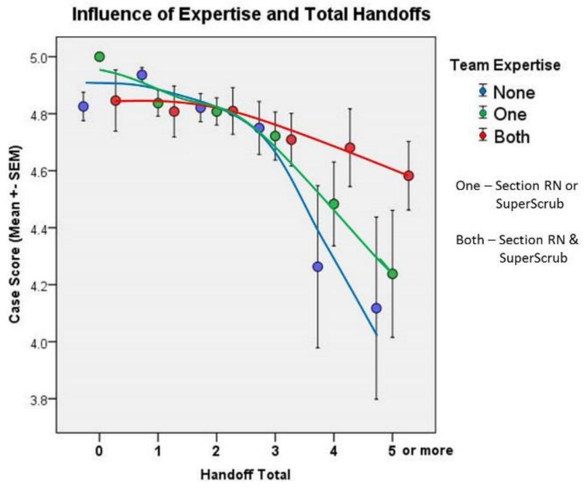


Figure 4

The section charge nurses were present in more complex cases (Section Charge vs Other, 31% vs 9%, P<0.0001, χ^2). Because of this, their cases usually were associated with more handoffs (Section Charge vs Other, 2.2 vs 1.5, P<0.0001, t-test). When case score was graphed as a function of handoffs, the presence of the section charge nurse was associated with improved team performance, and it seemed to counteract the negative influence of handoffs on case score. (Figure 3) Despite increasing number of handoffs, when the section charge nurse was present, team performance was rated better at each level of handoffs. The SuperScrubs were also more commonly placed into complex cases (SuperScrub vs Other, 38% vs 12%, P<0.0001, χ^2) with more handoffs (SuperScrub vs Other, 2.6 vs 1.5, P<0.0001, t-test). The presence of either the Section charge nurse or SuperScrub correlated with higher case score, and this was even higher when both were present (Figure 4).

Multivariate Analysis

On multivariate analysis, the following factors independently inversely correlated with case score (as a measure of team performance): increasing procedural complexity (P=0.0001), laparoscopic approach (P=0.0001), and increasing number of handoffs (P=0.01). The presence of the section charge nurse (P=0.0001) or either the section charge nurse or SuperScrub (P=0.007) independently positively correlated with case score.

DISCUSSION

This study demonstrates the complexity of team performance in the operating room, and the difficulty analysing OR team performance. It highlights the many variables that influence team performance including case complexity, approach (laparoscopic vs open), and number of handoffs. Complex cases and those requiring more specialized equipment and procedures (like laparoscopic cases) demonstrated significantly worse team performance than simple straight-forward cases. This was further compounded by the fact that presence of handoffs also correlated with worse team performance – and these complex and laparoscopic cases had more handoffs than other cases. This underscores the multivariate nature of OR teams and the need to consider this in any analysis of OR team performance.

This data agrees with our previous observations on handoffs. Not only are handoffs associated with increased work, increased stress - which interferes with teamwork (Arora et al, 2010, Driskell 1999, Sexton 2000), but increasing number of handoffs were independently associated with decreased team performance (case score). A number of studies have highlighted possible patient safety issues related to handoffs, focusing on in-hospital handoffs. (Borman, Jones, & Shea, 2012; Lee, Myers, Rehmani, et al., 2012; Kitch, Cooper, Zapol, et al, 2008; Charap 2004). But, these studies did not focus on OR handoffs, which are potentially a more salient issue. One reason is that handoff frequency is actually higher in the operating room than with patient care on hospital wards. Most hospital patient-related handoffs occur at the end of an 8-12 hour physician shift; while in the operating room, cases lasting a few hours can be associated with multiple nursing personal handoffs, especially if they occur near the lunch hour or change of shift (usually 3 pm). We noted that about 80% of General Surgery cases had one or more handoffs, and as case complexity increased, so did the number of handoffs. This means that surgeons ended up working with more fragmented teams in the most complex of cases. Not only can this increase the possibility for errors related to a handoff, but it disrupts the rhythm of the case, and increases the work of an already complex high-work case.

What was interesting was that the presence of a dedicated specialty cieculating nurse, the Section Charge nurse, seemed to mitigate this problem. This was the case even when the Section Charge nurse handed off the case to another circulating nurse at a later point in the case. The presence of a dedicated specialty nurse in the OR was independently associated with decreased case delays, increased documentation of good teamwork, and better team performance even in the face of increased handoffs. These findings support theories of team functioning and the importance of experience for enhancing team cognition and performance. Similarly, presence of an expert scrub technician, with high familiarity for the case, also improved team performance. And, in cases where the Section Charge nurse was not present, this person facilitated the case. The best outcomes were when both of these expert team members were present, underscoring the importance of expertise in team functioning.

The framework of implicit versus explicit communication may be illuminating when considering why these experts make a difference. In the operating room, teams rely on both explicit communication—in which everything that is being "said" is spoken aloud—and implicit communication, in which content is conveyed without requiring that all the words be actually spoken. For example, when a surgeon tells an experienced OR nurse, who is familiar with that surgeon, that they are doing a "laparoscopic Nissen", the nurse will know how to set up the room, the OR table needed, the instruments needed, the usual sequence of events, and, importantly, preparation for less common events that can occur during the case. In contrast, even an experienced nurse, who has limited experience with this surgeon, hearing the same words will not have the same information communicated to him/her. Similarly, the expert scrub technician knows the case and instrumentation, they work directly with the surgeon, and anticipate the moves and needs of the surgeon. Anticipatory cuing and sequence predictability are important to team coordination. The experienced nurse sets up not just for the expected case flow, but also for the various contingencies, so that even deviations from the usual sequence are prepared for, and do not need to be communicated at each step. The experience scrub technician similarly anticipates the moves, often without the need for direct communication, and sets up needed instrmentation. So, even in cases with many handoffs, case preparation (by Section nurse and Super scrub) was such that the case flowed more smoothly, even in their abscence.

Teams that work together frequently are better able to make use of implicit communication since they are familiar with each other's roles, needs, and the details of their procedures, while teams with higher turnover rates are known to rely more heavily on explicit communication (ElBardissi et al, 2008, Kenyon et al, 1997, Manser 2009, Muller et al, 2009, Stepaniak et al, 2012, Weaver et al, 2010). OR protocols require dedicated teams for such high risk procedures as Cardiac and Transplant surgery, but this study demonstrates that this approach may be needed for complex General Surgery procedures, including laparoscopic procedures. With the proliferation of unique surgical instrumentation, it is difficult for surgical teams to work in an ad-hoc manner. The time-out or preoperative briefing and ongoing verbal communication between the surgeon and the team can facilitate smooth coordination by providing a review of procedure steps, equipment, testing and other considerations, but it cannot replace expertise. And while CRM and medical team training can bring the team together (Hugh 2002, Karl 2009), it loses it's efficacy when the team completely changes during the case. This is one reason why surgical teams are very different than aircraft teams (which cannot change mid-flight). The use of explicit communication in the operating room like CRM-based team training, communication techniques such as closed loop, assertive communication, and the use of checklists, are useful but they only go so far. While it is important to teach effective explicit communication, these techniques cannot produce a comprehensive description of all the details for every case. This may explain why, even with our medical center's robust team training and briefing protocols, the effect of the section charge nurse's experience and expert scrub technician's expertise were still evident in improving OR functioning.

CONCLUSION

This study demonstrated the complex nature of surgical team performance. In a large-scale analysis of team performance, we found that the number of handoffs, procedural complexity, and complex instrumentation (laparoscopic cases) independently challenged team performance; while the presence of nursing expertise (surgical circulator/charge nurse or expert scrub technician) improved team performance even when these factors were present. This study underscores the value of expertise and the need to develop expert teams in many surgical domains.

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of the original MTT Implementation Team: Lygia Stewart, MD (Team Leader, Chief of General Surgery), Eric Huzko, MD (Chief of Perioperative Services), Debra Byrd, RN (OR nurse manager), David Doll (OR administrator), Bertha Ruiz (MTT/QI facilitator), Loreto D. Calara (Chief of SPD), and Mary Anne Johnson, MD (Associate Chief of Staff for Clinical Programs). We would also like to greatly acknowledge the data entry contributions of Susan Serrano.

REFERENCES

Anderson, C.E., & Stewart, L. (2015), The Influence of Operating Room Handoffs on Teamwork, Stress, and Work: a 360 degree Evaluation of Team Shared Situation Awareness; Proceedings of the 12th International Conference on Naturalistic Decision-Making, McLean, Virginia, June 2015

Arora, S., Sevdalis, N., Nestel, D., & Woloshynowych, M., (2010). The impact of stress on surgical performance: a systematic review of the literature. Surgery, 147, 318–330.

Arora, S., Hull, L., Sevdalis, N., et al, (2010). Factors compromising safety in surgery: stressful events in the operating room. Amer J Surgery, 199, 60–65.

Awad SS, Fagan SP, Bellows C, et al. (2005) Bridging the communication gap in the operating room with medical team training. Am J Surg 190:770-774.

Borman, K.R., Jones, A.T., & Shea, J.A., (2012) Duty Hours, Quality of Care, and Patient Safety: General Surgery Resident Perceptions. Journal of the American College of Surgeons, 215, 70-79.

Charap, M., (2004) Reducing resident work hours: unproven assumptions and unforeseen outcomes. Annals of Internal Medicine. 140, 814–815

De Leval MR, Carthey J, Wright DJ, Farewell VT, Reason JT. (2000) Human Factors and cardiac surgery: a multicenter study, J Thorac Cardiovasc Surg 119 (4 Pt 1):661-72.

Driskell, J.E. & Salas, E., (1999). Does stress lead to a loss of team perspective? Group Dyn, 4, 291–302.

Dunn EJ, Mills PD, Neily J, et al.(2007) Medical team training: applying crew resource management in the Veterans Health Administration. Jt Comm J Qual Patient Saf 33:317-325.

ElBardissi, A.W., Wiegmann, D.A., Henrickson, S., Wadhera, R., & Sundt, T.M., (2008). Identifying methods to improve heart surgery: an operative approach and strategy for implementation on an organizational level. Eur J Cardiothorac Surg, 34, 1027–1033

Haynes AB, Weiser TG, Berry WR, et al, (2009) A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med 360:491-499.

Hugh, T. B., (2002). New strategies to prevent laparoscopic bile duct injury - surgeons can learn from pilots. Surgery, 132, 826-835.

Karl, R.C., (2009). Aviation, J Gastrointest Surg, 13, 6–8.

Kenyon, T.A., Lenker, M.P., Bax, T.W., Swanstrom, L.L., (1997). Cost and benefit of the trained laparoscopic team. A comparative study of a designated nursing team vs a nontrained team. Surg Endosc, 11, 812–814.

Kitch, B.T., Cooper, J.B., Zapol, W.M., Marder, J.E., Karson, A., Hutter, M., Campbell, E.G., (2008), Handoffs causing patient harm: a survey of medical and surgical house staff. Joint Commission Journal on Quality and Patient Safety. 34, 563-570.

Lee, D.Y., Myers E.A., Rehmani S.S., Wexelman BA, Ross, R.E., Belsley, S.S., … Bhora, F.Y., (2012) Surgical residents' perception of the 16-hour work day restriction: concern for negative impact on resident education and patient care. J Am Coll Surg. 215, 868-877

Lingard L, Regehr G, Orser B, et al.(2008) Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. Arch Surg 143:12-7; disc 18

Makary MA, Sexton JB, Freischlag JA, et al. (2006) Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. J Am Coll Surg 202:746-752.

Manser T. (2009) Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. Acta Anaesthesiol Scand 53:143-151.

Muller, S., Zalunardo, M.P., Hubner, M., Clavien, P.A., & Demartines, N.A., (2009). Fixed team program reduces complications and length of hospital stay after open colonic surgery. Gastroenterology, 136, 842–847.

Neily, J., Mills, P.D., Young-Xu, Y., et al, (2010). Association between implementation of a medical team training program and surgical mortality. JAMA, 304, 1693–1700

Salas E, DiazGranados D, Klein C, et al. (2008) Does team training improve team performance? A meta-analysis. Hum Factors 50:903-933.

Salas E, Almeida SA, Salisbury M, et al. (2009) What are the critical success factors for team training in health care? J Comm J Qual Patient Saf 35: 398-405.

Sexton, B., Thomas, E., & Helmreich, R., (2000). Error, stress and teamwork in medicine and aviation: cross sectional surveys. British Medical Journal. 320, 745–749.

Stepaniak, P.S., Heij, C., Buise, M.P., Mannaerts, G.H.H., Smulders, J. F., & Nienhuijs, S. W., (2012). Bariatric Surgery with Operating Room Teams that Stayed Fixed During the Day: A Multicenter Study Analyzing the

Effects on Patient Outcomes, Teamwork and Safety Climate, and Procedure Duration. Anesth Analg, 115, 1384–1392.

Weaver, S.J., Rosen, M.A., Diaz-Granados, D., Lazzara, E.H., Lyons, R., Salas, E., et al, (2010), Does teamwork improve performance in the operating room? A multilevel evaluation. Jt Comm J Qual Patient Saf, 36,133–42

Wolf, F.A., Way, L.W., & Stewart, L, (2010). The Efficacy of Medical Team Training: Improved Team Performance and Decreased Operating Room Delays. A Detailed Analysis of 4,863 Cases. Ann Surg, 252, 477-483.

Young-Xu, Y., Neily, J., Mills, P.D., et al, (2011). Association between implementation of a medical team training program and surgical morbidity. Archives of surgery, 146, 1368-1373.

POSTER PRESENTATIONS

Diagnosis, forecast and sensemaking activities of a National Technical Support Team

Julia ALENGRY[a,b], Pierre FALZON[a], Cecilia DE LA GARZA[b] and Pierre LE BOT[b]
[a] *Research Centre on Work and Development, Conservatoire des Arts et Métiers, Paris, France*
[b] *EDF Lab – Human Factors Group, Palaiseau, France*

ABSTRACT
In crisis situations, urgency and uncertainty make it difficult for people to identify the current situation and risks incurred. Because of the unforeseen nature of the situations, off-site crisis cells, like the National Technical Support Team (NTST), are implemented in order to focus on situation sensemaking and anticipate future states of the facility.

The study's aim is to understand the reasoning of the NTST and how it builds a representation of the situation. A programme of observing crisis management simulations and interviews allowed to identify the diagnosis/forecast/recommendation activity as an iterative process that improves with interactions with the other stakeholders.

KEYWORDS
Crisis management, nuclear industry, expertise, resilience, sensemaking

INTRODUCTION

High-risk organisations are questioning their ability to face extreme and very rare accident situations, over and above those covered by the current design of technical and organisational urgency schemes. Crises are low-probability situations with major consequences (Weick, 1988). They require extensive interpretation and sensemaking efforts. The unique dimension of these situations requires the participants to engage in reasoning at higher levels of abstraction (Crichton & Flin, 2004).

Because of the unforeseen nature of the situations off-site crisis cells are implemented in order to focus on situation sensemaking and anticipate future states of the facility. During crisis management, these crisis cells are responsible for the anticipatory activity by remotely analysing the situation and preparing possible sequences of actions. Managing an incident requires setting new macro-goals, based on a means-ends analysis of the means and resources to be deployed. The crisis unit's work is centred on the identification of levers needed for implementing a new action plan.

Klein (1999) underlines the importance of this forecast activity for the revaluation of the actions to be implemented. The purpose of the forecast is to suggest actions, or even to change the current strategy, and to review the goals so that they match the development of the situation and of the probable future facility states. The anticipatory activity brings about a projection into the future. This activity is part of a different temporality than that of front-line teams. The forecast activity is critical in crisis management, especially in high-risk dynamic processes.

In the nuclear industry, crisis management has a number of specificities. Nuclear production is made by a complex and continous system, that changes slowly: the evolution can last from a couple to tens of hours. The system proceeds under its own impetus and partly escapes the participants' control (Hoc, 1989).

This communication aim is to understand the off-site crisis cells functioning thanks to observation methods during different crisis management simuations.

THE STUDIED SOCIO-TECHNICAL SYSTEM

Within the organisation under study, different operational states of the facility are anticipated.
Because of the severity of the situation, an in-house Emergency Plan is set off, leading to the mobilisation of a multi-level crisis organisation (Figure 22):

- Locally: Local Command Post, Local Crisis Team, Shift Manager.
- Nationwide: National Technical Support Team, National Command Post, AREVA Crisis Team, Institute for Radiological and Nuclear Safety (IRSN) Crisis Team (regularoty) and SEPTEN Crisis Team (another utility's team of experts).

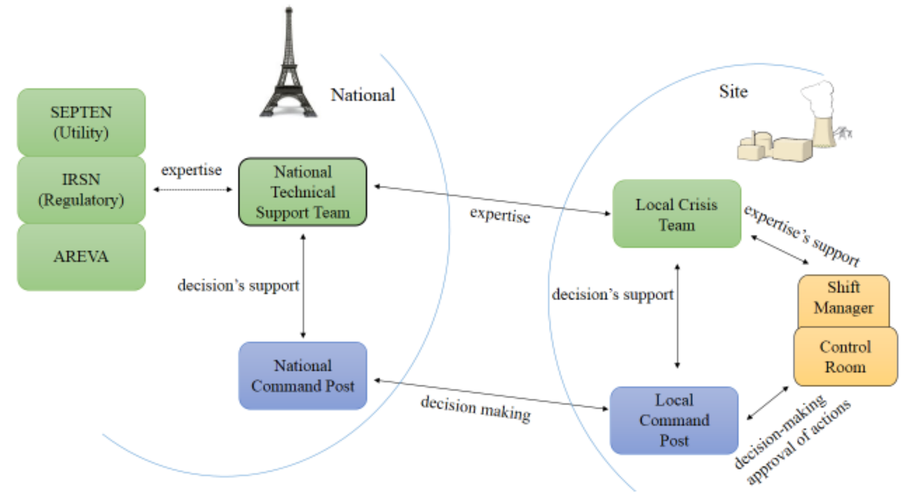


Figure 22: The crisis management organisation under study

This study focuses on the National Technical Support Team (NTST). The NTST comprises twelve general operations and environmental specialists. The NTST is not actually managing the plant. Its purpose is to make its expertise available to the Site by proposing a diagnosis and a forecast of the development of the situation. It also provides information, solutions, and recommendations for action to the Site (through the Local Crisis Team or Shift Manager). The NTST develops a global picture through its technical skills that complement those of the Local Crisis Team. The NTST is an essential resource for the Site to anticipate problems to come and tasks to be undertaken. The recommendations to the Site may concern changes to procedures, the implementation of preventive or mitigating actions. In this crisis organisation, the NTST prognosis has an additional objective : to provide the National Command Post with population protection measures, while confronting their diagnosis/prognosis to the other national expert entities (IRSN, Areva, SEPTEN).

Communications between the Local Crisis Team and the NTST are made through periodic messages received from the Site every 15 mins and describing the status of the facility and through audio-conferences with all expert entities held every 90 minutes approximately.

The entire crisis organisation is tested during realistic simulations, which NTST takes part in. It is during such simulations that we collected our data. The aim of this study is to understand the reasoning of the NTST, and how it builds a representation of the situation.

METHODOLOGY

Six drills were observed and eight post simulation interviews were conducted. The collection of data included four steps :

- Step 1 – Preparing the observation. We gathered information regarding prescripted activities of the NTST and regarding scenarios divised for each simulation.
- Step 2 – *In-situ* observation of six crisis drills. Observations focused on a pair of NTST participants, the Emergency Operations Specialists, who play a central role in team coordination. A table of observations was used to log the time, the team members' associated communications and activity "stages" (NTST's briefing, audio-conference, collection of data, recommendation drafting). Communications between team members were recorded and transcribed. Additionally, their search for information (documentation, data acquisition methods, handwritten notes, written questions ...) was collected.
- Step 3 – Drawing up the data. In order to understand the NTST functioning during the six simulations, time charts were drawn. Theses time charts focused on Emergency Operations Specialists activity. In these time charts, the main activites related to the diagnosis/prognosis missions, their duration and frequency were traced : drafting of diagnosis/prognosis documents - these are shared during audio-conferences with all expert entities, drafting recommendations and all communications established to build the diagnosis/prognosis.
- Focused time charts were then made to highlight recommendation process : from the first draft of a recommendation to the final drafting and the sending - or not, to the Site.
- Step 4 – Post-simulation semi-directed interviews. Eight interviews were conducted based on a table of questions related to the time charts. These time charts – global time charts and focused time charts, served as support during the interviews. The main goal was to make explicit Emergency Operations Specialists activities, thanks to the charts. More precisely, explicitation was asked on : how first diagnosis was set up,

how prognosis was reassessed throughout the simulation and how recommendations were prepared and debated between team members.

This method allowed us to characterise the diagnosis/forecast process, the making of recommendations, and their implications regarding the temporal dimension and the collective management of the situation.

RESULTS

Two phases in the construction of diagnosis/forecast/recommendation

Data analysis based on interviews and time charts yielded general results about the NTST reasoning during crisis simulation. The diagnosis/prognosis NTST activity appears to be split into two phases : 1) a preliminary diagnosis phase – systematically obseved ; 2) a simultaneous diagnostic/prognosis/recommendation construction phase.

As Figure 2 shows, the first phase begins with a voice message from the Site about the emergency situation, and ends when the first audio-conference with all expert entites occurs.

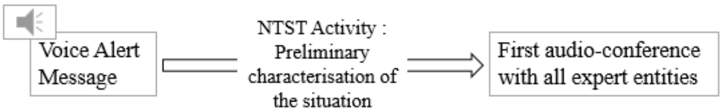


Figure 23 : Preliminary diagnosis phase schema

The voice alert message briefly defines the emergency situation declared in the Site and the facility states. From that voice message, Emergency Operations Specialists made their first hypothesis about the situation and look for clue. Experts mainly look into the procedure applied by the operation team and into the information system to collect data on the safety parameters of the facility. Before the first audio-conference, a briefing is set up inside the NTST by the head of team. According to the team members, these briefings are valuable to share the same representation of the situation and to identify key questions before the audio-conference.

After the first audio-conference, the second phase begins. Explicitation interviews with the Emergency Operations Specialists indicate that:

- Team members assess the situation, using pivotal topics (i.e. the major functions required to ensure the safety of the facility); they identify weak points and possible solutions.
- This analysis of weaknesses and solutions requires both a functional assessment (lost functions, remaining functions, and a forecast of functions that may be recovered depending on the development of the situation and on available resources), and a temporal assessment (the amount of time a function will remain in operation, and a forecast of time remaining before a new failure, or time remaining before an additional worsening leading to an irreversible status change).
- This analysis of weaknesses is a debative activity (Schmidt, 1991). An iterative process of identifying weak points and solutions makes it possible to issue a formal recommendation. Recommendations may concern preventive actions (e.g. managing functions that can be weakened in the medium/long term) or mitigation (e.g. managing a loss of redundance of a safety-critical function).

The purpose of the forecast activity is to bring proposals to front line actors to help them keeping the situation and the facility in a safe and controllable state.

The NTST activity is characterised by a double temporal dimension : it means coping with a constant oscillation between projecting the future state of the facility and updating NTST members' representation of the present situation. This double temporal offset requires phases of (re)synchronising the NTST and the Local Crisis Team that help with the sensemaking process. This double temporal dimension is highlighted throughout the recommendation process.

Recommendation process: a demanding cognitive activity with important synchronisation issues

Observations, time charts and interviews permit to underline that recommendation process is an iterative construction in wich Emergency Operations Specalists lead the team. They ask for specialist's advice, and manage different viewpoints in order to draft and send one clear recommendation to the Site.

In the 6 simulations observed, 21 recommendations were emitted. These recommendations were drafted by Emergency Operations Specialist 2. Recommendations are signed by the Head of team, and sent electronically to the Site. Then the Local Crisis Team relay to the Local Command Post for validation and application (cf. Fig. 3). Interviews and time charts allowed us to focus on each recommendation made by the Emergency Operations Specialists during these simulations. It indicated that the recommendation process is a complex and demanding activity:

- Experts have to assess and reduce uncertainties related to a recommendation. They take into account the recommendations potential effects, regarding the potential process evolution and the actions initiated on Site.

These variables may change the facility state expected. This projection is crucial to reduce incompatibilities or unreliability risks.

- Experts assessment requires oscillation between projection on the future facility state and present-time, to update NTST members' representation of the facility.
- Experts need to get the "right picture" of the facility state and of what is possible to operate on Site. In one case, a mismatch was identified between the actions requested by the NTST and the possibilities on Site.
- Experts try to identify the right moment for carrying them out, if the participants are not already carrying out these actions and if the resources currently on-site are adequate to perform the recommended actions. In the observed simulations, one recommendation was made too early. It did not make sense to the front-line teams and was put on hold; and one recommendation was sent too late, the Site operated by itself.

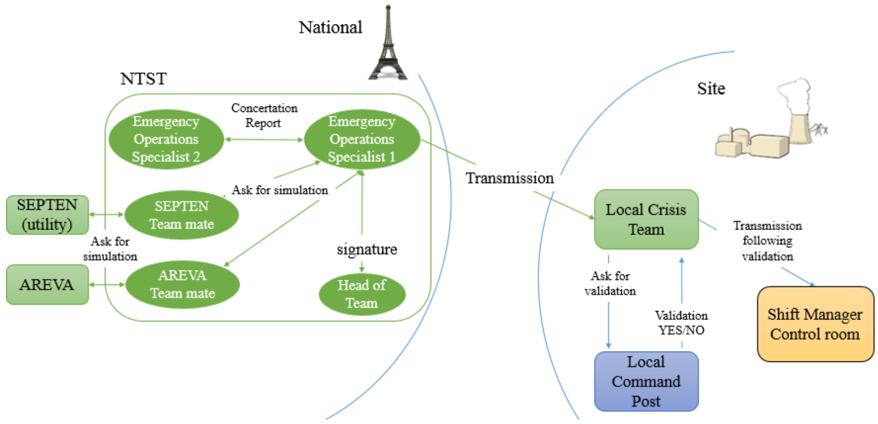


Figure 24 : Recommendation path example

CONCLUSION

NTST's role is to estimate the dynamics of the crisis, the consequences of the fluctuations in the process, and the ways to mitigate potential failures in the operations. This anticipatory activity involves requirements at two levels: at the NTST level for the construction of diagnosis/forecast, and at the collective level to support synchronisation between the NTST and the front-line teams. As regards the construction of the diagnosis/forecast, the NTST must oscillate continuously between future facility state and present state to update their representation of the situation. This oscillation is cognitively demanding but required for the development of a relevant diagnosis/prognosis (i.e. risks incurred, available resources). Concerning the collective level, the study underlined a constant need for synchronisation through the construction and transmission of recommendations. Recommendations have to be formulated at the convenient moment, and be in agreement with both the current state of the facility, and the available resources on Site. Synchronisation is essential to support the collective sensemaking.

Synchronisation seems difficult to maintain, and requires permanent adjustements between the NTST and the Local Crisis Team, achieved through the exchanges (audiconferences and written messages). Desynchronisation may lead to two risks: the recommendation content may not make sense for the Site, which may lead to non-application; the recommendation may be sent too late and be obsolete, or may be sent too early and not be applied. A further study will analyse those recommendations that are successfully implemented and those that are blocked at some point in the transmission chain between the different crisis centre. This will allow us to design training proposals complementary to those already existing and to approach differently these issues of synchronisation and sensemaking.

ACKNOWLEDGMENTS

We would like to thank the members of the NTST crisis team for their time and their involvement in this research.

REFERENCES

Crichton, M.T., & Flin, R. (2004). Identifying and training non-technical skills of nuclear emergency response teams. *Annals of Nuclear Energy, 31*, 1317-1330.

Hoc, J.M. (1989). Cognitive approaches to process control. In G. Tiberghien (Eds), *Advances in cognitive science: Vol. 2: Theory and applications* (pp. 178-202). Chichester: Horwood.

Klein, G. (1999). Chapitre 9 : Nonlinear Aspects of Problem Solving. In Gary Klein (Eds) *Source of Power, How People make Decisions* (pp.134-140). Massachussetts, US : The MIT Press.

Schmidt, K. (1991). Cooperative Work. A Conceptual Framework. In Jens Rasmussen, Berndt Brehmer, and Jacques Leplat (Eds) *Distributed Decision Making: Cognitive Models for Cooperative Work*, Wiley (pp. 75-109). Chichester, UK : Wiley.

Weick, K. (1988). Enacted sensemaking in crisis situations. *Journal of Management Studies, 25(4)*, 305-317.

How Analysts Think: What Triggers Insight?

Matylda GERBER[a], B.L. William WONG[b], Neesha KODAGODA[b]
[a]*Warsaw School of Economics*
[b]*Middlesex University London*

ABSTRACT

In this paper we try to demystify the process of gaining insight. It is important to design decision supporting systems that facilitate insightful decision-making but until we don't know how this process is formulated, it is hard to support it by a system. We propose that it is intuition and leap of faith that triggers the process of gaining insight. Intuition indicates an anchoring point and leap of faith directs further data collection and analysis that ends up in a form of insight. We propose a model that is complementary to triple path model of gaining insight (Klein, Jarosz, 2011). We provide results based on studying decision-making of seven experienced analysts from UK and Belgium. We hope this preliminary study will help to understand what contributes to the formation of insight and will enable the design of a system that supports insightful decision-making while solving criminal cases.

KEYWORDS

Insight, intuition, leap of faith, NDM, decision-making, intelligence analysis

INTRODUCTION

Insight, defined as *"sudden unexpected thoughts that solve problems"* (Hogarth 2001, p.251), is a very unique and important form of finding solution to a problem. Undoubtedly, it should be supported by systems that facilitate decision-making processes, especially in domains where decision-makers need to deal with situations where missing data are the order of the day, like criminal intelligence analysis. However, it is hard to design a support system if we do not know what triggers, precedes and causes its occurrence.

In earlier papers, we presented that police criminal intelligence analysts use intuition, leap of faith and insight to solve criminal cases and that those processes relate to each other (Gerber, Wong, Kodagoda, 2016). Afterwards, we found out that analysts use those processes in situations that are characterized by the lack of clear facts. Based on this finding, we have created a model that shows how decision-making in the absence of clear facts differs from familiar and unfamiliar situations (Gerber, Wong, Kodagoda, 2016).

In this paper, we would like to present how analysts use intuition to trigger the course of gaining insight and the role of intuition and leap of faith in the process of gaining insight with respect to Klein's triple path model of gaining insight (Klein, 2013).

Intuition: an obstacle or a base for insight

Studying insight as a form of problem solving has its roots in Gestalt psychology, which perceives insight as the sudden and unexpected understanding of a problem achieved by reorganization of thoughts (Wertheimer, 1945; Duncker, 1945, Metzger, 1953). Insight is perceived to be always preceded by impasse – self-imposed constraints that need to be relaxed so as to see the right solution (Knoblich et al., 2001). This view of insight matches the discontinuity model of insight, which perceives prior knowledge and experience to inhibit the occurrence of insight (Zander et al. 2016). Early intuitive responses are seen to lead to an impasse by misdirecting the formation of a correct solution. Insight is only possible when the problem-solver relaxes self-imposed constraints arising from intuition (Ohlsson, 1992).

As the discontinuity model treats insight to be the result of a mental restructuring process, the continuity model perceives insight to emerge from a gradual process that is initiated by intuition. From this point of view, the role of intuition in the process of gaining insight seems to be crucial. Bowers (1984, p. 256) perceives intuition as *"a sensitivity and responsiveness to information that is not consciously represented, but which nevertheless guides inquiry towards productive and sometimes profound insights."* According to the continuity model, the process of gaining insight proceeds in two stages. In the first stage, as a result of activation of tacit knowledge connected in semantic memory, a problem solver receives an intuitive feeling that cannot be explained verbally (Collins and Loftus, 1975). In the second stage, when the information becomes consciously available, the problem solver gains insight which explains the prior intuitive hunch (Bowers, 1990). From this perspective, intuitive processing is the non-conscious precursor of insight (Reber et al. 2007).

The source of divergence in opinions about the role of intuition in the process of gaining insight

The first reason for the divergence in opinions arises from the fact that different fields study intuition and insight. Intuition is in an interest of judgment and decision- making (Zander et al. 2016) which includes the fast-and-frugal heuristic approach (Gigerenzer, 2011) and the heuristics-and-biases approach (Kahneman and Tversky, 1974). Placing intuition only in the field of judgment and decision-making reduces intuition to a form of a final judgment, based on which decisions are made. However, intuition works also as an element of a more complex decision-making process (Bowers, 1990). For example, the Recognition-Primed Decision model (Klein, 2008) presents how intuition is utilized to accelerate the decision-making process, but at the same time how its accuracy is verified by a conscious mental simulation.

Insight is mainly investigated by a creativity and problem-solving domain. The main way insight is studied is in the form of laboratory experiments. College students try to solve tasks that are designed in a way to trap problem-solvers with their misleading intuition. A significant disadvantage of this form of study lies in the fact that it eliminates the role of experience in decision-making, which in real life situations plays a crucial role (Jung-Beeman et al., 2004).

In order to see when and how insight appears, we need to provide studies in real life situations. Klein and Jarosz (2011) conducted the first naturistic study to see how insight occurs in real life situations. They found a high significance of experience in the process of gaining insight and a manifestation of other ways of gaining insight than just the one resulting from overcoming impasse. Based on the research, they have defined three pathways that lead to insight: creative desperation, connection and contradiction. The creative desperation path is compatible with previous studies; a problem-solver needs a breakthrough to find a solution. By comparison, the connection path occurs when a person notices a connection among different elements. Finally, contradiction appears when a person faces an inconsistency within information.

METHOD

For this research we have taken the Naturalistic Decision Making (NDM) approach which centers on complex decision-making situations characterized by: complex tasks, complex environments and expert decision-makers (Cannon-Bowers, Salas & Pruitt, 1996; Salas, 2010). We used the Critical Decision Method (Klein, Calderwood et al. 1989), a retrospective interview technique for eliciting participants' knowledge, decision strategies and cues used in their work. We conducted interviews with seven experienced police criminal intelligence analysts from the UK and Belgium. Their average work experience was 12 years. The topics of the interviews were difficult criminal cases the analysts were able to solve independently from start to finish. The interviews were audio recorded and transcribed. We used the Emergent Themes Analysis – a qualitative data analysis approach (Wong, 2004) on transcripts of seven in-depth interviews conducted with experienced police criminal intelligence analysts. In paper (Gerber, Wong, Kodagoda, 2016) we have reported the first part of our analysis, where we have identified occurrence of intuition, leap of faith and insight in analysts' decision-making. The model we present in this paper (see Figure 1.) is a result of our further investigation. Based on our previous analysis, we have identified 33 cases where analysts gained insights that were preceeded by intuition and leap of faith. We have grouped those cases with respect they were initiated by a positive or a negative feeling provided by intuition. Each instance of intuitive judgment was verbally interpreted, what we call leap of faith. Below we present excerpts that reflect intuitions communicated by a positive and a negative feeling and their interpretations in form of leap of faith.

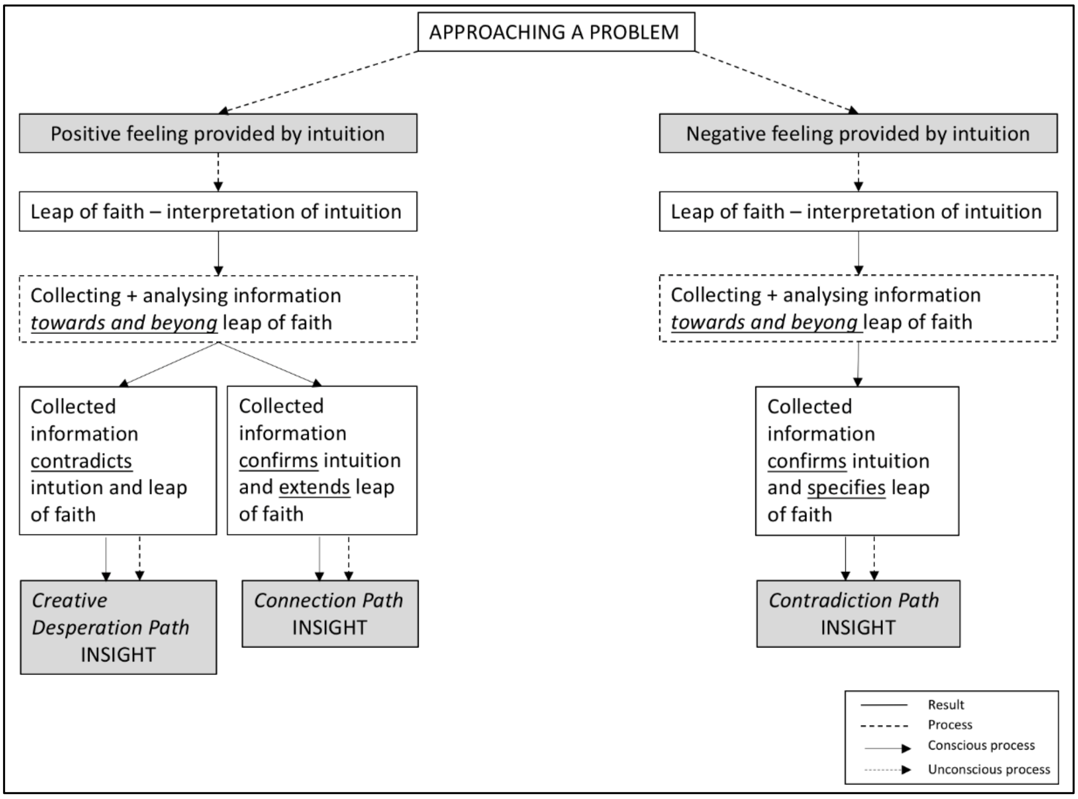
> Positive feeling provided by intuition - `330 "P2: You click the button and you have space in time and` <u>`what did I see is very straight forward`</u>`. Was a gap of an hour."` - leap of faith - `334 "P2: The lack of information is also information. I said he didn't want to be disturbed over there."`
> Negative feeling provided by intuition - `100 "P3: I've just read it and I've read the other one and I say,` <u>`umm… strange`</u>`."`- leap of faith – `142 "P3: He had a weird story, that he was on the roof because there was a dead pigeon. Really."`

At this moment we had 2 groups of insights - preceeded by a positive or a negative feeling provided by intuition. The further analysis allowed us to see that some insights preceeded by a possitive intuitive judgment occurred when collected data seemed to contradict the initial intuition and some were results of collecting information with accordance to intuition. Whereas, all instances of insights preluded by a negative intuitive judgment were results of collecting information consistent with intuition. This analysis allowed us to create a model presented in Figure 1.

RESULTS & DISCUSSION

Based on our analysis and Klein's (2013) triple path model of gaining insight, we have created a model that describes how analysts use intuition and leap of faith to gain insights while solving criminal cases. The model (see Figure 1) describes the process of gaining insights starting from the moment an analyst approaches a problem. Just after an analyst looks at data, he has a positive or negative feeling provided by intuition to focus on a piece information. Positive feeling informs about recognition of a familiar pattern that might be helpful to find a solution to a problem, whereas negative feeling alerts some inconsistency/anomaly that needs to be investigated. Right after intuition comes leap of faith – interpretation of intuition in terms of a possible result of investigating the recognized pattern. The leap of faith allows taking a direction to look for the solution to a problem. As leap of faith is formulated in general terms, an analyst provides analysis not only towards but also beyond leap of faith. At this point, there are three possibilities of gaining insight.

Figure 1. The role of intuition and leap of faith in the process of gaining insight



Creative Desperation Path of Gaining Insight

Creative Desperation Path is initiated when an analyst approaches a problem and receives a positive impulse from his intuition. This feeling motivates him to focus on a piece of information that might be helpful to solve the problem. Right after intuition, the analyst takes a leap of faith that interprets the feeling with respect to a possible solution to which the recognized piece of information may lead. Leap of faith, formulated in general terms, provides a direction where to search for more information. The analyst collects and analyses information towards and beyond leap of faith. If the collected information seems to contradict intuition, the analyst gains insight with a solution where contradictory information becomes consistent with intuition. The insight occurs as a result of both conscious and unconscious analysis.

Connection Path of Gaining Insight

Similar to Creative Desperation, the Connection path is also initiated by a positive feeling provided by intuition. It motivates the analyst to focus on a piece of information that might be helpful to solve a problem. Immediately after intuition the analyst takes leap of faith that provides interpretation of intuition in the form of a general direction where to search for the solution. While collecting and analysing information towards and beyond leap of faith, the analyst gains insight about the right solution. The solution obtained by insight is consistent with the initial intuition but is much more precise than the leap of faith. Both conscious and unconscious analyses are engaged to obtain the insight.

Contradiction Path of Gaining Insight

Contradiction Path begins when an analyst receives a negative impulse from intuition about recognition of some inconsistency/anomaly in the data. Without conscious analysis, he takes a leap of faith that interprets why investigating this inconsistency might be important. The leap of faith provides a general direction for collecting and analysing information. When the gathered information occurs to be consistent with intuition, the analyst gains insight into a solution that clarifies the initial leap of faith. Insight occurs as a result of conscious and unconscious analysis of experience and collected information.

CONCLUSION

It is important to design decision supporting systems that facilitate the process of gaining insight, because this form of finding solution enables successful decision-making in situation of missing data. As results of insights are well analysed, it is still not understood what influences and precedes its occurrence. In this paper we have presented tentative but promising proposition that it is intuition and leap of faith that triggers and precedes occurrence of insight. That is why, when we think about designing systems that facilitate the process of gaining insight, we need also to support occurrence of intuition and leap of faith.

We have identified three paths of gaining insight which come from a positive or a negative feeling provided by intuition. Those paths are compatible with the triple path model of gaining insight (Klein, Jarosz, 2011). Based on our analysis, we have created a model that describes the process of gaining insight with respect to intuition and leap of faith. Even though we have identified three paths of gaining insight, we assume that there might be more possibilities of gaining insight. However, we treat our results as a good starting point to look at insight as a part of a more complex process of decision-making than just an act of 'getting out of impasse'.

We consider essential to the study insight in natural decision-making environments where the role of intuition, leap of faith and conscious analysis are visibly important. We also find it significant, instead of studying intuition and insight as separate elements, to look at those cognitive acts as integrative parts of a more complex decision-making process. To do this, we need to conduct more field studies to identify occurrence of intuition, leap of faith and insight in natural decision-making environment. After those kinds of exploratory studies, we find it significant to create laboratory studies that will remind real life situations and will allow to detect mutual relations of intuition and insight.

We hope our study will be helpful in designing systems that support insightful decision-making. So far, we find it important that the system allows decision-makers to provide analysis towards their unsupported intuition and leap of faith, because it could lead to insights with a correct solution. However, the system should verify as soon as possible when intuition is misleading. The system should assume the role of users' individual experience (also the implicit ones) not as an obstacle but rather as a supportive resource that could be helpful in the absence of clear facts.

ACKNOWLEDGMENTS

The research reported here has received funding from the European Union Seventh Framework Programme (FP7/2007- 2013) through Project VALCRI, EC Grant Agreement N° FP7-IP-608142, awarded to Middlesex University and partners. We also gratefully acknowledge the support from the West Midlands Police Force, UK, the Lokale Politie Antwerpen, and the Service Public Federal Interieur in Belgium.

REFERENCES

Bowers, K.S. (1984). On being unconsciously influenced and informed. In: The Unconscious Reconsidered, eds K.S. Bowers and D. Meichenbaum. New York, NY: John Wiley&Sons.

Bowers, K.S., Regehr, G., Balthazard, C., and Parker, K. (1990). Intuition in the context of discovery. Cognitive Psychology.

Collins, A.M. & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. Psychological Review.

Damasio, A. (1994). Descartes' error. Emotion, reason, and the human brain. New York: Pengiun Books.

Duncker, K. (1945). On Problem Solving. Psychological Monographs.

Gerber, M. Wong, B.L.W.,Kodagoda, N. (2016). How analysts think: decision making in the absence of clear facts. Proceedings of 2016 European Intelligence and Security Informatics Conference.

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. Annual Review of Psychology.

Hogarth, R. M. (2001). Educating Intuition. University of Chicago Press.

Jung-Beeman, M. Bowden, E.M., Haberman, J., Frymiare, J.L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. PLoS Biol.

Kahneman, D. & Tversky, A. (1974). Judgment under uncertainty: heuristics and biases. Science.

Klein, G. & Jarosz, A. (2011). A Naturalistic Study of Insight. Journal of Cognitive Engineering and Decision Making.

Klein, G. A. (2013). Seeing What Others Don't: The Remarkable Ways We Gain Insights. New York, NY, USA: PublicAffairs, Perseus Book Group.

Ohlsson, S. (1992) Information-processing explanations of insight and related phenomena. In: Advances in the Psychology of Thinking, eds. M. Keane and K. Gilhooly. London: Harvester-Wheatsheaf.

Reber, R., Ruch-Monachon, M. A., and Perring, W.J. (2007). Decomposing intuitive components in a conteptual problem solving task. Conscious Cognition.

Sternberg, R., Davidson, J. (1996). The Nature of Insight. MIT.

Zander, T., Ollinger, M., Volz, K.G. (2016). Intuition and Insight: Two Processes That Build on Each Other or Fundamentally Differ? Frontiers in Psychology.

Wertheimer, M. (1959). Productive Thinking. New York, NY: Harper.

Wong, B. L. W. (2004). Critical Decision Method Data Analysis. In D. Diaper & N. Stanton (Eds.), Task Analysis for Human-Computer Interaction (pp. 327-346). Mahwah, NJ: Lawrence Erlbaum Associates.

Who Is the Champion? Performance Confidence and Actual Performance among Emergency Medical Services Crew Leaders

Jitka GURŇÁKOVÁ[a], Martin SEDLÁR[a], and Peter GRÖPEL[b]
[a]*Institute of Experimental Psychology, Centre of Social and Psychological Sciences, Slovak Academy of Sciences, Bratislava, jitka.gurnakova@savba.sk*
[b]*Department of Applied Psychology: Work, Education, Economy; University of Vienna, Austria*

ABSTRACT
This study tested the relationship between performance confidence and actual performance in a sample of emergency medical service (EMS) crew leaders. **Method:** 51 EMS professionals who competed in the international Rallye Rejviz competition participated in the study. Their self-reported confidence was compared with expert-based performance ratings. Their decision-making strategies were explored in qualitative interviews. **Results:** Performance confidence was unrelated to actual performance. However, cluster analysis revealed three types of EMS professionals. Well calibrated activer searchers with the best performance collected the most information before the diagnosis. Overconfident believers relied on their initial intuitions or assumptions. Careful thinkers perceived the complexity of near-field task.

KEYWORDS
expertise; health; emergency medicine; confidence; performance

INTRODUCTION

Research on confidence in medical diagnostics has revealed mixed results, which indicates that professionals and medical students do not necessarily know how well or badly they perform (e.g., Friedman et al., 2005; Meyer, Payne, Meeks, Rao, & Singh, 2013). The present study extended this line of research by testing a sample of emergency medical service (EMS) crew leaders using a mixed-method design. We sampled EMS crew leaders and measured their self-reported performance confidence and expert-based rating of their actual performance in a simulated near-field competition. We also interviewed the participants on their decision-making processes and looked for the reason as well as consequences of either miscalibrated or well calibrated performance. The present study thus combines naturalistic and normative approaches to decision making, trying to discuss the issue of expertise.

METHOD

Participants

Participants were 51 EMS crew leaders (30 men and 21 women) competing in the Rallye Rejvíz 2015 competition in the Czech Republic. All participants were leaders of either a physician crew ($n = 24$) or a paramedic crew ($n = 27$). Physicians were in average 38.4 years old ($SD = 8.74$), with 13.0 ($SD = 9.14$) years of experience in healthcare and 6.9 ($SD = 5.46$) years of experience in the emergency medical service. They attended in average 2.84 ($SD = 3.76$) similar competitions in the past. Paramedics were 29.9 years old ($SD = 5.88$), with 8.0 ($SD = 5.76$) years of experience in healthcare and 6.8 ($SD = 5.35$) years of experience in the EMS. Prior experience in EMS competitions was 5.29 ($SD = 3.81$) events.

Task and procedure

This study is a part of extensive data collection during the international competition Rallye Rejvíz 2015 organized in the Czech Republic. The competition consists of 11 simulated near-field tasks with different levels of difficulty. The tasks are performed within 24 hours, modelled just as close to actual situations as possible. Performance in each task was rated by at least two expert judges and acting patients on the quality of communications. Evaluation criteria were selected to reflect optimal solution for each task. Hence, we supposed the total performance as a measure of expertise. To understand decision making processes of EMS crew leaders, we focused more properly on a task called “Stone”. After arrival on the task, participants received the following instructions: “Emergency dispatch centre received an emergency call from and send you to a man, ca. 35 years old, sudden back pain and nausea.” Participants had to examine the patient properly and make the diagnosis of renal colic. Furthermore, they had to perform an adequate on-site treatment together with proper handling of an unexpected anaphylaxis reaction to the treatment. Time limit for the task was 12 minutes. Immediately after each crew finished this task, we conducted a 10-minutes interview with the crew leader which was focused on his/her decision-making processes and his/her performance confidence. We then compared participants' confidence with the expert ratings.

Analysis

Pearson correlations were computed to test the relationship between self-reported performance confidence and the performance score rated by the judges. Cluster analysis (K-means) based on the confidence level and actual performance was used to identify potential underestimators and overestimators. Differences between clusters were tested with one-way ANOVA and independent *t*-tests. All interviews were transcribed and analysed by two coders using the ATLAS.ti software.

RESULTS

Performance confidence was unrelated to actual performance in the task Stone (r = .20; p = .11). Similarly, the level of total performance confidence was not associated with the total competition performance (r = .12; p = .23). We therefore computed cluster analysis on confidence and performance in the task Stone to analyse the sample. Three clusters were found that we labelled active searcher (Cluster 1; N = 29), believer (Cluster 2; N = 12), and careful thinker (Cluster 3; N = 10). Figure 1 depicts the results of the cluster analysis.

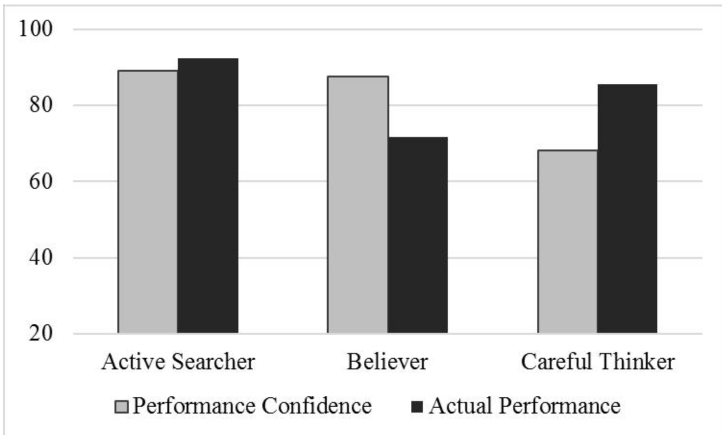


Figure 1. Types of EMS crew leaders based on their performance confidence and actual performance in selected task. The Y-axis represents percentages.

Participants in the clusters did not differ either in the healthcare experience, F(2, 37) = .22, p = .80, or in the EMS experience, F(2, 37) = 1.38, p = .26. However, differences in prior competition performance among the clusters close to signifficance were found, F(2, 37) = 2.85, p = .070. In particular, active searchers (Cluster 1) attended in average 5.23 (SD = 4.23) prior competitions, which was significantly more than believers (Cluster 2; M = 3.20, SD = 2.78), t(31) =2.42, p = .01. Physicians and paramedics were equally distributed across the clusters ($\chi2$ = 1.43, p = .49).

Active searcher (Cluster 1)

29 EMS professionals (15 physicians and 14 paramedics) showed both high performance confidence *(88.95%)* and high actual performance *(92.30%*). They had in average 9.62 (*SD* = 6.86) years of experience in healthcare, 5.91 (*SD* = 4.38) years of EMS experience and attended 5.23 (SD = 4.23) prior competitions (which was significantly more than Cluster 2, *t(31) = 2.42, p = .02*). They did the most proper examination of the patient *(90.95%), F (2, 48) = 12.31, p < .001*, made the most accurate diagnosis and its treatment *(90.35%), F (2, 48) = 5.972, p < .01,* and the on-site treatment of anaphylaxis (*93.10%), F (2, 48) = 21.08, p <.001*). Moreover, they obtained most favourable rating from the acting patients (*83.79%), F (2, 48) = 7.72, p = .001).* In addition, they achieved the highest actual performance not just in the task Stone, but also in the whole competition (*76.14%), F (2, 48) = 15.36, p < .001).*

They were labelled "active searchers" because, as found in qualitative interviews, they were likely to collect as much information as possible and listened to a patient's complaints carefully before making a diagnosis. If they had some initial idea about the source of the patient's complaints before all necessary examinations were done, they registered it but followed the medical guidelines properly and did all examinations as they were aware of the risk of selection bias.

Believer (Cluster 2)

EMS crew leaders with higher performance confidence *(87.55%)* than actual performance in the task Stone *(71.60%)* as well as in the whole competition *(65.23%).* In the selected task they received the lowest score in quality of examination (*73.15%*), diagnosis and therapy (*63.33 %*); treatment of anaphylaxis (*52.78%*) and the lowest rating from acting patients (*63.75%*). Despite similar overall results, this type comprised two completely different kinds of leaders: the group of 6 most experienced physicians with mean healthcare experience 18.33 years ($SD = 10.32$, $t(9) = 3.28$, $p = .01$) and 10.50 years of EMS experience ($SD = 5.79$, $t(9) = 2.93$, $p = .04$) but with the weakest experience with EMS competitions ($M = .83$; $SD = .98$, $t(9) = -1.96$, $p = .02$). And on the other hand – the group of 6 paramedics with the lowest healthcare experience ($M = 2.80$ *years*, $SD = 2.05$) as well as EMS experience ($M = 2.50$ *years*, $SD = 2.00$), who had experience from 3.20 ($SD = 2.78$) competitions.

Further analysis indicated that the Cluster 2 physicians and paramedics' processes of solving the task were different. Physicians were more frequently focused on seeking the key signs of renal colic than paramedics (problems with urination & tapottement, $t(10) = 1.53$, $p = .01$), while some others – to this hypothesis less relevant procedures – were omitted. As the result – they set the diagnosis of renal colic in this task correctly, $t(10) = 1.58$, $p < .000$. However, they needed longer time to identify unexpected anaphylactic reaction to selected treatment, $t(10) = -1.00$, $p = .03$.

On the contrary, paramedics in the same cluster used more frequently just the basic routine sequence of examination (blood pressure, saturation of O2, pulsation) and more frequently than physicians in the same cluster the ECG, $t(10) = -1.58$, $p < .000$, to reject the most threatening possibilities (like heart-attack). However, they omitted more specific examination procedures which could lead them to the diagnosis of renal colic. They set less accurate diagnosis more frequently and sometimes they just stopped at the level of descriptions of the symptoms instead of looking for their origin.

The qualitative analysis of interviews showed that physicians in the second cluster tended to trust their intuition ("diagnosis from the door") about the patient's troubles without special effort to verify it by proper finishing of all examination procedures. Moreover, it seems that their attention to the patient troubles after the conclusion about the diagnosis decreases as they overlooked more frequently the first signs of anaphylactic reaction to the renal colic treatment. On the other hand – paramedics in the second cluster had just a simple general idea what the reason behind the patient's complaints could be. After excluding the first possible life-threatening diagnosis, like heart-attack, they finished their search and did not proceed with the more specific examination to understand the patient's current troubles more properly. We called this cluster "believers" even if the origin of the high confidence of physicians and paramedics seemed different.

Careful thinker (Cluster 3)

EMS crew leaders (3 physicians and 7 paramedics) had lower performance confidence (*68.21%*) in comparison to middle level of actual performance in the whole competition (*71.69%*) as well as in the selected task (*85.59%*). More specifically, they showed middle level of performance in the examination (*87.78%),* diagnosis and its treatment (*78%*) and the treatment of anaphylaxis (*83.33%*). Since neither significant differences in age, length of praxis or competitions, nor special differences in the performance of the selected task between physicians and paramedics were found, we present descriptive characteristics of this group together: the mean healthcare experience was 11.43 years ($SD = 5.68$) and the mean EMS experience was 9.71 years ($SD = 6.70$). Members of this group participated in 4.14 competitions ($SD = 4.02$).

Comparison of quantitative as well as qualitative data showed that members of the third cluster have solid experience – either from real praxis and/or from competitions. They expressed more uncertainty accompanied with awareness of complexity of the task (possible diagnosis, what information are missing, simulation limits and competitions specifics). They set the diagnosis of renal colic just after or in the course of careful examination. Based on these characteristics we called this type "careful thinkers."

DISCUSSION

Described typology is in line with confidence in diagnostic correctness (e.g. Friedman et al., 2005; Meyer et al., 2013): most successful active searchers as calibrated, least successful believers as overconfidently miscalibrated, and average careful thinkers as underconfidently miscalibrated. We suppose that one of the key factors responsible for the success of active searchers was feedback about the performance, evaluation criteria, and recommended standards received in previous competitions (see Kahneman & Klein, 2009). Active searchers were also able to use a combination of intuitive and deliberative approach to decision making that is close to optimal (Croskerry & Norman, 2008).

With experience, health professionals gain intuition which is especially important for rapid recognition of cases. Physicians "believers" - who had many years of experience - relied heavily on intuition. But, as indicated by their lowest general performance, they sometimes err. Moreover, older leaders may fail at the competitions due to problems with adopting current recommended standards (see Choudhry, Fletcher, & Soumrai, 2005). On the other hand, young leaders may lack time to fully adopt the required standards. They were inadequately confident in their rather simple assumptions because of a lack of knowledge together with missing feedback.

The ability to accept the inevitability of some level of uncertainty is characteristic only for great commanders (Klein, 1999), because of that they can act decisively and prudently at the same time. Careful thinkers are not exactly like this. Due to their uncertainty they are too deliberative (hesitant), which is a disadvantage in dynamic time pressured settings.

All suggested factors (feedback/participation at competitions, years of experience, and uncertainty acceptance), their various levels, combinations and interrelationships seem to differentiate the EMS crew leaders according to their confidence and performance. It is important to state that this typology may not be final and represents rather working styles than stable personality types. We think that the same person may prefer different working styles depending on other person-related factors (fatigue, social desirability, etc.), and situational factors (familiarity, time pressure, etc.). Our typology stems from a specific case that needed diagnosing as opposed to clear cases that are straightforwardly recognised.

It is necessary to mention two kinds of limits of this study. Firstly, the limits related to the competition and simulation (not-representative sample, artificial character of the tasks, high achievement motivation, or possible leakage of information among the crews regarding the tasks). Secondly, the limits of used methodology (more superficial interviews, simple measure of performance confidence, specific case etc.). Although over- and underconfidence of EMS crew leaders may be viewed as cognitive bias, their confidence may depend on knowledge about the evaluation criteria (see Gigerenzer, Hoffrage, & Kleinbölting, 1991). Active searchers who participated more frequently in competitions knew how the judges rated their performance. Hence, they were able to include more relevant evaluation criteria into their self-rating of performance. On the other hand, performance in different kinds of competition tasks was rated in respect to their optimal solutions. Thus, we assumed that the competition results can serve as a valid measure of the quality of their real performance.

Further investigation of EMS crews is necessary for verification of our findings. Nevertheless, we believe that similar competitions are not just a great opportunity for EMS crews to learn, but they are a unique source of otherwise unobtainable data which can complement other NDM methods.

CONCLUSION

Our findings suggest that the champion is the experienced active searcher with appropriate feedback who is aware of the risk of incorrect intuitions. Thanks to his/her extensive knowledge he/she properly looks for information during the whole process of the patient's examination and treatment in accordance with current guidelines. As the example of believers shows, even rich experience in emergency medical services without proper feedback about their performance does not guarantee proper healthcare for a patient. Regular training in simulations as close to real-life as possible and/or supervision is needed to improve the annual statistics of diagnostics and treatment errors.

ACKNOWLEDGMENTS

We thank all organizers and participants of Rallye Rejviz 2015 for their precious collaboration. This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0361-12.

REFERENCES

Croskerry, P., & Norman, G. (2008). Overconfidence in clinical decision making. The American Journal of Medicine, 121(5A), S24-S29.

Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: the relationship between clinical experience and quality of health care. Annals of Internal Medicine, 142, 260-273.

Friedman, C. P., Gatti, G. G., Franz, T. M., Murphy, G, C., Wolf, F. M., Heckerling, P. S., . . .

Elstein, A. S. (2005). Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. Journal of General Internal Medicine, 20, 337-339.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. Psychological Review, 98, 506-528.

Kahneman, D., & Klein, G. (2009). Condition for intuitive expertise: A failure to disagree.

American Psychologist, 64, 515-526.

Klein, G. A. (1999). Sources of power: How people make decisions. Cambridge, MA: The MIT Press.

Meyer, A. N. D., Payne, V. L., Meeks, D. W., Rao, R., & Singh, H. (2013). Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. JAMA Internal Medicine, 173, 1952-1958.

Visualizing Decision Spaces for Option Awareness

Gary L. KLEIN, Mark S. PFAFF, and Jill L. DRURY
The MITRE Corporation

ABSTRACT

Understanding the situation space (facts about the situation) is necessary but not sufficient for developing and assessing decision options in complicated decision environments. Developing options under complexity requires translating from this situation space to a decision space (analytical information that facilitates option comparison). Properly visualized the decision space provides an understanding of the complex interactions between actions and conditions that yield mission success and failure -- enabling decision makers to plan options that will facilitate conditions yielding success and mitigate those conditions that will yield failure. We have called this capability *Option Awareness.* This paper will present many different interactive visualizations of decision spaces we have developed for varied decision environments. It will also elucidate the principles that yield better comprehension of the relationships among options, conditions and outcomes in those decision spaces.

KEYWORDS

Practical application, Decision Making, Sensemaking, Option Awareness.

DECISION SPACE EXAMPLES

Figure 25. Decision Space Interactive Visualization Envrionment (DECISIVE)

In this example of our interactive visualization tool, the top window presents the outcome of forcast model as a set of box-plots. While this is sufficient to compare options based solely on their outcomes (option awareness level 1), additional analysis and visualization is needed to yield a deeper option awareness to discriminate those factors that yield better vs. worse outcomes. In this tool, as the red criteria line is moved in the upper window by the user to indicate the bounder between better and worse, a statistical analysis executes in real time to produce the classifier tree in the lower window, which is a hierarchical structure starting with one node representing a scenario attribute that provides the greatest discrimination among outcomes. At each successively lower level tree branches either lead to another discriminatory factor or to a leaf node: the leaf nodes in the tree represent a set of outcomes that occur under a set of conditions that are described by the discrimination path traversed through the tree to get to the node. The results from this type of data mining can be used as part of pre-planning. To be used in real-time, the classifier tree can be translated into a playbook: the decision maker can assess the situation, find those conditions in the playbook and see what is recommend to do to obtain good outcomes.

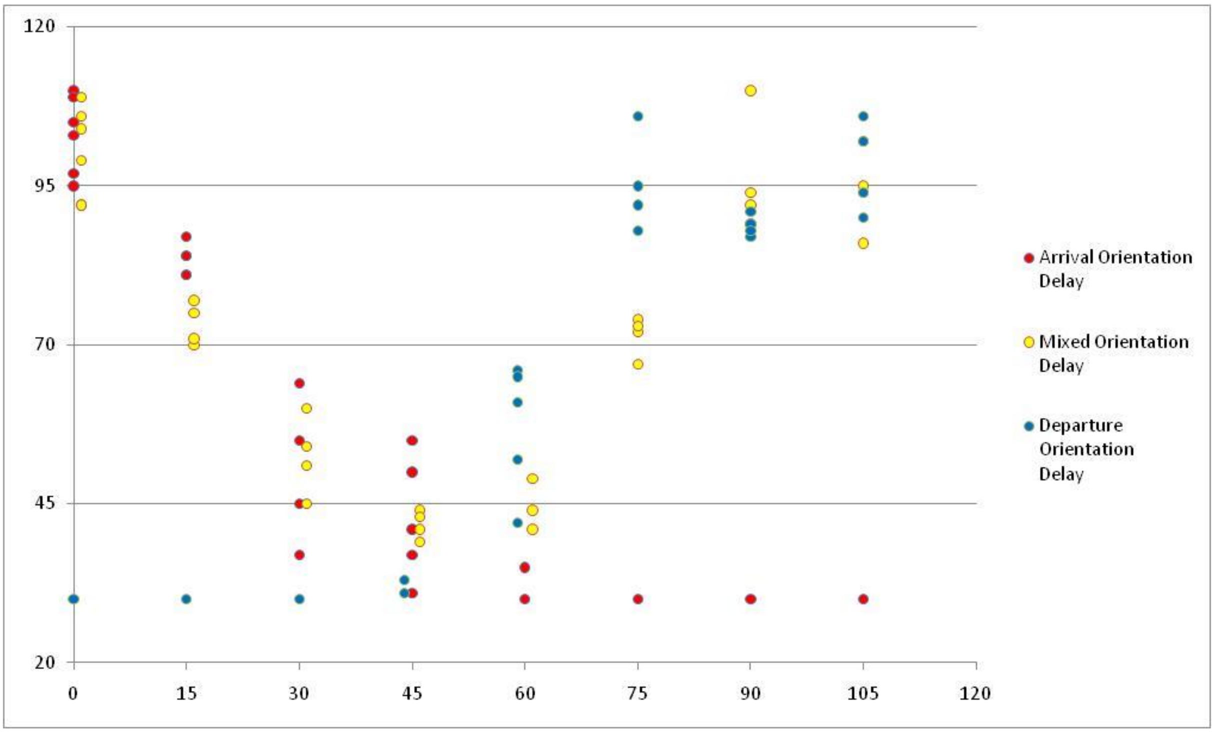


Figure 26. A Temporal Decision Sapce for Airport Reorientation Decisions

In this decision environment, airport managers must decide when would be the best time to change the orientation of flight processing from an arrival orientation to an departure orientation. Time is represented on the x-axis, while accumulated delay is represented on the y-axis. The range of delay that would occur at each time, for each orientation is represented by the columns of dots on the graph, which were generated by a forecast model.

At time 0, we are already in the departure orientation (the blue dots). Up through 30 minutes, this is the most robust option – the minimum delay we can get under current weather conditions. However, at 45 minutes the delay from this orientation begins to grow and become brittle.
Around 60 minutes from now, the arrival orientation (the red dots) becomes the most robust, and results in maintaining the minimum delay. So, the range of viable times for an orientation changeover is somewhere between 45 and 60 minutes from now.

With this information, we can begin preparing for that range now. In 30 minutes, based on better information on arriving and departing flights, the airport tower controllers nd Terminal Radar Approach Control Facilities can coordinate and shape the air traffic to achieve an exact time for the changeover.

These are just a couple of examples of the decision space examples that will be presented in the paper and demonstrated at the conference. A full paper with detailed explanations will be submitted at your request.

Team Workload Demands Influence on Cyber Detection Performance

Ricardo LUGO[a], Prosper KWEI-NAHR[a], Øyvind JØSOK[b], Benjamin J. KNOX[b], Kirsi HELKALA and Stefan SŰTTERLIN[ac]
[a]*Department of Psychology PACE CybORG, Innland Norway University of Applied Sciences, Norway*
[b]*Norwegian Defence Cyber Academy*
[c]*Oslo University Hospital*
[c]*The Research Centre for Child And Youth Competence Development, Inland Norway University of Applied Sciences*

ABSTRACT

The emerging nature of the cyber domain and limited research is posing new and challenging questions for military operations concerning human behaviour and performance in socio-technical complexity. The Hybrid Space is a conceptual framework that describes individual cyber operator's behaviors and team macrocognition within the complex environment - occupied by human and computer networks - between the cyber domain and the physical domain and the strategic and tactical domains. This applies the Hybrid Space conceptual framework by investigating how team workload demands are associated with operator's cognitive focus movements within the Hybrid Space, i.e., between the physical/cyber domain and tactical/strategical levels of thinking. Findings suggest that different team workload demands are associated with alternating movement patterns of the cyber defence operator's cognitive focus with the Hybrid Space.

KEYWORDS

Coordination, military, cyber, security, education and training

INTRODUCTION

This project applies the Hybrid Space conceptual framework by investigating on how team workload demands influence operator's movement within the Hybrid Space. The Hybrid Space (Jøsok, Knox, Helkala, Lugo, Sütterlin, & Ward, 2016) is a conceptual framework that describes a multi-domain environment where cyber as the key enabler intensify complexity that results in high cognitive load for individuals and teams. Jøsok et al. (2016) discusses that individuals and teams need to move in-between the cyber domain and physical domain on the x-axis, as well as in-between tactical and strategic considerations on the y-axis (See fig. 1), in order to ensure efficiency, communication and performance in a cyber-defense context (Knox et al., 2017). One recurring problem in researching cyber defense individuals and teams is that there is generally little knowledge about the metrics and measurements of performance in such an environment. "*For the most part, there currently exists no quantitative basis for assessing the performance of cyber defenders, whether at the individual, team, group or organizational levels*" (Forsythe, Silva, Stevens-Adams, & Bradshaw, 2013, p. 419). The nature of cyber operations merge human and technology, with humans operating at "human speed" cyber operations are increasingly automated processes operating at computational speed, reducing, or even removing, time to lead or time to ground communication between human agents during operations (Knox et al., 2017). This may require the cyber operator to shift domains more rapidly, moving from a cyber-environment to a more traditional physical military environment, taking different perspectives, while taking into consideration long-term strategic demands and more short-term tactical decisions to be able to support the current mission commander effectively (Williams, 2014). While studying individual and team cognition and processes in a controlled laboratory environments has its limitations (Klein et al., 2003), naturalistic approaches focus on generalizing findings from applied interventions that cannot be investigated at a micro level. We argue that the complexity of cyber operations, and the significance of these challenges, needs to be understood through their interdependency and the reciprocal processes that occur for effective team functioning in order to be transferrable to realistic educational and training efforts for cyber personnel (Jøsok et al., 2016; Knox et.al., 2017; Tikk-Ringas et al., 2014). Through operationalizing the Hybrid Space conceptual framework, we investigate how a naturalistic approach will help to gain a better understanding of the processes cyber operators and teams undertake in a cyber defence exercise by mapping individuals and their interactions and allowing for testable hypotheses on team performance outcomes.

METHODS

This study operationalized and quantified cyber operators' subjective movements in the Hybrid Space as a function of team workload demands during a Cyber Defense Exercise (CDX).
H1: Higher workload demands are associated with restricted movement in the Hybrid Space
H2: Higher cyber-physical and strategic-tactical movement will be associated with higher team workload demands

H3: Lower cyber-physical and strategic-tactical movement will be associated with increased Task-team demands
H4: Operators' movement between quadrants will be associated with Team Workload Demands

Cyber Defense Exercise

Data was collected during the Norwegian Defense Cyber Academy's (NDCA) annual Cyber Defense Exercise (CDX): Exercise Cold Matrix. This arena facilitates the opportunity for students to train in tactics, techniques and procedures for handling various types of cyberattacks. The exercise contributes to improving appreciation for the human and technical competences necessary to establish, manage and defend a military information infrastructure under simulated operational conditions. The students work in four teams of 9 or 10 members (of 37 students, 31 participated in the study), take decisions and act in order to strengthen operational freedom and control in the cyber domain. The four teams participating in the exercise work independently from each other but not against each other. Success is given as feedback to the decisions and actions taken during the exercise. Intrusions are initiated by an affiliated agency who are engaged to help the NDCA with their education. The teams are composed of 31 (M=22.7 years, SD=0.71) resembling a complete cohort enlisted in the NDCA. The exercise lasted four days where data was collected on the third day, which lasted 12 hours and data was collected every hour.

Measurements

Hybrid Space metrics and measurements

The Hybrid Space is mapped on a Cartesian plane and cyber operators marked their position simultaneously every hour during the third day of exercise (see Jøsok et al., 2016 for description). In addition, students noted their current task at each position, to give context to further analysis.

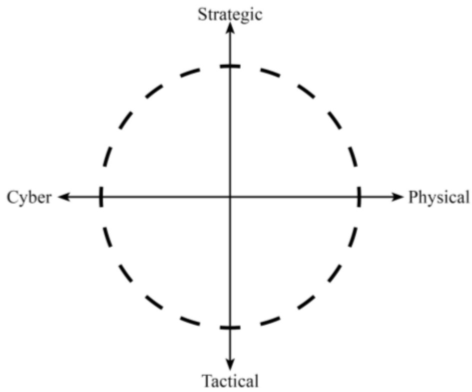


Figure 1: The Hybrid Space operationalized for measuring individual cognitive movement (x-axis represents cyber-physical considerations movement, and y-axis represents strategic tactical considerations movement.)

Hybrid Space operationalization

Movement in the Hybrid Space is operationalized through four constructs and represent the dependant variables in the study. Four dependent variables were created:
HSDT: distance traveled in the Cartesian Plane measured by Euclidian distance
HSxM: Movement along the cyber-physical domain (x-axis)
HSyM: Movement along the strategic-tactical domain (y-axis)
HSQC: Number of quadrant changes
Students reported movement within the Hybrid space was measured every hour during the exercise

Team Workload Questionnaire (TWLQ; Sellers et al, 2014)

The TWLQ was used to assess the workload demand in team tasks. The Task Workload Scale was omitted in this study due to the scale's focus on personal (individual) demands, but included the Teamwork and Task-Team component. The Team Workload Questionnaire (Sellers et al., 2014) was administered electronically by Qualtrics on the fourth day of the exercises. Time were set aside to answer the questionnaires.

RESULTS

To test the hypotheses, multiple regressions were performed.
Distance Travelled in the Hybrid Space was not influenced by Workload demands.

X-Axis movement was facilitated by team communication and coordination demands and restricted by dissatisfaction with team performance. These factors explained 12.4% of the movement along the cyber-physical axis of the Hybrid space (R^2 = .212, Adjusted R^2 = .124, F = 2.42, p = .044).

Y-Axis movement was facilitated through Communication Demands but restricted through higher time-share demands. These factors explained 24.5% of the variance of movement on the Strategic-tactical axis (R^2 = .295, Adjusted R^2 = .245, F = 5.87, p = .007). Communication demands in itself explained 29.2% of the variance of movement on the Strategic-tactical axis.

Quadrant changes was restricted by Performance Monitoring Demands (r= -.389; R^2 = 15.1%).

DISCUSSION

Results show that team workload factors can predict movement in the Hybrid Space supporting three of the four hypotheses (Hypotheses 2 & 3). Where only the total distance travelled could not be predicted by team workload demands (Hypothesis 1).

CONCLUSION

This research showed that team workload demands can influence movement in the Hybrid Space, but only on the cyber-physical axis (x-axis) and strategic-tactical axis (y-axis), and quadrant changes. While this research only focused on team workload demands, research focusing on individual characteristics and task requirements will be needed to further validate operationalizations and identify factors that can influence performance in the Hybrid Space.

REFERENCES

Forsythe, C., Silva, A., Stevens-Adams, S., & Bradshaw, J. (2013, July). Human Dimension in Cyber Operations Research and Development Priorities. In International Conference on Augmented Cognition (pp. 418-422). Springer Berlin Heidelberg.

Jøsok, O., Knox, B.J., Helkala, K., Lugo, R.G., Sütterlin, S., & Ward, P. (2016). Exploring the Hybrid Space - Theoretical framework applying cognitive science in military cyberspace operations. Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience Volume 9744 of the series Lecture Notes in Computer Science, 178-188. doi: 10.1007/978-3-319-39952-2_18

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. IEEE intelligent systems, 18(3), 81-85.

Knox, B., Jøsok, Ø., Helkala, K., Khooshabeh, P., Ødegaard, T., Kwei-Nahr, P., Lugo, R.G., & Sütterlin, S. (submitted). Socio-technical communication: The Hybrid Space and the OLB-Model for science-based cyber education. Manuscript submitted for publication Military Psychology

Tikk-Ringas, E., Kerttunen, M., & Christopher, S. (2014). Cyber Security as a Field of Military Education and Study. Joint Force Quarterly, 75.

Williams, B. T. (2014). The Joint Force Commander's Guide to Cyberspace Operations. Joint Force Quarterly, 73(2nd quarter).

Cognition at the Crime Scene: Identifying Cognitive Demands on Professional Judgement & Decision Making Expertise of Crime Scene Examiners

Amanda MARTINDALE[a], Dave COLLINS[b], and Victoria MORTON[c]
[a]University of Edinburgh
[b]University of Central Lancashire
[c]Scottish Police Authority Forensic Services

ABSTRACT
This research highlights the importance of professional judgement and decision making expertise within a skilled Scene Examination workforce. The use of Applied Cognitive Task Analysis allowed for illumination and provision of feedback on the thought processes of experienced Scene Examiners, including detailed observable information about actions taken, situation assessment, and use of critical cues. This information carries a range of operational uses, and can be a valuable addition to existing training. Implicit findings indicated the potential to move training towards a more adaptive expertise base, with consequent gains for the quality of service, and a more flexible and adaptable workforce.

KEYWORDS
Decision Making; Crime scene examination; Accessing expert cognition; Cognitive demands.

INTRODUCTION

Scene Examiners (SEs) meticulously search the crime scene, recording and recovering forensic evidence, which can ultimately be used to prove or disprove if a crime has occurred. These stressful field conditions are characteristic of the high stakes which typify naturalistic environments including ill-defined and competing goals, conditions of uncertainty, and time pressured decision making (Ross, Shafer & Klein, 2006).

Given the inherent complexities of crime scenes, and the huge amount of visual information available, the professional judgment and decision making (PJDM) expertise of SEs is of the utmost importance to the successful delivery of this service. From an organisational and business perspective, international standards for professional judgement have recently been established with governing bodies considering how they can demonstrate that they meet these standards through accreditation criteria (e.g., ISO/IEC17020). Although professional judgement is widely recognised as a necessary and intrinsic part of daily scene examination activity (and indeed of practice across many professions) it is surprisingly difficult to define and operationalize. We suggest that PJDM offers insight into how practitioners ‘think’ in action. Service delivery is a series of judgments and decisions and SEs are required to process vast amounts of information, to be able to think on micro and macro levels (often at the same time) and to rapidly formulate and enact coherent plans of action (Martindale & Collins, 2012).

The development of PJDM expertise has been studied in several high performance domains (e.g., psychological support, elite sport and coaching; Martindale & Collins, 2013) and is readily applicable to parallel domains of human performance. Indeed, as SEs develop this expertise, so their cognitive development, knowledge structures, and reasoning processes become more sophisticated and enhanced (Hoffman, 1996). Yet, these processes are by their very nature ‘covert’ - making them very difficult to ‘see’ and therefore to understand and train in novice and developing practitioners.

This work investigating PJDM expertise in Scene Examination attempted to ‘make thinking visible’ by accessing and capturing the thought processes of experienced SEs, identify the cognitive demands on PJDM, provide individual feedback to SEs involved in this work, and to develop a simulation scenario based training tool as a means of accessing and documenting expert knowledge. We were also interested to gain the SE’s perspectives of the benefits and limitations of using the Applied Cognitive Task Analysis methodology.

METHOD

Participants

Following institutional ethical approval and informed consent, 6 scene examiners were recruited (2 male and 4 female). This exceeded the recommendation of 3 – 5 participants by Militello and Hutton (1998). Participants were recruited across three regional areas (2 from each area). All participants were deemed competent and monitored by their National govering body. Participants had worked in the field for an average of 18 years (range 6-31 yrs). Participants interviewed were either experienced Scene Examiners (3) or Scene Examination Supervisors (3) and were selected by the National governing body senior management on the basis of their experience and expertise (e.g., Militello & Hutton, 1998). In addition to their Scene Examination role, participants also had experience in a range of other roles including: fingerprint officer/expert, footwear & tool impression comparison, ballistic examiner, photography assistant, and medical photographer.

Materials

Cognitive Task Analysis (CTA) techniques have been shown to produce greater amounts and qualitatively more meaningful data than observations of the task or self-generated explanations of the task (Tofel-Grehl & Feldon, 2013). This work involved the use of Applied Cognitive Task Analysis (ACTA, Militello & Hutton, 1998) to understand the cognitive task demands on PJDM of SEs, and to identify the key cognitive elements required to perform proficiently. This meta-method comprises of three techniques: task diagram, knowledge audit, and simulation scenario (crime scene examination response to a complex and major incident) which complement each other, but which also tap into different elements of cognitive skill (McAndrew & Gore, 2013). The method therefore offers a unique window on the thought processes and PJDM expertise of Scene Examiners, and transforms covert thinking into detailed observable information about actions taken, situation assessment, and the use of critical cues. Previous findings have indicated high levels of validity and reliability with modal statistics occurring in the range of 90% to 95% (See Militello & Hutton, 1998, for a full description of each ACTA method).

Procedure

Each interview took approx. 3 hours and was conducted on site at a time and location that was operationally viable. Throughout the interview process the researcher made detailed notes following the outlines provided on the ACTA Job Aid cards (ARA Inc, 2005). The interviews were recorded on a digital voice recorder and transcribed verbatim. Audio recordings of the interviews were utilized afterwards to ensure the completeness and accuracy of the data. The provision of individual feedback to SEs involved the presentation, read-through, and verification of each stage of the ACTA protocol as well as the resulting Cognitive Demands Table. In particular, SEs were asked about the 'completeness' and 'accuracy' of the information at each stage and were actively encouraged to highlight anything which was missing or incorrect. Following the provision of feedback and data verification process, SEs were asked to report on the perceived benefits and limitations of accessing expert cognition in this way. Raw data quotations were qualitatively analysed using inductive analysis to build up a series of lower order and higher order themes.

RESULTS

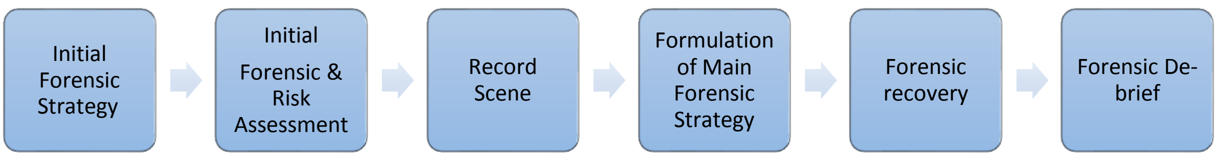
Due to the richness and depth of the qualitative data collected, illustrative examples will be used to outline the results from the task diagram and knowledge audit stages. Due to the restricted nature of the data, an overview of the broad themes from the Cognitive Demands Table will summarise the findings. These findings are used to illustrate the cognitive demands on professional judgment and decision making expertise in Scene Examination.

Illustrative example: Scene Examination Supervisor (32 years experience)

Stage 1: Task Diagram

The task diagram for Scene Examination Supervisor A focuses on the strategy for forensic recovery at a crime scene (see Figure 1). The first stage involves the formation of an initial forensic strategy which includes gaining background information, and establishing health and safety issues. The second stage is the initial forensic assessment in which the SE attends the scene with others and undertakes an initial forensic and risk assessment of the situation. The third stage is to record the scene as is (e.g., briefing and instructing the SE team to video the entire scene). The fourth stage involves the formulation of the main forensic strategy, and the fifth is the forensic recovery of evidence itself. The sixth and final stage is a forensic de-brief with the SE team. Scene Examiner A identified the forensic recovery stage as the most cognitively demanding stage within the overall task.

Figure 1. Exemplar Task Diagram for Scene Examination forensic recovery



Stage 2: Knowledge audit

Table 1 provides an extract from Scene Examination Supervisor A's knowledge audit. The example included in this table is discussed in more detail below.

Table 1. Exemplar Opportunities/Improvising from Knowledge Audit

Cognitive Element of Expertise	Cues & Strategies	Why Difficult? (for someone less experienced)
Opportunities/Improvising Experts are comfortable improvising – seeing what will work in this particular situation; they are able to shift directions to take advantages of opportunities. *Can you think of an example where you have improvised in this task or noticed an opportunity to do something better?*	• Prioritise forensic opportunities (fibres, shoe impressions, pieces of door, blood) • Resist 'quick fix' frame of mind due to time pressure to catch up • Resist temptation to bypass certain aspects to get to aspects that would yield quicker results (fingerprints as quick fix, but destructive process) • Conscious decision to operate in a structured manner • Sense making/ story building/ running mental simulations • Ensure compromises are not to the detriment of the forensic recovery • Where compromise is necessary (e.g., major road) weigh up consequences/ implications	• May be tempted to go for the 'quick fix' to catch up – potentially compromising the scene • Where there are ways to cut a corner or improvise they may not resist in order to do the job properly

Opportunities/Improvising

Scene Examiners were asked to provide an example of where they have improvised in the task of forensic recovery or noticed an opportunity to do something better. In terms of cues and strategies, scene examiners may look to prioritize forensic opportunities (e.g., fibres, shoe impressions, and blood) whilst resisting the 'quick fix' temptation to bypass certain aspects of the recovery to get to aspects that would yield quicker results. For example, fingerprints may yield a 'quick fix' opportunity, but the collection of prints is a destructive process to other potential forensic material. The conscious decision to operate in a structured manner allows for the opportunity to make sense of the situation, build a bigger picture of the crime scene, and ensure that compromises are not to the detriment of the forensic recovery. This may be difficult for someone less experienced, as a novice is more likely to go for the 'quick fix', especially if under time pressure.

Cognitive Demands

A cognitive demands table was compiled to consolidate and synthesize the data from the six Scene Examiners. Further cognitive demands tables were compiled to represent responses from SE's in each of the three regions. These tables contained information on the difficult cognitive elements of the job, why this would be difficult for someone less experienced, the potential errors someone less experienced may make, and the cues and strategies that more experienced personnel may rely on. Overall, four broad areas of scene examination (formulation of initial forensic strategy, initial forensic assessment, formulation of main forensic strategy, and forensic recovery) revealed between 26 and 31 cognitive demands, across regions. Forensic recovery as the central task revealed between 8 and 12 cognitive demands. These findings highlighted potentially important regional differences in the effectiveness of the approach to the examination of a scene by SEs.

Perceived Benefits & Limitations

Participants reported a number of perceived benefits and limtations to the use of the ACTA method for accessing SE cogntion in this way. Benefits included the effectiveness of ACTA in generating useful information (e.g., encourages reflection, useful addition to SOPs, concise description of the job), potential uses for information generated using ACTA (e.g., initial and ongoing training tool, provision/exchange of vicarious knowledge, useful to explain scene examination to new police officers, useful in developing a National model for major incidents) and as a valuable addition to existing training (e.g., more concise way of reflecting on/progressing through scenarios of increasing difficulty). Limitations included difficulties with knowledge elictiation (e.g., difficulty generating examples 'on the spot', uncertainty over whether responses were appropriate, inherent time constraints on explaining all aspects of the job), the combination of volume and serious crime (e.g., different strategic approaches to volume and serious crime), and the realism/scope of scenarios (e.g., second-guessing in simulation scenario what others have already done, wouldn't be assessing cold – would have much more information).

DISCUSSION

Accessing and documenting the cognitive demands on PJDM in SEs illuminates the cognitive processes of these skilled professionals. This information feeds directly into the necessary training systems and structures required to develop and accelerate expertise in Scene Examination. This research shows that using a naturalistic approach to investigating PJDM expertise in Scene Examination can yield rich and detailed insights into the cognitive demands inherent in this complex task, and offer a unique window on 'cognition at the crime scene'.

Implicit Findings and Interpretation

Several implicit findings emerged from the data in addition to the explicit and generally positive perceptions reported earlier. For example, the potential to move training towards a more adaptive expertise base, with consequent gains for the quality of service emerged as a consistent if implicit theme. The current culture within Scene Examination appears to be overly geared towards 'competent' practice. This is characterised by the prevalence of SOPs as markers of 'quality' and the use of competency-based assessment in order for SEs to reach a baseline (competent) level of service provision. Some of the responses generated by participants reflect this culture (e.g., difficulty in 'thinking on the spot' and being 'unsure' if responses were 'correct') suggest that perceptions of 'what is expected' are procedurally based.

In contrast, further responses from participants in this study suggest some inherent drivers (indeed desire) for the Scene Examination culture to be expertise-based. This would be characterised by on-going training and development opportunities for SEs (largely embedded within practice rather than bolt-on 'top-up' courses), plus the development and enhancement of practice through a community of practice approach (Wegner, 2011). Professional learning systems and structures that would allow SEs to share experiences and explore the 'shades of grey' inherent in practice would relieve some of the competency-based constraints and make for a more flexible and adaptable workforce. Recent positive exemplars of this approach are apparent from a wide variety of professional domains, including nursing, military and elite refereeing (e.g., Collins et al., 2015).

Limitations

There are some important limitations to note, many of which were identified by the participants in reflecting on the process undertaken. Although this research follows other similar studies (e.g., McAndrew & Gore, 2013) in an attempt to capture 'cognition in the wild', the recall of past events/circumstances and the simulation of crime scene scenarios are not directly parallel to conducting scene examination in the real-time present. Of course, the logistics of this are complex, not least because of the difficulties with introducing researchers to crime scenes. Although 'think aloud' technology may go some way to overcoming some of the practical constraints on accessing expert cognition, the threat to compromising a crime scene would need to be firmly established and negated prior to any data collection at live crime scenes, as well as authorisation from the criminal justice system.

CONCLUSION

This study makes a unique contribution to the literature by outlining the cognitive demands on PJDM expertise in scene examination. These findings have been useful from an operational perspective (e.g., allocating staff to respective crime scenes) and in terms of developing training and ongoing professional learning systems and structures (e.g., in designing a scenario-based training tool for scene examination response to a complex and major incident). This data is also contributing to the development of a National proficiency scale for scene examination, which in turn has implications for policy, such as a National career grade framework for scene examination.

ACKNOWLEDGMENTS

The authors wish to thank the participants and senior management of the Scottish Police Authority Forensic Services for their time, willingness, and insight, and the Scottish Institute of Policing Research and the University of Edinburgh College of Arts Humanities and Social Sciences for funding this research.

REFERENCES

Collins, D., Burke, V., Martindale, A. & Cruickshank, A. (2015). The illusion of competency versus the desirability of expertise: Seeking a common standard for support professions in sport. *Sports Medicine, 45*, 1–7.

Hoffman, R. R. (1996). How can expertise be defined? Implications of research from cognitive psychology. In: Williams R, Faulkner W, Fleck J. eds. *Exploring expertise*. Edinburgh, Scotland: University of Edinburgh Press, 81–100.

Martindale, A. & Collins, D. (2012). A professional judgment and decision making case study: Reflection-in-action research [Special issue]. *The Sport Psychologist, 26*, 500-18.

Martindale, A. & Collins, D. (2013). The development of professional judgment and decision making expertise in applied sport psychology. *The Sport Psychologist, 27*, 390–98.

McAndrew, C. & Gore, J. (2013). Understanding preferences in experience-based choice: A study of cognition in the "wild". *Journal of Cognitive Engineering and Decision Making, 7*, 179-97.

Militello, L. G. & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*, 1618– 41.

Ross, K. G., Shafer, J. L. & Klein, G. (2006). Professional judgments and naturalistic decision making. In Ericsson KA, Charness N, Hoffman RR, Feltovich PJ. eds. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press, 403-20.

Tofel-Grehel, C. & Feldon, D. F. (2013). Cognitive task analysis-based training: A meta-analysis of studies. *Journal of Cognitive Engineering and Decision Making, 7*, 293-304.

Wenger, E. (2011). *Communities of practice: A brief introduction*. National Science Foundation (U.S.)

Development of a Scenario-Based Training Tool for Crime Scene Examination Response to a Complex and Major Incident

Amanda MARTINDALE[a], Dave COLLINS[b], and Victoria MORTON[c]
[a]*University of Edinburgh*
[b]*University of Central Lancashire*
[c]*Scottish Police Authority Forensic Services*

ABSTRACT
Following the use of knowledge elicitation techniques with Scene Examiners, and in collaboration with the Forensic Services Scottish Multimedia Unit, a scenario-based training tool was devised to simulate four separate but related scenes of crime. Exemplar responses to this complex and major incident were generated by experienced Scene Examiners across three National regional areas. This is an example of the type of training tool that could be generated as part of a wider expertise-based training environment for Scene Examination services.

KEYWORDS
Learning and Training; Crime scene examination; Scenario-based training; Judgment and decision making.

INTRODUCTION

The aims of this research were to develop a simulation scenario training tool as a means of accessing and documenting knowledge in Scene Examination, particularly with regards to Professional Judgment and Decision Making (PJDM), and to ascertain how knowledge accessed via simulation scenarios could be packaged to facilitate on-going professional learning.

This research investigating PJDM expertise in Scene Examination (SE) attempted to 'make thinking visible' by accessing and capturing the thought processes of experienced SEs. This involved the use of Applied Cognitive Task Analysis (ACTA, Militello & Hutton, 1998) to elicit knowledge, understand the cognitive demands on PJDM in Scene Examination, and to identify the key cognitive elements required to perform proficiently.

The data generated via ACTA offered a unique window on the thought processes and PJDM of Scene Examiners, and transformed covert thinking into detailed observable information about actions taken, situation assessment, and the use of critical cues. This information was then used to create a scenario-based tranining tool for scene examination response to a complex and major incident. Synthesised responses to a complex simulation scenario formed 'exemplar responses' from experienced scene examiners in three National regions for use as part of a training tool.

These phases of work were designed to establish cognitive authenticity (the emulation of the features an expert would perceive in the performance environment that support perception and decision making; Ross & Pierce, 2000) within the Scene Examination profession and therefore contribute to sustainability for the future in terms of retaining expert knowledge and training PJDM expertise within the Scene Examination workforce.

METHOD

Participants

Following institutional ethical approval and informed consent, 6 scene examiners were recruited (2 male and 4 female). This exceeded the recommendation of 3 – 5 participants by Militello and Hutton (1998). Participants were recruited across three regional areas (2 from each area). All participants were deemed competent and monitored by their National govering body. Participants had worked in the field for an average of 18 years (range 6-31 yrs). Participants interviewed were either experienced Scene Examiners (3) or Scene Examination Supervisors (3) and were selected by the National governing body senior management on the basis of their experience and expertise (e.g., Militello & Hutton, 1998). In addition to their Scene Examination role, participants also had experience in a range of other roles including: fingerprint officer/expert, footwear and tool impression comparison, ballistic examiner, photography assistant, and medical photographer.

Materials

This project involved the use of Applied Cognitive Task Analysis (ACTA, Militello & Hutton, 1998) to understand the cognitive task demands on PJDM of SEs, and to identify the key cognitive elements required to perform proficiently. This meta-method comprises of three techniques (task diagram, knowledge audit, and simulation interview) which complement each other, but which also tap into different elements of cognitive skill (McAndrew & Gore, 2013). For the purposes of developing a scenario-based training tool the third stage of ACTA (Simulation Interview) was utilised. This method offers a unique window on the thought processes and PJDM expertise of Scene Examiners, and transforms covert thinking into detailed observable information about actions taken, situation assessment, and the use of critical cues. Previous findings have indicated high levels of validity and reliability with modal statistics occurring in the range of 90% to 95% (Militello & Hutton, 1998).

Scenario Design

A simulated scenario of a complex and major incident was developed with support from the Forensic Services Scottish Multimedia Unit. This simulation incorporated four separate but related scenes of crime (first deposition site, deceased's home address, second deposition site, and suspect's home address) and included location maps, photographic stills, and panoramas for each of the scenes of crime. Each scene included briefing notes and the simulation developed over time to incorporate the four separate but related scenes of crime. This approach to scenario design could be adopted by other forensic services governing bodies, especially as the process of scene examination inherently involves video and photographic capture, providing a uniquely advantageous position from which to re-create crime scenes using high and low fidelity techniques, and therefore to offer cognitively authentic training experiences.

Simulation Interview

The third ACTA technique, the Simulation Interview, presented the participants with each challenging scenario in which participants were told "*As you experience this simulation, imagine you are the scene examiner in the incident. Afterwards, I am going to ask you a series of questions about how you would approach this situation.*" (ACTA Job Aid Card, ARA Inc, 2005). This simulation of a complex and serious incident allowed for an in depth exploration of the *major events* that would occur in addition to *actions*, *situation assessment*, *critical cues*, and *potential errors* that a less experienced scene examiner may make.

RESULTS

SEs responses to the simulation scenario (including major events, actions, situation assessment, critical cues, and potential errors someone less experienced may make) were synthesised to form exemplar responses from experienced scene examiners across three National regions. These findings highlighted potentially important regional differences in the effectiveness of the approach to the examination of a scene by SEs and provide an exemplar of the type of training tool that could be generated as part of a wider expertise-based training environment (Alison et al., 2013).

Simulation Interview Overview

A simulation interview overview table was compiled to consolidate and synthesize the data from the six Scene Examiners. Further simulation interview tables were compiled to represent responses from SE's in each of the three regions. These tables contained information on the major events that would occur, in addition to actions, situation assessment, critical cues, and potential errors someone less experienced may make. Overall, the major events that would occur were collated for each scene (first deposition site, deceased's home address, second deposition site, and supsect's home address) revealing between 18 and 20 major events, across regions. These findings highlighted potentially important regional differences in the effectiveness of the approach to the examination of a scene by SEs.

Due to the restricted nature of this data, it is not possible to provide illustrative examples of extracts from the Simulation Interview; however, these findings offer insight to the cognitive demands on professional judgment and decision making expertise in Scene Examination, and offer exemplar responses to a cognitively authentic simulation scenario for use as a training tool.

DISCUSSION

Simulation Scenarios: A Future Training Method

Scene Examiner responses to the simulation scenario (including major events, actions, situation assessment, critical cues and potential errors someone less experienced may make) were synthesised to form exemplar responses from experienced SEs in three National regions. The products of the ACTA simulation interview method were then utilised to develop a scenario-based training tool for crime scene examination response to a complex and major incident.

This simulation tool highlighted potentially important regional differences in the effectiveness of the approach to the examination of a scene by SEs. Going forward, it will provide the National governing body with a tool to access and collate the knowledge of the SE workforce, track development and changes in practice over time, and can be used as a training tool as part of a wider expertise-based training environment.

An Expertise-Based Approach to Scene Examination PJDM Training and Practice

Simulation interviews provide an opportunity for the understanding and development of declarative knowledge and reasoning in scene examination. Further, scenario-based tranining can facilitate expert learning by enabling professionals to form more complete metal models of practice as they move through the four cognitive steps of judgment, elaboration, flexibility, and decision making (Collins et al., 2015). This window of opportunity to make thinking 'visible' to supervisors, peers, and supervisees can provide a 'cognitive apprenticeship' model (Collins, Brown & Holum, 1991) for elaborating on the underpinning rationale for decisions, what other options were considered and rejected, and for explicitly exploring the choices made.

This focus on the development of professional judgment and decision making expertise would help to develop the thinking structures, adaptabiltiy, and critical analysis necessary to allow scene examiners to function with high levels of proficiency in their complex and dynamic envoronment. As Smith and colleagues so eloquently state: "academic research generally and our society particularly have largely neglected the fact that sound judgment and decision making are the crux of many professions. By understanding and communicating what professional decision makers do and how they do it well, we make valuable contributions both to our field and to the professional community at large" (p.4). Thus, the development of a scenario-based training tool for crime scene examination response to a complex and major incident enables the understanding and communication of PJDM witihn Scene Examination, and to wider criminal justice partners.

ACKNOWLEDGMENTS

The authors wish to thank the participants and senior management of the Scottish Police Authority Forensic Services for their time, willingness, and insight, the Forensic Services Scottish Multimedia Unit for assistance with scenario design, and the Scottish Institute of Policing Research and the University of Edinburgh College of Arts Humanities and Social Sciences for funding the initial research. The development of the Scenario-Based Training Tool was supported by an ESRC Impact Acceleration Account Grant.

REFERENCES

Alison, L., van den Heuval, C., Waring, S., Power, N., Long, A., O'Hara, T. & Crego, J. (2013). Immersive simulated learning environment for researching critical incidents: A knowledge synthesis of literature and experiences of studying high-risk strategic decision making. *Journal of Cognitive Engineering and Decision Making, 7*, 255-72.

Collins, D., Burke, V. Martindale, A. & Cruikshank, A. (2015). The illusion of competency versus the desirability of expertise: Seeking a common standard for support professionals in sport. *Sports Med, 45*, 1 – 7.

Collins, A., Browm, S. J. & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Education, 6(11),* 38-46.

Militello, L. G. & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*, 1618– 41.

McAndrew, C. & Gore, J. (2013). Understanding preferences in experience-based choice: A study of cognition in the "wild". *Journal of Cognitive Engineering and Decision Making, 7*, 179-97.

Phillips, J. K., & Battaglia, D. A. (2003). Instructional methods for training sense-making skills. In Phillips, J. K., Klein, G., & Sieck. W. R. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. In D. K. Koehler and N. Harvey (Eds.). Blackwell handbook of judgment and decision making. Wiley-Blackwell.

Phillips, J. K., Klein, G. & Sieck. W. R. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. In D. K. Koehler and N. Harvey (Eds.). Blackwell handbook of judgment and decision making. Wiley-Blackwell.

Ross, K. G. & Pierce, L. G. (2000). Cognitive engineering of training for adaptive battlefield thinking. In IEA 14th Triennial Congress and HFES 44th Annual Meeting (Vol. 2, pp. 410-413). Santa Monica, CA: Human Factors.

Quenching the Thirst for Human-Machine Teaming Guidance: Helping Military Systems Acquisition Leverage Cognitive Engineering Research

Patricia L. McDermott[A], Katherine E. Walker[A], Cynthia O. Dominguez, Ph.D.[a], Alex Nelson[B] and Nicholas Kasdaglis, Ph.D.[a]
[a]*The MITRE Corporation*
[b] *Air Force Research Laboratory*

ABSTRACT
There is a need in systems acquisition that is not currently being met. The insights from cognitive engineering research in human-automation interaction are not being systematically applied to acquisition processes associated with operational military systems. To address this gap, we synthesized guidance from the literature and translated it into a set of general cognitive interface requirements for human-machine teaming. By presenting the guidance as requirements, we are attempting to remove barriers from effective insights being used in implementation. This paper describes ten themes of human-machine teaming that need to be supported: Observability, Predictability, Directing Attention, Exploring the Solution Space, Directability, Adaptability, Common Ground, Calibrated Trust, Design Process, and Information Presentation. Example requirements are provided for Exploring the Solution Space. The general set of requirements can be tailored to specific systems as needed. To support this tailoring, we are developing and piloting cognitive task analysis techniques focused on human-machine teaming.

KEYWORDS
Coordination; military; cognitive engineering; common ground; systems development; human automation interaction

INTRODUCTION

The cognitive engineering (CE) communitity has produced useful research findings and related insights for designing effective collaboration between automation and humans. These findings and insights are not making their way into the design of operational systems. We are finding a general lack of familiarity and application of these findings among program managers, systems engineers and developers. There is no systematic process for applying this human-automation research when engineering complex human-machine systems. Users and designers talk about concepts like "transparency" but it is not clear what this means in terms of design requirements for a system. The systems acquisition community may not be familiar with the CE literature or they may not have the necessary background to apply the findings to envisioned or upgraded systems. Regardless, a gap exists between the CE research and the system accquistion process for fielding new systems.

The lack of CE insight in highly autonomous systems is apparent. So-called "clumsy" automation has many negative consequences, such as brittle performance, miscalibrated trust in the system, and lack of user acceptance (Wiener, 1989; Sarter, Woods, & Billings, 1997; Lee & See, 2004; Parasuraman & Riley, 1997). Although the research community has answered the call to provide guidance on how to improve interactions between humans and advanced technological systems (Lyons, 2013), surprisingly little work has been done to translate the literature into materials that non-researchers can use as guidance for design of highly automated and autonomous systems.

Our analysis has aggregated and packaged the human-machine teaming (HMT) literature into several interconnected themes for the design of human-autonomy interaction.

METHODS

Researchers completed three main tasks: literature review, analysis, and creation of general cognitive interface requirements. First, to leverage and package existing research for system acquision personnel and program managers, we reviewed the large applied literature on human-automation interaction, human-machine trust, human-robot teaming, flight deck automation surprises, and other related work. Seventy-six papers were triaged to identify the papers most likely to yield citable guidance for HMT. We summarized 39 of these papers, extracting key information on HMT philosophy, principles, key terms, and requirements and considerations (The authors can be contacted for a list and summary of papers reviewed). The research findings were analyzed for common themes and elements of cognitive support. The analysis yielded ten themes that capture unique aspects relevant for HMT design across the literature and domains. Guidance was sorted into these themes and then translated into requirements language.

RESULTS

Themes for Design of Effective HMT

Below are the ten themes that emerged from our bottom-up analysis of HMT guidance in the literature. The first eight reflect guidance on design content and thus are ripe for tailoring through cognitive task analysis (CTA). The last two, design process and information presentation, are of a different nature. They capture insights about how design teams work together and processes they use for analysis and validation, as well as how to apply basic information presentation principles. These processes are consistent regardless of domain and therefore are not directly addressed in a CTA.

Design Content Themes to be Addressed through CTA

- Common Ground. Support for common ground means pertinent beliefs, assumptions, and intentions are shared among team members. Common ground is constantly and actively updated and maintained as team members maintain a shared picture of what's happening in the world and the status of the overall plan (Bradshaw et al., 2004; Klein, Feltovich, Bradshaw, & Woods, 2005). Teammates engage in proactive backup behavior and proactive communications (Johnson, Bradshaw, Hoffman, Feltovich, & Woods, 2014).
- Observability. Observability is defined as transparency into what an automated partner is doing relative to task progress. Observability can be best achieved by supporting shared understanding of the problem to be solved and progress towards goals (Bradshaw et al., 2004; Joe, OHara, Medema, & Oxstrand, 2014). This understanding must be provided early in use in a way that is comprehensible to operators, and it must be informative enough to enable effective intervention (Hoff & Bashir, 2015; Parasuraman & Riley, 1997).
- Predictability. Future intentions and activities of the partner are discernible and understandable. It is a subset of observability, and based on an understanding of the automation's goals, abilities, and limitations (Klein et al., 2005). Supporting predictability aids the user in anticipating changes in the system and gives them an idea of where to look next (Bradshaw et al., 2004; Lyons, 2013).
- Directability. Directability is supported when humans are enabled to direct and redirect an automated partner's resources, activities, and priorities. The people on the scene will inevitably know about information associated with the particular situation that the automated system will have no knowledge of (Roth et al., 1987). Ultimately then, the human must be in control and be able to stop the process, toggle between levels of autonomy, or override and manually control the automation (Joe et al., 2014; Lyons, 2013).
- Directing Attention. Each partner must be able to direct the attention of the other partners to critical problem features, cues, indications, and warnings. Automation must be able to flag additional and relevant information it cannot process so the human can redirect. It should also communicate proactively when information it has becomes relevant, such as when goal obstacles are encountered (Joe et al., 2014).
- Exploring the Solution Space. Exploring the Solution Space is supported when partners can leverage multiple views, knowledge, and solutions to jointly understand the problem space. Automation should be able to rapidly generate multiple distinct courses of action and give people ways to rapidly compare those solutions. Both humans and technology should be able to expand or constrict the solution considerations and shift perspectives (Woods & Hollnagel, 2006).
- Adaptability. Adaptability is defined as partners being able to recognize and adapt fluidly to unexpected characteristics of the situation. Supporting this requires multiple options to both recognize an unexpected situation and to address and recover from it (Johnson et al., 2015).
- Calibrated Trust. Calibrated trust is supported when human operators have a strong understanding of when and how much to trust an automated partner in context. Users should understand clearly when automation can be relied upon, when more oversight is needed, and when performance is unacceptable (Joe et al., 2014). Calibrating trust requires providing information on the source of the diagnostic information, and the credibility of the alert (Johnson et al., 2014).

Design Process Themes (not addressed through CTA)

- Information Presentation: This category was created to capture published recommendations on how to present information to support simplicity and understandability.
- Design Process. Elements of the cognitive engineering-driven design process of systems engineering that must be incorporated to enable effective HMT.

Cognitive Interface Requirements

Cognitive Interface Requirements were drafted for each theme. These themes go beyond informaiton presentation to address how the automation and the human(s) can work together to achieve task and mission goals. The requirements are purposefully general in nature. The goal is to provide a comprehensive general set of requirements that can be tailored to specific domains. Toward this end, the requirements contain rationale - italicized statements that convey the purpose behind the requirement – so the rationale is not lost when the requirements are tailored. Each requirement is tagged with a numbers that link it to the relevant research article(s) that inspired it. The requirements have hyperlinks (shown in blue underlined text) which provide an example or more details about the requirement.

Observability Example:

- The system shall provide information about how automation activities may evolve in the future. [18] Example: Auto GCAS merging chevrons

Calibrated Trust Example:

- The system shall convey system reliability in different contexts *to promote calibrated trust*:
    - situations in which the automation can be relied upon;
    - situations that require increased oversight by personnel;
    - situations in which the automation's performance is not acceptable. [3]

The set of general requirements can be tailored for specific systems. Designers can identify which requirements in the general set are applicable to a given system. For example, if the envisioned automation is for a commander's virtual assistant, the requirements on remote perception may not be applicable. Likewise, the requirements relating to what-if analysis may not be applicable for small autonomous vehicles. Another way to tailor the requirements is to rewrite them using more specific language. Consider this requirement for Exporing the Solution Space:

- The automation shall allow a human to input information that wasn't initially taken into account by the algorithms *so the algorithm can incorporate new information that the human has or expertise of the human* [74].

For a new automated planning system with a logistics tool for calculating battery and fuel expenditures, it could be rewritten as follows:

- The automation shall allow a logistician or commander to input their knowledge of troop speed, given the expected impact of casualties and the weight of supplies to be transported. This information will be used in the calculation of the fuel expenditure.

[3] Joe, J., O'Hara, J., Medema, H., & Oxstrand, J. (2014, June).

[18] Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. Advances in human performance and cognitive engineering research, 2, 1-12.

[74] Scott, R., Roth, E., Truxler, R., Ostwald, J., & Wampler, J. (2009, October). Techniques for effective collaborative automation for air mission replanning. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 53, No. 4, pp. 202-206). Sage CA: Los Angeles, CA: SAGE Publications.

CTA Methods focused on HMT

The first eight HMT themes are ripe for for exploration via CTA with subject matter experts (SMEs). We conducted preliminary work to develop a CTA methodology to address HMT. We identified building blocks for knowledge elicitation techniques and developed a technique to tailor the scope and focus of a knowledge elicitation to elicit critical HMT requirements.

After a survey of CTA methods, we concluded that two knowledge elicitation techniques are ripe for modifications that would enable addressing the eight HMT requirements themes which require SME input. The Critical Decision Method (Hoffman, Crandall, & Shadbolt, 1998) can capture expert insights on how they and their teammates performed tasks in past situations (either with or without automation) and can be used to follow the thread of how a situation would be different with automation. The Knowledge Audit (Militello & Hutton, 1998), a survey method of elicitation, could be used to probe specific cognitive elements or themes and may be well-suited to envisioned systems. We plan to use the Critical Decision Method and the Knowledge Audit as starting points but to tailor them to address the HMT themes. For exmaple, the Big Picture probe in the Knowledge Audit could be modified. Instead of asking about the major elements of the situation that you need to know and keep track of, the question could ask about the major elements of "what the automation is doing," that need to be known to understand the situation. This would address Observability and could lead to follow-up questions on Calibrated Trust. A modified Knowledge Audit paired with the Critical Decision Method could allow researchers to elicit information to tailor the requirements and characterize the HMT interactions and the challenges.

Addressing eight HMT themes may be too large a scope for a single CTA interview. To help focus the interview, we are developing a HMT Triage technique. The idea is to focus on the themes that that warrant the most attention. A theme may warrant attention because it is prevalent (e.g., the automation performs numerous calculations and assessments and there is a question about how many of the assumptions and information sources should be observable to the human), it is challenging (e.g., to enhance common ground, the automation needs just in time information from the human but system developers are reluctant to distract the human from his or her main task), or because it is critical to success (e.g., at different points in a mission, the unmanned aerial vehicle operator needs to direct his or her attention to different pieces of information in order to meet dynamically changing surveillance goals). Some of the triage questions are technology based. For example, “Is the automation in motion (actively controlling something in the environment, such as where a telescope is pointing) or at rest (operating virtually, such as a planning system)?” The implication is that autonomy in motion may warrant more focus on Observability so the human can see the effects on the object being controlled and Directability because the human may need to intervene. Other triage questions are designed to help the interviewee determine if the theme is likely to be applicable and a challenge. For Common Ground, the interviewee could ask if beliefs and assumptions are likely to change during a mission. The goal of the HMT Triage is narrow the focus to 3-5 HMT themes. This is not to say that other themes are not important, it simply acts as a screening technique to focus on the areas most in need of cognitive support. The questions in the HMT Triage can be used prior to an interview if the interviewing team has access to background material, SMEs, or system developers than can help answer questions. Otherwise, it can be used with a SME at the beginning of an interview.

CONCLUSION

As the prevalence of autonomous and automated components of military systems increases, issues with HMT often surface and can impact trust in technology and create a barrier to operational acceptance. This effort bridges the gap from research to technology development to better enable systems that synergize the joint strengths of humans and technology towards asymmetric advantages. These research efforts are critical for complex systems to go beyond decision aiding to achieve true teaming in which the joint performance of human and machine is greater than either alone.

Our next step is to develop a set of tailored requirements in two domains. To do this, we will tailor CTA methods to specifically address the themes in HMT. We will document the CTA methods used and develop a “how to” guide for knowledge elicitiation with experts and using the results to tailor cognitive interface requirements for HMT. The result will be a repeatable method that development teams can apply to ensure their systems support human-automation teaming.

REFERENCES

Bradshaw, J. M., Acquisti, A., Allen, J., Breedy, M. R., Bunch, L., Chambers, N., & Johnson, M. (2004, March). Teamwork-centered autonomy for extended human-agent interaction in space applications. In *AAAI 2004 Spring Symposium*, 22-24.

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, 57(3), 407–434.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 40(2), 254-276.

Joe, J., O’Hara, J., Medema, H., & Oxstrand, J. (2014, June). Identifying requirements for effective human-automation teamwork. In *Proceedings of the 12th international conference on probabilistic safety assessment and management* (PSAM 12, Paper# 371),(INL/CON-14-31340).

Johnson, M., Bradshaw, J. M., Hoffman, R. R., Feltovich, P. J., & Woods, D. D. (2014). Seven Cardinal Virtues of Human-Machine Teamwork: Examples from the DARPA Robotic Challenge. *IEEE Intelligent Systems*, 74-80.

Johnson, M., Shrewsbury, B., Bertrand, S., Wu, T., Duran, D., Floyd, M., et al. (2015). Team IHMC's Lessons Learned from the DARPA Robotics Challenge Trials. *Journal of Field Robotics,* 32(2), 192–208.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational simulation*, 53.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.

Lyons, J. B. (2013). Being Transparent about Transparency: A Model for Human-Robot Interaction. In *AAAI Spring Symposium Series*.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, 41(11), 1618-1641.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, 39(2), 230–253.

Sarter, NB, Woods, DD, & Billings, C. (1997). Automation Surprises. In G. Salvendy (Ed), *Handbook of human factors and ergonomics,* 2nd Edition. Wiley.

Woods, D., & Hollnagel, E. (2006). *Joint Cognitive Systems*. CRC Press.

Wiener, E.L. (1989). Human factors of advanced technology ("glass cockpit") transport aircraft. (NASA Contractor Report No. 177528). Moffett Field, CA: NASA Ames Research Center.

Identifying the Cognitive Demands on Experts' Decision Making in Liver Transplantation

Gala MOROZOVA[a], Amanda MARTINDALE[a], and Ian CURRIE[b]
[a] *Human Performance Science research group, The University of Edinburgh*
[b] *Transplant Unit, Royal Infirmary of Edinburgh, NHS Lothian*

ABSTRACT

Introduction. Cognitive task analysis (CTA) has recently gained the attention of surgical educators and the present study is investigating the cognitive demands of the Liver Transplantation procedure. **Methods.** In-depth interviews, following the Applied Cognitive Task Analysis protocol with four consultant transplant surgeons. **Results.** Eleven elements that show evidence of significant cognitive demands were extracted across the dataset. **Conclusion.** This study begins to reveal the origin and contents of transplant surgeons' decision-making expertise. Applying CTA techniques to this domain is an essential step to modernisation of surgical training and possesses value for both decision-making researchers and medical practitioners.

KEYWORDS

Expertise; surgery; applied cognitive task analysis; expert performance.

INTRODUCTION

Surgery is an increasingly complex performance domain, where decision making skills are of paramount importance (Alderson, 2010; Cuschieri, Francis, Crosby, & Hanna, 2001; Yates & Tschirhart, 2006). Several studies have focused on expertise acquisition and examined surgical trainees to determine correlates of surgical performance (e.g., Francis, Hanna, Cresswell, Carter, & Cuschieri, 2001; Wanzel et al., 2003). The results showed that innate technical abilities (e.g. steady hand, visuo-spatial ability) may help young surgeons to obtain surgical skills more quickly; however, it is experience and competent judgment that makes a difference to surgical performance (Norman, Eva, Brooks, & Hamstra, 2006; Smink et al., 2012). Previous research has shown that senior surgeons rank decision making and cognitive abilities as the most important non-technical skills for a surgical trainee (Cuschieri et al., 2001; Jacklin, Sevdalis, Darzi, & Vincent, 2008). Despite its importance, decision making receives little attention in surgical training models (Flin et al., 2007; Jaffer, Bednarz, Challacombe, & Sriprasad, 2009). A large portion of surgical education is based on assisting and observing more experienced colleagues, however, as senior staff develop expertise, they automate their procedural knowledge, making it difficult to articulate the steps taken in their decision-making process (Jaffer et al., 2009; Smink et al., 2012).

Clark, Pugh, Yates, Inaba, Green and Sullivan (2012) compared 3 methods for capturing surgeons' descriptions of how to perform a complex task and found that when experts were asked to free-recall the procedure they unintentionally omitted almost 70% of the information that novices need to successfully perform a task. In their study, they found that interviews following Cognitive Task Analysis (CTA) methodology were able to capture more decision and action steps from expert surgeons than unaided free-recall methods. Their findings complement previous research, suggesting that most forms of CTA show substantial benefits with respect to the accuracy and completeness of data obtained (Clark et al., 2012; Smink et al., 2012; Tofel-Grehl & Feldon, 2013). Moreover, it has also been shown in the literature that performance improvements in training for a number of surgical procedures can be attributed to CTA-based instruction (Clark et al., 2012; Sullivan et al., 2008; Tofel-Grehl & Feldon, 2013; Wingfield, Kulendran, Chow, Nehme, & Purkayastha, 2015). Additionally, research suggests that CTA-based instruction can increase the learning curve and accelerate the acquisition of expertise among trainees (Clark et al., 2012). Understanding the cognitive demands underpinning surgical decision making is necessary to ensure that training can prepare young surgeons to meet the increasing demands of their profession with flexibility and innovation (Cristancho, Vanstone, Lingard, LeBel, & Ott, 2013; Flin et al., 2007). This study, therefore, adopts a CTA methodology to improve the understanding of decision making expertise in transplant surgery and answer the following research question – what are the cognitive demands on experts' decision making in liver transplantation?

METHODS

A qualitative approach was chosen to collect rich data and capture high quality descriptions of surgeons' expertise. The Applied Cognitive Task Analysis (ACTA; Militello, Hutton, Pliske, Knight, & Klein, 1997) technique was adopted to collect information about experts' decision making processes. This method was chosen over other CTA techniques for a number of reasons. First of all, it comprises a combination of different techniques that complement each other and elaborate on different aspects of expertise. Secondly, study by Militello and Hutton (1998) showed that it requires relatively little prior training for the researcher and detailed instructional materials are available. Furthermore, ACTA has been used in NDM research with experts from a variety of knowledge domains and indicated high levels of validity and reliability in terms of its ability to generate relative data across participants (McAndrew & Gore, 2013; Militello & Hutton, 1998).

Participants

Purposeful convenience sampling was used to recruit four expert transplant surgeons (1 female, 3 male) from the Edinburgh Royal Infirmary. The number of participants was chosen according to Militello and Hutton's (1998) suggestion that three to five subject matter experts usually exhaust the domain of analysis. Professional occupation, expertise level, and willingness to participate in a study served as selection criteria. Recognition by fellow colleagues and years of experience were used as indicators of the expertise level. These criteria were chosen in accordance with ACTA's methodological recommendation and were also in line with Hoffman's (1998) expertise model, suggesting that all consultants can be considered experts.

At the time of data collection participants held the position of Consultant Transplant Surgeon performing both retrieval and transplant surgeries for liver, kidneys and pancreas. All participants had worked in healthcare for a minimum of 16 years (mean = 27,5 years) and had acquired a minimum of 4 years (mean = 13,7) of experience within their current position.

Materials

ACTA instructional materials and Job Aids were used for this study (Militello et al., 1997). These in-depth interviews utilise a combination of knowledge elicitation techniques to uncover different elements of expertise, and consist of three stages: Task Diagram, Knowledge Audit and Simulation Interview.

The first stage of the interview, task diagram, prompts the surgeon to give a broad overview of the task and to indicate the difficult cognitive elements. The interview opened with the question: "Think about what you do when you perform a liver transplant surgery. Can you break this task down into between three and six steps?" (Militello et al., 1997). After the diagram is drawn, the experts were asked to identify areas of the task that demand complex cognitive skills. In this study 'Liver Transplantation' was chosen for analysis, as a cognitively challenging task in which all participants have expertise.

The 'Knowledge audit' builds upon the information received in the first stage and uncovers different elements of expertise using a set of 8 probes (Militello et al., 1997). This technique allows the researcher to elicit examples of cognitive skills, detailed information about the selected task, and to contrast expert and novice performance.
Although, a simulation scenario was developed for this study due to time constraints and limited availability of experts it wasn't possible to conduct this stage with all participants and therefore it was omitted. A study by McAndrew and Gore (2013) examined financial traders' decision making using only Stage 1 and Stage 2 of the ACTA protocol and showed that it is sufficient to identify cognitive demands and compile practical recommendations.

RESULTS

To demonstrate the richness and complexity of qualitative data generated, a number of illustrative examples are presented to outline the results derived from Stages 1 and 2 of ACTA. Finally, extracts from an overall cognitive demands table are presented and discussed to summarise the findings of the study.

Stage 1: Task Diagram

The number of steps and level of detail varied among participants. The Combined Task Diagram (see Figure 1) is presented to summarise steps identified across the sample. This task diagram provides a broad overview of the intraoperative phase of liver transplant surgery and consists of five steps.

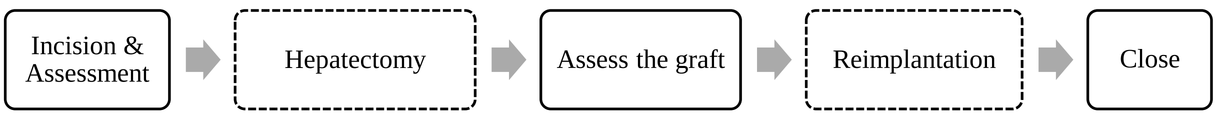


Figure 1. Combined task diagram

Participants, in one form or the other, said that all of its steps require a lot of cognitive skills. Although some of the participants identified more than two cognitively challenging elements, Hepatectomy and Reimplantation were highlighted as cognitively challenging by all participants.

Stage 2: Knowledge Audit

Table 1 provides an extract from Transplant Surgeon B's knowledge audit. This probe from the ACTA protocol elicits the overarching elements that form the 'big picture' for the liver transplantation task. Surgeons were asked to give an example and name major elements they have to know and keep track of. Four components were identified – estimated difficulty of operation, timing, patient characteristics, and quality of the graft. As Transplant Surgeon B explained, these are the elements that a surgeon needs to be aware of throughout the entire process, as they will affect all decisions that one makes. He noted that it is difficult for novices to understand the whole situation and maintain a big picture view. This is due to novices often focusing solely on the part that they are trying to learn at the moment, and not on the entire operation.

Table 1. Knowledge Audit Table Illustrative Example Transplant Surgeon B

Example	Cues & Strategies	Why Difficult?
Big picture... · Difficulty of operation · Timing · Patient · Quality of the organ	Patient characteristics; Time available; Quality of the organ; Difficulty of the operation; Operating staff (seniority, number, fatigue) · Depending on these factors decide how to approach the operation · Accomplish hepatectomy safely (without too much blood loss; instability in the patient) · Start implantation phase when the patient is stable · Accomplish both in reasonable time (no excessive ischemia time, but not in a rushed manner)	· Focused on the part that they are trying to learn at the moment, not on entire operation · Struggle to anticipate the difficulty as they haven't seen a lot of cases

Cognitive Demands

To integrate data drawn across the four task diagrams and knowledge audits from experts and summarise the results of the study a cognitive demands table (see Table 2) was compiled. Overall, eleven elements that show evidence of significant cognitive demands were extracted across the dataset: (1) Anticipate difficulty of the operation and choose how to approach it; (2) Assemble appropriate team & equipment; (3) Decide if portacaval shunt is feasible and necessary; (4) Complete hepatectomy in reasonable time; (5) Choose appropriate technique and pace for dissection; (6) Spot abnormalities in the liver graft; (7) Decide what to do if patient becomes unstable after perfusion; (8) Control bleeding; (9) Assess appearance of the liver; (10) Decide if blood supply for the graft is adequate; (11) Self-monitoring. There were no contradicting themes and examples across the sample. The cognitive demands which emerges had common elements and should not be considered on their own, but rather as a whole, to provide a complete picture of cognitive challenges in transplant surgery.

Cognitive Element	Why Difficult?	Common Errors	Cues & Strategies Used
(2) Assemble appropriate team & equipment	• Don't know and therefore cannot assess the team • Don't realise how difficult operation will be • Want to acquire more experience, therefore are reluctant to delegate • Not used to having 100% responsibility • Don't want to get bad reputation by demanding more senior assistants (e.g. scrub nurses)	• Blaming yourself for what is a team failure • Not realising when you need more senior/additional assistant • Losing time and concentration waiting for additional equipment/ staff	• How sick is the patient? • What complications to expect • Who is available (fatigue) • Equipment (do you need anything extra) • Make sure everything is ready before the operation begins, that you have the right team, right equipment
(4) Complete hepatectomy in reasonable time	• Difficult to judge the time during the operation • Confidence, don't want to ask for help • Might not have technical skills to complete hepatectomy safely and quickly • Don't know surgeons responsible for retrieval	• Not realising how long the operation will take given the 'factors' • Not asking for help • Asking for help when a lot of time is consumed • Wasting too much time and compromising graft's quality	• Familiarity of the surgeon responsible for the retrieval • Condition of the patient and donor liver • Anticipated difficulty of the hepatectomy • Start the operation as early as possible to minimise ischemic time • If confident in the retrieval surgeon and graft's quality, start before the liver has arrived in the building • Ask for a senior/additional assistant, if you expect a difficult hepatectomy

Table 2. Extract from the Cognitive Demands Table

DISCUSSION

The results of this study make a unique contribution to the literature through investigation of the cognitive demands of liver transplantation and support previous research on decision-making expertise in surgery. The ability to anticipate operation’s difficulty and foresee complications, which experts often referred to, as well as the majority of the cognitive demands identified, are closely related to the situation awareness concept proposed by Endsley (1997). Understanding the situation and being aware of patient and donor characteristics coupled with the proper assessment of one’s team and one’s own capabilities were found to be crucial for making effective decisions. This supports previous research suggesting that situation awareness is one of the driving factors in the decision making process in surgery and other performance domains (Cuschieri et al., 2001; Endsley, 1997; Flin et al., 2007; Yule, Flin, Paterson-Brown, & Maran, 2006). Strategies reported by experts in this study also fall into Cristancho and colleagues’ (2013) model of surgical decision-making, however, more research is needed to fully test this model.

This study has several implications for educators and surgeons themselves. Educators could use these results to enrich existing training programmes with expertise-based knowledge. Causer, Barach and Williams (2014) reviewed how medical education can benefit from the systematic use of the expert performance approach as a framework for measuring and enhancing clinical practice, and came to a conclusion that in order to optimise the training of medical professionals, both instructional materials and training approaches have to be guided by empirical research from the learning and cognitive sciences. In terms of application of the results to surgical education and training, expert surgeons could use the Cognitive Demands Table and task diagrams to focus more on articulating their decision making while performing cognitively challenging steps of the operation for teaching purposes (e.g. cognitive apprenticeship). Furthermore, SMEs noted that the materials generated in this study would be particularly useful for newly appointed Consultant Surgeons, as they enter a role which puts the responsibility for the operation on their shoulders for the first time. The cognitive demands and strategies described in the Cognitive Demands Table could help maintain awareness of the key elements and big picture of the procedure. Future work could focus on translating these materials into a checklist for use immediately prior to the procedure, to ensure all necessary preparations are in place and serve as a reminder of the key elements to keep track of during the operation.

Limitations

This research has some important limitations that have to be taken into consideration. First of all, despite two decades of empirical enquiry NDM still requires further theoretical and methodological refining to achieve more advanced level in studies of expertise in real-life settings. It also needs to be noted that one interview structure does not always suit all participants. Although a great volume of relevant data was generated following the ACTA methodology, it was sometimes difficult for experts to recall a specific situation because, as surgeons mentioned, cases in which they managed to avoid complications using expertise didn’t register in their memory.

Although it is worth noting, that the Simulation Interview method was omitted in this study, which could have addressed this issue. However, there are certain challenges associated with using the Simulation Interview part of ACTA. First of all, it takes a lot of time from both the research team and SMEs to create and pilot appropriate simulation materials and then to schedule and conduct additional 1,5 hours of interviews. Secondly, it is hard to estimate the amount of detail that needs to be included in the scenario to avoid purely hypothetical answers, as many decision and actions will depend on the context (e.g., resources available, team members experience, patient characteristics). Perhaps, complimenting Stages 1 and 2 of ACTA protocol with the analysis of documentation or observations would be a more feasible solution.

Another limitation of this study is generalizability, as it only includes four participants from a single institution. This is consistent with ACTA recommendations (Militello et al., 1997) and CTA literature in general (Schraagen, 2006; Smink et al., 2012; Sullivan et al., 2008; Tofel-Grehl & Feldon, 2013). Nonetheless, these surgeons all practice at the same hospital and local factors may generate similarity in opinion and techniques. This could explain why no contradicting descriptions of cognitive demands were found among participants. Expanding the sample to a national group of experts might yield more generalizable results. However, chosen methods are consistent with other published CTA studies and may, in fact, present greater value for implementation at the Edinburgh Royal Infirmary as its findings are generated from participants whose decision making is already framed by the organisational constraints of this hospital.

Future Research

Despite the limitations, this study creates opportunities for further research on expertise in general and in transplant surgery in particular. First of all, as this study was the first documenting cognitive demands on decision making during liver transplant surgery, the cognitive steps and demands identified must be examined in more detail, to ensure accuracy and completeness of these findings. Another important direction that future researchers and educators should take is to analyse the current teaching programme and find opportunities to integrate experience-based knowledge and expert performance approach into surgical training.

Participants often mentioned the importance of situation awareness, along with teamwork factors, confidence, and coping skills, in overcoming cognitive challenges of their profession. These influence and implications should be investigated further in order to enhance surgeons' performance and nurture expertise among junior staff.

CONCLUSION

The main objective of the present study was to identify the cognitive demands on expert decision making in transplant surgery. Conducting a field study of expertise using a detailed cognitive task analysis, this research begins to reveal the origin and contents of transplant surgeons' decision-making expertise. Examples and findings presented here illustrate how experts use cues and strategies to overcome cognitive challenges during liver transplantation. Although these findings cannot be contrast with similar research, as there is no literature directly investigating cognitive demands in this discipline, results of this study relate to the previous research on decision making expertise in surgery and possess value for both decision making researchers and medical practitioners.

REFERENCES

Alderson, D. (2010). Developing expertise in surgery. *Medical teacher, 32*(10), 830-836.

Causer, J., Barach, P., & Williams, A. M. (2014). Expertise in medicine: using the expert performance approach to improve simulation training. *Medical education, 48*(2), 115-123.

Clark, R. E., Pugh, C. M., Yates, K. A., Inaba, K., Green, D. J., & Sullivan, M. E. (2012). The use of cognitive task analysis to improve instructional descriptions of procedures. *Journal of Surgical Research, 173*(1), 37-42.

Cristancho, S. M., Vanstone, M., Lingard, L., LeBel, M.-E., & Ott, M. (2013). When surgeons face intraoperative challenges: a naturalistic model of surgical decision making. *The American Journal of Surgery, 205*(2), 156-162.

Cuschieri, A., Francis, N., Crosby, J., & Hanna, G. B. (2001). What do master surgeons think of surgical competence and revalidation? *The American Journal of Surgery, 182*(2), 110-116.

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making.

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., & Youngson, G. (2007). Teaching surgeons about non-technical skills. *The Surgeon, 5*(2), 86-89.

Francis, N. K., Hanna, G. B., Cresswell, A. B., Carter, F. J., & Cuschieri, A. (2001). The performance of master surgeons on standard aptitude testing. *The American Journal of Surgery, 182*(1), 30-33

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology *Exploring expertise* (pp. 81-100): Springer.

Jacklin, R., Sevdalis, N., Darzi, A., & Vincent, C. (2008). Mapping surgical practice decision making: an interview study to evaluate decisions in surgical care. *The American Journal of Surgery, 195*(5), 689-696.

Jaffer, A., Bednarz, B., Challacombe, B., & Sriprasad, S. (2009). The assessment of surgical competency in the UK. *International Journal of Surgery, 7*(1), 12-15.

McAndrew, C., & Gore, J. (2013). Understanding Preferences in Experience-Based Choice A Study of Cognition in the "Wild". *Journal of Cognitive Engineering and Decision Making, 7*(2), 179-197.

Militello, L. G., & Hutton, R. J. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

Militello, L. G., Hutton, R. J., Pliske, R. M., Knight, B. J., & Klein, G. (1997). Applied cognitive task analysis (ACTA) methodology: *DTIC Document.*

Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. *The Cambridge handbook of expertise and expert performance, 2006*, 339-353.

Schraagen, J. (2006). Task analysis. *The Cambridge handbook of expertise and expert performance*, 185-201.

Smink, D. S., Peyre, S. E., Soybel, D. I., Tavakkolizadeh, A., Vernon, A. H., & Anastakis, D. J. (2012). Utilization of a cognitive task analysis for laparoscopic appendectomy to identify differentiated intraoperative teaching objectives. *The American Journal of Surgery, 203*(4), 540-545.

Sullivan, M. E., Ortega, A., Wasserberg, N., Kaufman, H., Nyquist, J., & Clark, R. (2008). Assessing the teaching of procedural skills: can cognitive task analysis add to our traditional teaching methods? *The American Journal of Surgery, 195*(1), 20-23.

Tofel-Grehl, C., & Feldon, D. F. (2013). Cognitive Task Analysis–Based Training A Meta-Analysis of Studies. *Journal of Cognitive Engineering and Decision Making, 7*(3), 293-304.

Wanzel, K. R., Hamstra, S. J., Caminiti, M. F., Anastakis, D. J., Grober, E. D., & Reznick, R. K. (2003). Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery, 134*(5), 750-757.

Wingfield, L. R., Kulendran, M., Chow, A., Nehme, J., & Purkayastha, S. (2015). Cognitive Task Analysis Bringing Olympic Athlete Style Training to Surgical Education. *Surgical innovation, 22*(4), 406-417.

Yates, J. F., & Tschirhart, M. D. (2006). Decision-making expertise. *The Cambridge handbook of expertise and expert performance*, 421-438.

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006). Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery, 139*(2), 140-149.

Cognitive Skills Training for Frontline Social Workers: A Pilot Study

Emily NEWSOME[a] and Gary KLEIN[a]
[a]*ShadowBox LLC*

ABSTRACT
This paper summarizes an initial effort to design and deliver cognitive skills training to frontline social workers. The characteristics of social work expertise are not widely agreed upon, but investigators chose to design a cognitive training program with the goal of shifting the mindsets of social workers from a procedural to a problem-solving mindset. Investigators designed a pilot training intervention based on the ShadowBox approach to cognitive skills training (Klein & Borders, 2016). Investigators pilot tested the intervention at three public child welfare agencies across the country. Preliminary results based on reactions of participants indicate that participants found scenarios to be realistic and discussions to prove beneficial. Areas of improvement include the realism of expert feedback.

KEYWORDS
Decision making; education and training; social work; expertise.

INTRODUCTION

Frontline child protective workers, like practitioners in other high-pressure domains, must make life-altering decisions while managing time pressure and uncertainty. Their primary obligation is to investigate allegations of child abuse and neglect, which involves such cognitive skills as assessing a child's physical environment, evaluating parental capacities, and investigating specific incidents of abuse and neglect. Their work depends heavily on anticipatory thinking — anticipating the potential for harm if the child is allowed to remain with its family. Entire investigations must take place within mandated timeframes, and caseworkers usually carry many cases at a time, creating a fast-paced work environment.

It is difficult to aquire expertise as a child protection worker. The field is plagued by chronic understaffing and high turnover, so many are not in the job long enough to develop expertise. For those who build tenure as child protection workers, feedback is often inconsistent or even absent at times. Often the repercussions of a specific action (or inaction) are never known. Even in cases that are tracked over time, it is often unclear what the best course of action in a specific situation might have been. In an attempt to encourage more consistent decisions, and help inexperienced workers get up to speed more quickly, the field of child welfare has introduced the use of structuralized, survey-like tools to evaluate the safety and risk of children (Schwalbe, 2004). These assessment tools are based on the probabilistic analysis of environmental factors that are associated with detrimental outcomes. A worker must determine which factors are present in a given case, input them into the tool, and the tool's output will determine the agency's level of response to the family. While in theory, these tools promote consistency and efficiency, the reality is that questions remain about the effects of standardized tools on decision making, specifically on the role they play in developing expertise. Research in other domains indicates that over-reliance on procedures and tools like these may even inhibit the development of expertise (Hoffman, Ward, Feltovich, DiBello, Fiore & Andrews, 2013). Expertise is developed through routine practice; because of the high turnover, there is little room for the development of expertise. Thus, it is important to implement strategies for promoting the development of expertise in child welfare. This paper describes an initial effort to design such a program for promoting the development of expertise in this domain.

Expertise in child welfare

There is little consensus in the child welfare domain as to how expertise is defined in frontline child welfare. Many public child welfare organizations value and emphasize compliance with specific rules and procedures, such as completing investigations within mandated timeframes (Munro, 2011). However, the RPD model of expertise suggests that expertise is defined by an expert's ability to recognize patterns and mentally simulate potential responses (Klein, Calderwood & Clinton-Cirocco, 1986). Recent research based on this understanding of expertise has hypothesized that there is a cognitive difference in how highly skilled social workers and social workers of average skill approach their work (Newsome, Wright, Klein, Flory & Baker, 2015). Average social workers approach their work with a focus on following rules and procedures, which tends to guide their decision-making to be more rigid and procedurally focused. Highly skilled social workers approach their work with the goal of solving problems, which drives their decision-making to be more flexible, as well as family- and situation-specific. This approach posits that expertise in social work has more to do with quality of investigations rather than complying with mandates.

Expertise is developed through the formation of tacit knowledge and deepening of mental models. Cognitive transformation theory (Klein & Baxter, 2009) posits that in order to enrich one's mental model through training, flaws in current mental models must be challenged. Hoffman et al. (2013) suggest that one way to rapidly shift people from one mental model to another is through an accelerated expertise program. With the goal of shifting social workers' mindsets from procedurally-oriented to problem-solving-oriented, we implemented a pilot cognitive skills training program to promote the development of expertise.

ShadowBox training

For this effort, we employed a technique called ShadowBox, which is derived from lessons learned from research into cognitive skills training (Klein & Borders, 2016). ShadowBox training begins with presenting trainees with complex scenarios, which are often based on real life experiences of social workers. Embedded within these scenarios are four to six decision points, which ask trainees about their priorities, next steps, cues that stand out to them, or goals. Trainees then rank order a set of pre-determined options and write their reasons for their selections. After making their own decisions, trainees see the answers a panel of experts gave when they (the experts) worked through the scenario. Trainees compare both their rankings and their rationale for their decisions. Trainees are encouraged to identify insights about details the experts noticed, tasks experts prioritized, or the goals experts identified that trainees may have missed.

The following sections of this paper describe an effort to design and implement a ShadowBox-based cognitive skills training program for frontline social workers. We have just completed the pilot phase of this project; therefore, we describe our approach, anticipated results, and next steps.

PRACTICE INNOVATION

Our initial effort to develop a cognitive skills training program for child welfare practitioners has three main components: identifying scenario objectives, developing complex scenarios, and integrating facilitated discussion.

Identifying scenario objectives

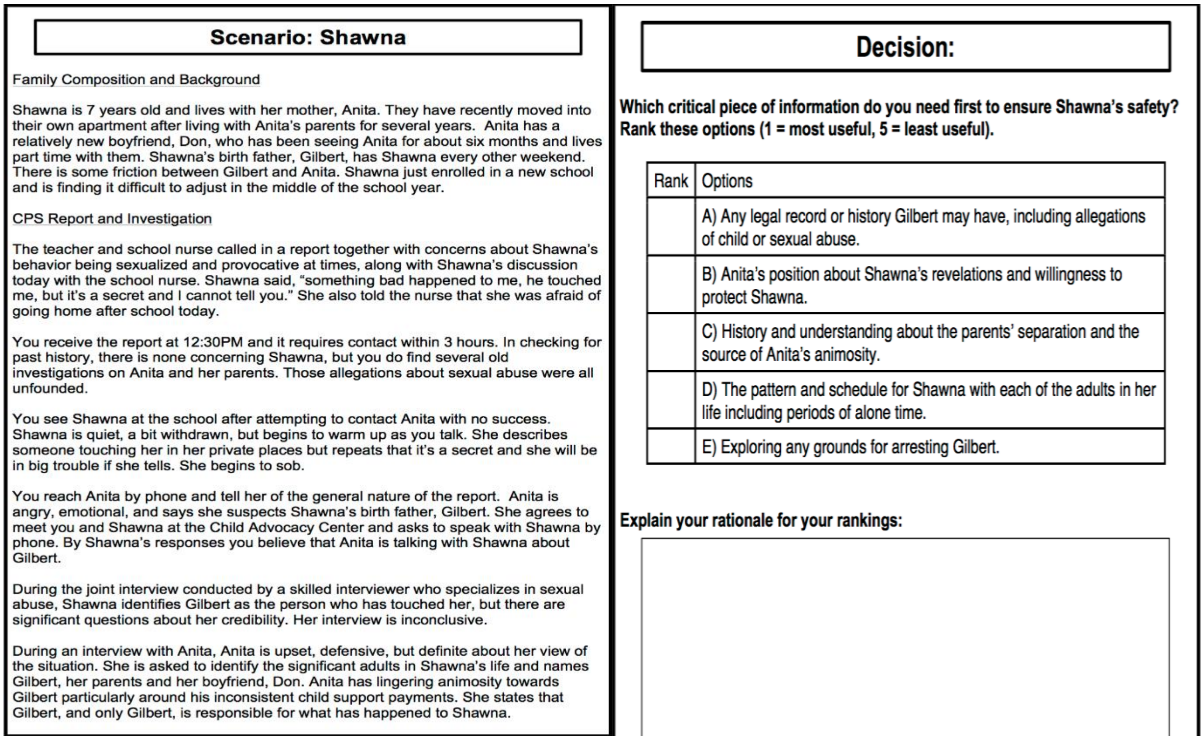
The overall goal of ShadowBox training is to shift mindsets. ShadowBox shifts mindsets by providing a space for trainees to have their core beliefs about decision making challenged (e.g. that following all rules and procedures will always lead to the best decisions for children and families). By comparing their own thought processes to those of a panel of experts, participants are also able to understand how experts approach a scenario -- what cues stand out to them, what their priorities are, and how they develop strategies for intervening on families. Over time, trainees will begin to think more like the experts and shift from a procedural to problem-solving mindset.

Developing complex scenarios

In this pilot project, the ShadowBox scenarios came from actual cases experienced by frontline social workers. We used the critical decision method, a form of cognitive task analysis, to elicit critical incidents from social workers (Crandall, Klein & Hoffman, 2006). This interview approach allowed us to probe the decisions made at different points in a worker's investigation, including how they evaluated various priorities, addressed competing goals, and developed strategies throughout the investigation. These decisions ultimately become decision points in ShadowBox scenarios.

We generated decision options by probing hypotheticals during our interviews – asking interviewees how newer colleagues would handle the situation. Additionally, we provided at least one option at each decision point that would appeal to someone with a "problem-solving" mindset, and another option that would appeal to someone with a "proceduralist" mindset. We hypothesize that options that would appeal to problem-solvers have to do with investigating all relevant safety concerns, rather than focusing exclusively on safety concerns specifically mentioned in the allegation. We hypothesized that options that would appeal to proceduralists are directly related to specific incidents or reported allegations of abuse, rather than the general well-being of children/families. Figure 1 provides an example scenario and decision point. In the sample decision point, a trainee with a proceduralist mindset may choose Option A or E, because they are a direct response to the reported concern, which is that the father is the abuser.

Option B may appeal to someone with a problem-sloving mindset, because it would involve ensuring the child's safety; if the mother shows willingness to protect Shawna, she will be safe no matter who the abuser is.
To date, we have generated eleven social work scenarios, which cover a range of social work cases, including allegations related to domestic violence, potential drug abuse, sexual abuse, mental health issues, neglect, and custody battles.



Scenario: Shawna

Family Composition and Background

Shawna is 7 years old and lives with her mother, Anita. They have recently moved into their own apartment after living with Anita's parents for several years. Anita has a relatively new boyfriend, Don, who has been seeing Anita for about six months and lives part time with them. Shawna's birth father, Gilbert, has Shawna every other weekend. There is some friction between Gilbert and Anita. Shawna just enrolled in a new school and is finding it difficult to adjust in the middle of the school year.

CPS Report and Investigation

The teacher and school nurse called in a report together with concerns about Shawna's behavior being sexualized and provocative at times, along with Shawna's discussion today with the school nurse. Shawna said, "something bad happened to me, he touched me, but it's a secret and I cannot tell you." She also told the nurse that she was afraid of going home after school today.

You receive the report at 12:30PM and it requires contact within 3 hours. In checking for past history, there is none concerning Shawna, but you do find several old investigations on Anita and her parents. Those allegations about sexual abuse were all unfounded.

You see Shawna at the school after attempting to contact Anita with no success. Shawna is quiet, a bit withdrawn, but begins to warm up as you talk. She describes someone touching her in her private places but repeats that it's a secret and she will be in big trouble if she tells. She begins to sob.

You reach Anita by phone and tell her of the general nature of the report. Anita is angry, emotional, and says she suspects Shawna's birth father, Gilbert. She agrees to meet you and Shawna at the Child Advocacy Center and asks to speak with Shawna by phone. By Shawna's responses you believe that Anita is talking with Shawna about Gilbert.

During the joint interview conducted by a skilled interviewer who specializes in sexual abuse, Shawna identifies Gilbert as the person who has touched her, but there are significant questions about her credibility. Her interview is inconclusive.

During an interview with Anita, Anita is upset, defensive, but definite about her view of the situation. She is asked to identify the significant adults in Shawna's life and names Gilbert, her parents and her boyfriend, Don. Anita has lingering animosity towards Gilbert particularly around his inconsistent child support payments. She states that Gilbert, and only Gilbert, is responsible for what has happened to Shawna.

Decision:

Which critical piece of information do you need first to ensure Shawna's safety? Rank these options (1 = most useful, 5 = least useful).

Rank	Options
	A) Any legal record or history Gilbert may have, including allegations of child or sexual abuse.
	B) Anita's position about Shawna's revelations and willingness to protect Shawna.
	C) History and understanding about the parents' separation and the source of Anita's animosity.
	D) The pattern and schedule for Shawna with each of the adults in her life including periods of alone time.
	E) Exploring any grounds for arresting Gilbert.

Explain your rationale for your rankings:

Figure 1. Sample scenario background and decision point.

Integrating facilitated discussions

Prior versions of ShadowBox training consisted of trainees completing scenarios independently, using web- or paper-based scenarios. This delivery format has many benefits, including allowing trainees to complete training on their own time. When designing a training intervention for child protective services, however, administrators strongly recommended that we include small group discussion as a core component of the training. In this domain, discussions among colleagues about their interpretations and points of view are highly valued as a means to maintain common ground, share lessons learned, and stay apprised of emerging sociocultural trends and changing regulations. Additionally, decisions involving the potential removal of children from the home are usually made as part of a team comprised of at least one supervisor and one worker. Thus, it was critical that we integrate discussions into our implementation plan.

PILOT TEST

We recently completed a pilot test of a ShadowBox training intervention at three different child welfare agencies. Two of these agencies were large, urban agencies, and the third was a smaller agency whose jurisdicition was suburban and rural. To encourage group discussion, participants were divided into groups of five to eight workers or supervisors. Prior to the training intervention, investigators trained facilitators about how to conduct ShadowBox-related discussions at each site. Trained facilitators at each of the three sites then led each group through training scenarios, and facilitated discussion at each decision point. The two larger agencies had five participant groups each, and held group sessions once per month. During each session, two scenarios were completed. The smaller agency had three groups, which met monthly for one hour and completed one scenario per meeting. Groups at all agencies completed six ShadowBox scenarios in total. Participants were asked to provide demographic data, as well as their perceptions of the utility ShadowBox training and feedback about the presentation of ShadowBox training after completing their final scenario.

Results

Table 1 summarizes responses to the evaluation survey completed by participants at each site. All questions were scored on a 5-point scale in which 1 was the most positive response and 5 was the least positive response, except for question 3 which was reverse scored. For questions 1, 2, 4, 5, 6, 7, a score lower than 2.5 was considered a positive response, scores between 2.5 and 3.5 were considered neutral, and scores above 3.5 were considered negative. Perhaps the most interesting finding is that participants benefitted from hearing others' perspectives during discussions (Q4) with an overall mean score of 1.87. Participants also rated scenarios positively. Specifically, they found scenarios to be realistic and relevant (Q1, m=1.98), as well as challenging and interesting (Q2, m=2.31). The evaluations also highlighted areas for improvement. Mean ratings for question 5 were in the neutral range (m=3.19), suggesting that the presentation of the expert model could be improved, and that the participants valued the group discussions more than the expert panel responses. Similarly, neutral response to Questions 3 and 6 suggest that there was variability in the perceived benefits of the training with regard to time investment (Q3, m=3.18) and preparedness (Q6, m=2.84); the participants at site 3 were less enthusiastic about the training than participants at sites 1 and 2.

Table 1. Mean ratings of ShadowBox effectiveness

Question	**Site 1 N=31 Mean (SD)**	**Site 2 N=17 Mean (SD)**	**Site 3 N=36 Mean (SD)**	**All Sites N=84 (Mean SD)**
Q1. These scenarios were realistic and relevant to frontline child protective work.	2.08 (0.98)	1.84 (0.60)	2.00 (0.94)	1.98 (0.86)
Q2. These scenarios were challenging and interesting to me	2.23 (0.91)	2.26 (0.93)	2.47 (0.72)	2.31 (0.86)
Q3. This training was too time-consuming for the benefit provided	3.40 (1.08)	3.05 (1.18)	3.00 (0.82)	3.18 (1.05)
Q4. I benefitted from hearing others' perspectives during discussions	1.77 (0.95)	1.95 (1.22)	1.94 (0.90)	1.87 (1.02)
Q5. The expert panel feedback provided me new insights.	3.19 (0.98)	3.11 (0.88)	3.29 (0.85)	3.19 (0.90)
Q6. I'm better prepared for complex child protection challenges after this training.	2.69 (0.93)	2.68 (0.95)	3.24 (0.83)	2.84 (0.93)
Q7. Overall, I benefitted from this training	2.12 (0.73)	2.37 (1.16)	2.82 (0.81)	2.39 (0.94)

Additional analyses will include looking at improvement on various ShadowBox performance metrics, such as how closely participants aligned with the experts, and whether they shifted from a "proceduralist" mindset to a "problem-solving" mindset. We will also look at participants' reactions to the training program as indicated by their responses to our post-intervention survey. Similar evaluations of ShadowBox have yielded promising results. For example, Klein & Borders (2016) found a 28% improvement in half-day training workshops with military personnel. We expect that since this intervention involved extended exposure to scenarios, these results can be replicated and even improved upon.

NEXT STEPS

Integrating lessons learned from this pilot study, we plan to conduct a full-scale evaluation of the ShadowBox training at two of the three pilot sites. We will modify specifics, such as frequency and duration of training, based on the analysis of our pilot data and feedback from facilitators. In the full-scale evaluation, we will include performance-related measures, including: .outcomes for workers, outcomes for systems, and outcomes for children. Specific outcomes of interest will include fewer repeat reports on open investigations, as well as more trust between workers and supervisors about the quality and consistency of their decision-making.

ACKNOWLEDGEMENTS

This research was funded by the Annie E. Casey Foundation. We thank them for their support but acknowledge that the findings and conclusions presented in this report ar those of the authors alone, and do not necessarily reflect the opinions of the Foundation. The authors would like to thank Gretchen Test, Amy Baker, Jan Flory, Pat Rideout, Stacey Gerber, John Mattingly, Tracey Feild, and Rodney Brittingham for their contributions to this effort. Additional thanks to Laura Militello and Joseph Borders for their reflections on the manuscript.

REFERENCES

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2013). *Accelerated expertise: Training for high proficiency in a complex world.* Psychology Press.

Klein, G., & Baxter, H. C. (2009). Cognitive transformation theory: Contrasting cognitive and behavioral learning. In D. Schmorrow, J. Cohn, & D. Nicholson (Eds.), *The PSI handbook of virtual environments for training and education: Developments for the military and beyond. Volume I: Learning, requirements and metrics* (pp. 50-65). Westport, CT: Praeger Security International.

Klein, G., & Borders, J. (2016). The ShadowBox approach to cognitive skills training: An empirical evaluation. *Journal of Cognitive Engineering and Decision Making, 10*(3), 268-280.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986, September). Rapid decision making on the fire ground. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 30*(6), 576-580. SAGE Publications.

Munro, E. (2011). *The Munro review of child protection: final report, a child-centered system* (Vol. 8062). London, UK: The Stationery Office.

Newsome, E., Wright, C., Klein, G., Flory, J., & Baker, A. (2015, September). Using the ShadowBox™ method to detect the 'Investigator'and 'Proceduralist' mindsets in frontline social workers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 816-820. SAGE Publications.

Schwalbe, C. (2004). Re-visioning risk assessment for human service decision making. *Children and Youth Services Review, 26*(6), 561-576.

Naturalistic Decision Making and Adaptive Expertise in Trainee Teacher Pedagogy

Stewart OLLIS
Fife Education

ABSTRACT
This paper explores the adoption of adaptive expertise, macro-cognition and principles of naturalistic decision making in the development of a trainee teacher over a two year period. Where traditional approaches to curriculum and pedagogical development focus on routine, rationalism, structure and certainty, the present study embraced uncertainty, emergence, complexity and improvisation required in the classroom. The study adopted a Dionysian/ rhizo-analytical case study approach to research design which embraced a multi-method framework. The whole study macro-analysis (2 years, year) was embedded with meso- (term, months, weeks) and micro- (days, lessons, episodes) studies. Initial findings suggest that developing teacher pedagogy and student learning under principles of adaptive expertise, macro-cognition and naturalistic decision making enhances depth of learning and understanding. These principles and the Dionysian model of expertise mirrored trainee teacher development and would support the 'complexity turn' and 'adaptive expertise' desired in curriculum and pedagogy.

KEYWORDS
Expertise; learning and training; education and training.

INTRODUCTION

The Scottish Government introduced the Curriculum for Excellence in 2008. As a constructivist design, the CfE was implemented to embrace the dynamic nature of learning ecology, uncertainty and the complexity of school and teacher practice. Indeed, the CfE had been a catalyst for a 'theoretical' complexity turn in how we understand learning, teaching and curriculum design. The University of Glasgow formulated 'Assessment is for Learning' (Hayward & Spencer, 2010) as a complexity framed model, the Scottish Primary School PE program committed to a non-linear pedagogy (Jess et al, 2012) and curriculum designers (e.g. Biesta, 2013; Priestley & Biesta, 2013) were forerunners in promoting the complexity turn on an international stage. The PGDE teacher training course itself gave lectures on the 'complexity turn' and promoted literature to advocate the adoption of complexity design in understand the various syllabus areas such as Mike Askew's *Transforming Mathematics* and Ovens et al's *Complexity Thinking in Physical Education*.

Literature on teaching and pedagogic development as 'adaptive expertise' is frequent (Baroody & Dowker, 2003) whilst the theory of 'deliberate practice' and 'visible learning' most common in the literature. Uncertainty is the new constant for teachers (Evans, 2013) and the importance of balancing structure and improvisation, and ability to adapt to the shifting needs of a class and lesson are clearly recognised. In addition, prominent readers such as John Hattie (2014) and Dylan William (2016) even adopt research such as Klein and Klein (1981) to highlight the necessity for not reducing teaching to a set of rules, as well as advocate the adoption of training methods based upon cognitive task analysis and naturalistic decision making.

The aim of the present work was to make sense of the CfE, associated literature and models of practice. More importantly, the trainee wished to embrace associated interventions in classroom practice during a two year trainee teacher period.

METHOD

The trainee teacher has completed a year long PGDE teacher training course and at the time of writing was near completion of the GTC probationer year at a Fife Primary School. The trainee teacher has to conduct personal reflection on teaching practice assisted by a personal supporter at his/her Primary School where his/her development is discussed during (at least) weekly meetings. The teacher also has to be observed on a series of lessons throughout the year. To assist in this process, the trainee teacher adopted a Dionysian/ rhizo-analytical case study approach (Coleman & Ringrose, 2013; Kidd, 2015; Sellers, 2013; Semetsky & Masny, 2013) to this process. The trainee teacher had embraced previous real world research using ethnography and action research as a main frame, embedded with multi-methods including quasi-experimental studies over various timescales. A rolling literature review process is inherent to the Dionysian/ rhizo-analytic design. The research methods were designed to balance the rational and planning of pedagogic practice with the emergent and dynamic nature of practice. Indeed, the research was a real world methodology committed to seeing the research field as a complex adaptive system.

FINDINGS

Findings suggested that teaching and educational management is fast paced and open with a commitment to enhancing practice highly evident. However, the traditional and rational models of practice and development of teacher practice are still dominant and the 'complexity turn' has faced what appears as a resistance or lack of attention in schools. Ethnographic dialogue suggests that the school environment is so fast paced and demanding that there is little space for a complexity turn to occur and resistance seems most evident at managerial/ leadership levels. It is suggested that present practice of high rationality and evidence based practice creates a glass ceiling (e.g. levels of competence and proficiency in the Dreyfus model of expertise) which requires an embracing of 'invisible learning' to break through and attain expertise and mastery which require the ability to make intuitive decisions and absorbed awareness.

The trainee teacher embraced the principles of adaptive expertise and complexity science to understand his/her development as well as his/her students. Contrary to the adoption of deliberate practice as the main frame for teacher development and 'visible learning', recent research suggests that the theory of deliberate practice only accounts for 4% of education expertise and 1% of professional development as an expert teacher (Macnamara et al, 2014). This discrepancy was identified in the real world experience of the trainee teacher. The present study resonated with the principles of the Dionysian model of expertise already adopted in Scottish PE (see figure 1). The model adopts many of the principles of naturalistic decision making and macro-cognition to ensure that we give recognition to both the 'streetlights and shadows' (visible and invisible learning), experiential framed learning and macro-cognition as a means of developing deliberate and non-deliberate mindfulness.

Figure 1. Dionysian Development: Complex adaptive model of expertise (Sproule, et al, 2011)

The trainee teacher continually returned to the principles of naturalistic decision making for addressing pedagogic issues which were deemed paradoxical. Commonalities were found in the more dynamic educational terminology of reviewed literature such as threshold concepts, liminality and interleaving which would be termed in the NDM world as critical transitions, leverage points and variable practice. However, there was clear indication that order and rationalism and the desire for visible and evidence based planning and development was dominant in working school practice and lesson design. The thinking of practitioners such as David Didau who claim learning is predominantly non-linear and invisible with a necessity to embrace complexity framed perspectives of learning (e.g. adopt the Robert Bjork understanding of learning), requires support with the work conducted in the field of NDM to date and was deemed as a catalyst for identifying shared issues.

CONCLUSION

The domains of educational governance, teaching and learning are environments where the capability to make decisions in uncertainty and demanding situations are necessary and theorists suggest are going to get greater.

"Learning is a messy, complicated business- a luminal process at the boundary between control and chaos" where there is a desire to "map the unmappable and plot pupils' journeys within subject domains: the threshold concept" (Didau, 2015, p159). Naturalistic decision making has researched the complexities of decision making in various fields and has much to offer educationalists in making the 'complexity turn' in practice and embracing the 'adaptive expertise' their field desires. The Dionysian/ rhizoanalysis methodologies and Dionysian Development model proved beneficial when researching complex models of expertise and long term naturalistic decision making studies in real world practice.

ACKNOWLEDGMENTS

The study gives gratitude and thanks to the Fife Education Probationer Team (led by Sheona Goodall) and Trainee Teacher Support/Mentor (Susan Mackenzie) for all the support and guidance over the last two years which made this paper feasible.

REFERENCES

Askew, M. (2016). Transforming Primary Mathematics. London, Routledge.

Baroody, A.J., & Dowker, A. (2003). The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise. Mahwah, NJ, taylor and Francis.

Biesta, G.J.J. (2013). The Beautiful Risk of Education. London, Paradigm.

Coleman, R. & Ringrose, J. (2013). Deleuze and Research Methodologies. Edinburgh University Press.

Crandall, B., Klein, G. & Hoffman, R.R. (2006). Working Minds. London, Bradford Books.

Didau, D. (2015). What if Everything You Knew About Education Was Wrong. Carmarthen, Wales, Crown House Publishing.

Dreyfus, S.E. (2004). The five-stage model of adult skill acquisition. Bulletin of Science Technology and Society, 24(3), 177-181.

Evans, L. (2013). The professional status of educational research: Professionalism and developmentalism in the twenty-first-century working life. British Journal of Educational Studies, 61(4), 471-490.

Hattie, J. & Yates, G. (2014). Visible Learning and the Science of How We Learn. London, Routledge.

Hayward, L. & Spencer, E. (2010). The complexities of change: formative assessment in Scotland. The Curriculum Journal, 21(2), 166-177.

Jess, M., Attencio, M., & Carse, N. (2012). Introducing conditions of complexity in the context of Scottish physical education. In A.Ovens, T.Hopper, & J.Butler (Eds.), Complexity Thinking in Physical Education. (pp 27-41). Abingdon.

Kidd, D. (2015). Becoming Mobius. Carmarthen, Wales, Independent Thinking Press.

Klein, H.A. & Klein, G.A. (1981). Perceptual/ cognitive analysis of proficient CPR performance. Paper presented at the Annual Meeting of the Midwestern Psychological Association, Detroit, MI.

Klein, G.A. (2011). Streetlights and Shadows. London, Bradford Books.

Macnamara, B., Hambrick, D., & Oswald, F. (2014). Deliberate practice and performance in music, games, sports, education and professions: A meta-analysis. Psychological Science, 1-11.

Ovens, A., Hopper, T. & Butler, J. (2012). Complexity Thinking in Physical Education, Abingdon.

Priestley, M. & Biesta, G. (2013). Reinventing the Curriculum. London, Bloomsbury.

Salas, E. & Klein, G. (2001). Linking Expertise and Naturalistic Decision Making. Mahwah, NJ, Lawrence Erlbaum.

Sellers, M. (2013). Young Children Becoming Curriculum. London, Routledge.

Semetsky, I. & Masny, D. (2013). Deleuze and Education. Edinburgh University Press.

Sproule, J., Ollis, S., Gray, S., Thorburn, M., Allison, P. and Harton, P. (2011). Promoting perseverance and chakllenge in physical education, Sport, Education and Society, 16(5), 665-684.

Stobart, G. (2014). The Expert Learner. London, Open University Press.

William, D. (2016). Leadership [for] Teacher Learning. Florida, USA, Learning Sciences International.

Critical Care Decision Making: A pilot study to explore and compare the decision making processes used by critical care and non-critical care doctors when referring patients for admission

Nicola POWER, Jacqueline BALDWIN[b], Nick PLUMMER[b], Shondipon LAHA[b]
[a]*Department of Psychology, Lancaster University, UK, Email: n.power3@lancaster.ac.uk*
[b]*Critical Care Medicine and Anaesthesia, Lancashire Teaching Hospital Trust NHS Foundation*

ABSTRACT
Decision-making regarding admission to critical care in UK hospitals is challenging. Not only does demand for beds exceed capacity, but the requirement to cover emergency admission creates pressure to build redundancy into the system. There are no clear guidelines to aid clinical decision-making, resulting in an over-reliance on professional judgement. Although experts are highly skilled decision makers, there is great variability in such decision-making, especially at the multi-speciality level wherein cognitive biases contribute to disagreement. This research is the first to explore multi-disciplinary decision making regarding critical care admission using the Critical Decision Method, interviewing non-critical care (n~12) and critical care doctors (n~12). This pilot research provides the foundation to the development of an intervention to improve multi-speciality decision-making.

KEYWORDS
Decision Making; Health; Team and Organizational Factors in Complex Work Environments; Sensemaking.

INTRODUCTION

Decision-making in busy hospital environments is fast paced, high-stakes and complex. This is especially true for emergency care, wherein the critical status of a patient is high. The critical care, or intensive care, unit is a self-contained area of a hospital with specially trained staff and equipment dedicated to the management and monitoring of patients with life-threatening conditions who need more support than can be offered on a general ward. Patients in critical care need organ support and a higher ratio of nursing, be this after an emergency, a major trauma, or after a large elective operation. Critical care units are therefore essential in supporting emergency and elective medical care, but are expensive to run due to equipment costs and nursing support (approx. £2,000 per day, per bed). National guidelines recommend that such units should run at 75% bed capacity, but the typical demand vastly exceeds this, reaching over 95% utilisation in many units nationwide.

Generally, admission to critical care should be reserved for patients who are likely to survive if admitted for care, but unlikely to survive, or will develop significant morbidity, if they are not admitted (Blanch, Abillama, Amin et al., 2016). However the capacity to apply such rules in the real-world is difficult, especially when these decisions rely on subjective clinicial judgements that can vary considerably, both within and between specialities. To put this into context, imagine being the consultant doctor for the critical care unit who is treating multiple sick patients with one bed remaining. It is a Saturday night, when unanticipated emergency cases are likely from the emergency department. You receive a referral from the ward to admit a geriatric patient in his 80s who has terminal cancer and has developed pneumonia. The consultant oncologist argues that the patient's life will be prolonged, despite having terminal cancer, if provided with critical care. Do you admit a geriatric patient who is terminally ill already? Or refuse admittance to save for the possibility of emergency cover ? Do you assess other patients to see if they have improved and risk downgrading them to a lower level of care, such as on a general ward with a lower nurse:patient ratio? Or choose to downgrade another patient who has shown few signs of improvement, but would be likely to die if critical care is removed? The ultimate responsibility for this decision lies with the clinician responsible for the critical care unit. How do you make this choice?

Although the above scenario is fictitious, it provides an illuminating example of the types of high-risk, ambiguous and complex choices that both admitting critical care physicians, and non-critical care doctors referring these patients, must make. Clinicians are faced with the unavoidable task of 'rationing' patient care, whereby treatment to selected patients must sometimes be witheld due uncontrollable limitations (Truog, Brock, Cook et al., 2006).

This creates a difficult set of decisions for clinicians deciding who should be referred and/or admitted to critical care. Critical care doctors are responsible for deciding who to *admit* to the unit and when to move patients off the unit. Meanwhile non-critical care doctors are responsible for *referring* patients from non-critical care wards to the unit. Both sets of doctors are required to do this in the absence of clear guidelines on how to make patient referral or admission decisions (Pattison & O'Gara, 2014), therefore they are likely to base such choices on subjective judgements that, depending upon their clincial experience, can be weighted by their own professional cognitive biases and perspectives. Yet, despite the importance of these decisions, psychological research to explore and compare multi-speciality cognitive processing is limited. This pilot research will be the first to address this gap, by qualitatively unpacking the cognitive processes that influence multi-speciality decision making in critical care contexts.

Decision Making in Complex Environments

A lack of guidelines when working in ambiguous, pressurised and risky contexts can derail decision making. In emergency response contexts, it was found that decision inertia, defined as the redundant deliberation of choice for no cognitive gain, derailed decision-making when commanders traded off 'save life' and 'prevent harm' goals (Power & Alison, 2017). Due to the high-risk context, commanders were aware that actions to 'save life' might counterintuitively risk harm (e.g. to emergency responders) and vice versa, leading commanders to redundantly deliberate about the potential negative short- and long-term consequences that might arise, rather than satisficing to a 'least worst' choice (Power & Alison, 2017). When making decisions in complex and ambiguous environments, individuals rely on their cognitive biases. In their study of 'rationing' in critical care admissions, Trueg et al., (2006) warned that reliance on clinical judgements to make these decisions was risky, due to the potential for using irrational and prejudiced cognitive biases. Indeed, a study comparing Australian and New Zealand critical care doctors' decision-making showed cultural differences in their response to critical care vignettes; New Zealand doctors were more selective and liklely to refuse admittance than their Australian counterparts (Young & Arnold, 2010). A lack of consensus on whether to admit patients to critical care can be further problematic when considering the perspectives of clinicians from different speciliaties outside of critical care. In their analyses of multi-disciplinary decision making in emergency response contexts, Power and Alison (2016) identified that, although responders from different agencies *assumed* common goals, their self-reported goals were highly inconsistent and role specific. Individuals interpretted the situation through their own role-specific and cognitively biased lens. Furthermore, they were unaware of these differences and assumed that colleagues shared the same goals, risking duplicated efforts and contradictory behaviours. This research will explore multi-speciality critical care decision-making to identify whether similar cognitive biases exist in this equally high-risk and complex environment.

What do we know about critical care decision making ?

There is limited published research exploring multi-disciplinary critical care referral decisions in the UK. Much of the research has focused on decisions about *removing* patients from critical care (i.e., end-of-life decision making; McAndrew & Leske, 2015), rather than *admitting* patients to the service. Moreover, the existence of the NHS means that research in non-UK settings is incomparable; as care in the UK is funded by the NHS, which creates larger bed and fiscal pressures compared to non-UK hospitals (Rhodes, Ferdinande, Flaatten, Guidet, Metnitz & Moreno, 2012). Critical care doctors in the UK are the 'gatekeepers' who decide when *not* to admit patients, which contrasts to privately funded healthcare systems where patients/family have a greater role in deciding whether they want to use high cost treatment. This absence of research, and multi-speciality involvement, makes it difficult for decision-making at the individual level due to a risk of cognitive biases, and at the multi-disciplinary level as different departments in the hospital perceive the purpose of critical care in different ways. Recent research in the UK suggested the treatment of referral decisions as binary 'admit or not' decisions was unhelpful, and that there were muliple pathways that doctors could choose between when considering patient referral (Charlesworth, Mort & Smith, 2016). However, we disgree with the authors' conclusion that such multi-faceted decisions therefore cannot be framed by clinical guidelines as guidelines will not outperform mature clinical judgements. In line with the methodological ethos of NDM, we argue that it is possible to build flexible clinical guidelines that can facilitate decision making 'in the wild' by drawing from this very same clinical expertise. Furthermore, it is possible to explore expertise across specialities to contribute to a greater understanding of conflicting and counter-productive cognitive drivers that can be usefully translated into training packages to facilitate the development of accelerated expertise. Ultimately, critical care decision making is complex and exacerbated by a lack of guidance at both individual and multi-specialty levels. This research will explore cognitive processing in this domain to help work towards a solution.

AIMS AND OBJECTIVES

The purpose of this research was to qualitatively explore the decision making processes used by doctors from different specialities when referring and receiving patients for admission to critical care, with a view to generate a greater understanding, and comparison, of how they assess patient risk. It had four aims, to identify:

- The main pressures and cognitive biases that influence decision making for non-critical care doctors referring patients to critical care.
- The main pressures and cognitive biases that influence decision making for critical care doctors admitting patients to critical care.
- The level of understanding between these two groups of doctors.
- Methods for developing a greater shared understanding to improve multi-speciality decision making.

METHODOLOGY

The research aims were met through semi-structured individual interviews with non-critical care doctors (n~12) and critical care doctors (n~12). Doctors were recruited via email and word of mouth from a hospital in the North West of England with a mixed general and neurosciences critical care unit. Participants were interviewed in a quiet and private room located in the Clinical Research Facility at their place of work. Each interview had three persons present: the interviewee, the primary interviewer and the second interviewer/note taker. The interviews used the Critical Decision Method (CDM) (Crandall et al., 2006) interview technique to identify the core challenges, and possible solutions, to critical care referral decisions. A CDM interview involves a multi-pass, retrospective discussion of a challenging decision made by the doctor in the past. Specifically, we asked clinicians to discuss a challenging incident whereby they referred a patient who was refused admission to critical care (non-critical care doctors), or who received a referral for a patient that they refused to admit to the unit (critical care doctors). It involved four phases: (i) incident identification (free narrative recall of the event); (ii) timeline verification (identification of crucial decision points); (iii) cognitive probing (identifying the factors that guided or hindered their cognitive processing); and (iv) hypothetical consideration (consideration of how a novice or external team member may have interpreted the decision). The interview protocol is displayed in Table 1. Using this method, each interview lasted approximately 60 minutes (research is currently in progress, and so a final SD and M is unavailable at the time of writing). The interviews were audio recorded on a Dictaphone and later anonymously transcribed for qualitative analyses. Thematic analyses were used to analyse the data (Braun & Clark, 2006) using the qualitative analyses software 'NVivo'. This involved a process of inductive coding (i.e., bottom-up coding) to identify emerging themes, refinement of themes into codes and further deductive coding (i.e., top-down coding) to produce a rich and detailed description of the data to answer our research questions.

Table 1. CDM Interview protocol (adapted from Power & Alison, 2017)

PROBE TOPIC	PROBES [To be used flexibly and in response to the discussion on the interviewee. Do NOT use probes to 'lead' the interview]
Basis of choice	Why did you [Non-CrC] refer the patient? [CrC] not admit the patient? What did you believe the consequences of your choice may be? What were these beliefs based upon? Were you following any standard rules or operating procedures? Had you been trained to deal with this type of event? Were you reminded of any previous experiences? Did you consider any other courses of action?
Information and Cues	What information did you have available to you when making this choice? Which pieces of information were most/least important? Did you use all the information you had available to you? Was there any additional information you would have liked? Did you seek guidance from someone else at this point? How did you know to trust the information you had? Was there any information that you found unhelpful?
Goals	What would you say was at the forefront of your mind when making this choice? What was your main goal? Did you have any competing goals or objectives?
Decision timing	How did you know when to make this decision? How long did it take you to reach the decision? Did you feel under any time pressure? Could you have delayed your choice? Did you try to avoid or defer your decision at all?
Influence of uncertainty	How certain or unsure were you about your choice? At any point did you find it difficult to process the information you had? Did you feel confident in your decision at the time? Did you feel satisfied with your decision?
Decision barriers	In your opinion, what were the biggest barriers to your decision making on this case? Were there any organisational issues that made your choice difficult? Ward context (bed space, n of patients)? Were there any human factors that made your choice difficult? Shift patterns (tiredness)? Were there any team issues that made your choice difficult? Within your team? Differences of opinion (treatment styles)? From external team members/other specialities? Previous inter-personal relationships?

	Were there any patient-specific issues that made your choice difficult? E.g. Patient's family? Religious reasons? Unsure of patients' preference for end of life?
Decision strategy	What types of strategies did you adopt to try and make this decision? Do you think that you could develop any rules or tricks that could assist another person to make this decision?

RESULTS

Research is currently 'in progress', and so results cannot be reported at this time. However, by the date of the conference in June, data will have been analysed and preliminary findings presented in our poster.

CONCLUSION

The results of this study will be used in a number of ways. In the immediate term, the results will be used to inform the basis of a larger grant application focused on expanding what we have learned about critical care referral decisions from this research. The aim of the larger project is to ultimately develop a novel decision protocol to guide multi-specialty critical care decision making, either via the development of flexible clinical guidelines or the development of a novel training programme to help increase understanding about critical care decision-making between different specialities. The results of the proposed research will also be disseminated via a research paper for submission to the journal *Anaesthesia*. In the longer term, it is hoped that this research will be used to generate real impact on the clinical decision-making of medical practitioners, by the uptake of the protocol at local or national levels. Fundamentally, this research will create a novel insight into multi-speciality decision-making in critical care contexts, evidencing the strength and importance of NDM methods to ensure that context and end-user application are at the centre of research.

ACKNOWLEDGMENTS

This research has been funded by Lancaster University's Faculty of Science and Technology Research Grant.

REFERENCES

Blanch, L., Abillama, F.F., Amin, P., Christian, M., Joynt, G.M., Myburgh, J., Nates, J.L., Pelosi, P., Sprung, C., Topeli, A., Vincent, J-L., Yeager, S., & Zimmerman, J. (2016). Triage decisions for ICU admission : Report from the Task Force of the World Federation of Societies of Intensive and Critical Care Medicine. *Journal of Critical Care, 36,* 301-305.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3, 77-101.

Crandall, B., Klein, G., & Hoffman, R.R. (2005). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis.* Cambrdige, MA: MIT Press.

McAndrew, N.S., & Leske, J.S. (2015). A balancing act: Experiences of nurses and physicians when making end-of-life decisions in intensive care units. *Clinical Nursing Research, 24(4),* 357-374.

Pattison, N., & O'Gara, G. (2014). Making appropriate decisions about admission to critical care: the role of critical care outreach and medical emergency teams. *British Association of Critical Care Nurses, 19(1),* 4-6.

Power, N., & Alison, L. (2016). Offence or defence?: approach and avoid goals in the multi-agency emergency response to a simulated terrorism attack. *Journal of Occupational and Organizational Psychology.* Online first.

Power, N., & Alison, L. (2017). Redundant deliberation about negative consequences: Decision inertia in emergency responders. *Psychology, Public Policy & Law.* Online first.

Rhodes, A., Ferdinande, P., Flaatten, H., Guidet, B., Metnitz, P.G., & Moreno, R.P. (2012). The variability of critical care bed numbers in Europe. *Intensive Care Medicine, 38,* 1647-1653.

Truog, R.D., Brock, D.W., Cook, D.J., Danis, M., Luce, J.M., Rubenfeld, G.D., Levy, M.M., Task Force on Values and Ethics, and Rationing in Critical Care (VERICC). (2006). Rationing in the intensive care unit. *Critical Care Medicine, 34(4),* 958-963.

Focus shift: Differences in reasons generated using Premortem and Worst Case Scenario plan evaluation methods

Madeline PEABODY and Elizabeth S. VEINOTT Ph.D.
Michigan Technological University
Houghton, MI USA
mpeabody@mtu.edu;
eveinott@mtu.edu

ABSTRACT

Motivation: In this paper, we describe a field test comparing the Premortem and Worst-case scenario plan-evaluation techniques with real teams that generated different plans to meet the same goal. Research Approach: Sixty-eight members of five different teams competing in a yearly snow statue competition at Michigan Technology University completed an on-line evaluation of their team's plan using either the Premortem or Worst Case Scenario technique. Results: Participants generated a similar numbers of reasons for failure and solutions in the two methods, but the type of reasons generated were different. Chi-square analyses indicated that participants in the Worst-case scenario method generated more reasons outside their control, while those using the Premortem generated more reasons that were within their control. Similar results were found for solutions. Plan confidence changes over the course of the plan evaluation did not statistically differ by method. Originality: This paper provides one of the first experimental, direct comparisons of the Premortem and Worst Case Scenario methods with individuals on teams executing their own plans. It suggests that while the plan evaluation methods are similar, the information they generate differs in important ways.

INTRODUCTION

People are often overconfident and optimistic that their plans will be executed in a timely and effective manner (Buehler, Griffin, & Ross, 1994). Many strategies have been developed to assess plans in organizations, but the successes of these approaches remain largely anecdotal (Meissner & Wulf, 2015). We empirically compare two plan evaluation methods, the Premortem and the Worst-case Scenario method (Sunstein, 2009) with teams who have developed their own plan. The Premortem (Klein, 1998; 2007) combines a prospective hindsight mechanism (Mitchell, Russo, & Pennington, 1989) and a failure frame (Tversky & Kahneman, 1981) by assuming the plan failed catastrophically. A previous empirical evaluation of the Premortem method found a greater reduction in plan confidence than using a Pro/Con, Con Only, and Critique plan-evaluation method (Veinott, Klein, & Wiggins, 2010). Using this concrete strategy, research has shown that people calibrate their plan confidence more effectively with the Premortem and develop more resilient plans (Peabody, 2017; Veinott & Peabody, 2017; Veinott, Klein, & Wiggins, 2010). In contrast, worst-case scenarios tend to focus on a single, worst case. Both techniques topically focus on plan failure, but differ in terms of their strategy. In this paper, we examine whether this difference in strategy affects people's plan evaluations. This work extends this research to a new, novel, real-world planning exercise with individuals.

METHODS

Sixty-eight members from five organizations participated in this experiment while competing in a month-long snow statue contest during a university winter carnival. Participants included 75% males, and ranged in age from 18 to 33 years-old (M=20 years, SD=1.99). Each year about 30 competing organizations design, engineer, and construct ice statues that represents a theme (e.g., space travel, automotive history). These outdoor statues are large, often 10 meters tall by 20 meters wide and teams spend months planning, and one month executing the plan in the middle of winter. Participants from the planning teams of five organizations responded to an on-line questionnaire that randomly assigned them to either a Premortem or Worst-case scenario plan-evaluation process. They reviewed their team's plan and rated their initial confidence in the plan (0-100% confidence). Next, those in the Premortem condition were asked

to imagine an outcome where your snow statue is a complete fiasco- your group has failed miserably. List as many reasons as you can as to why this happened for a couple minutes. Participants in the Worst-case scenario condition were instructed to imagine the worst-case scenario for your group's snow statue. List as many reasons as you can as to how this scenario might happen. After this, participants provided their second plan confidence rating, generated solutions to the problems for a few minutes, and provided a final plan confidence rating.

Two independent coders rated the 199 reasons and 155 solutions generated and assigned each to one of the following four categories: a) External factors (time, weather), b) Equipment (pipes, supports), c) Leadership/Knowledge (communication, team experience), and d) Human factors (errors). Overall, inter-rater agreement was high with a Cohen's Kappa=.81.

RESULTS

Confidence

A 2 Method (Premortem, Worst-case Scenario) x 3 Time (Initial, Post Reasons, Post Solution) mixed factorial ANOVA was used to evaluate plan confidence ratings. Time was the within-subjects variable. A statistical significant main effect of Time on plan confidence ratings, $F(1, 243)=133$, $p<.001$, indicated that team members' confidence changed between ratings. However, the main effect of Method on plan confidence was not statistically significant, $F(1, 243)=.376$, $p=.540$.

Reasons and Solutions

Participants in the Premortem condition generated about three reasons (M=3.2), while those in the Worst-case generated two reasons (M=2.6. A 2 Method (Premortem, Worst-case Scenario) x 2 (reasons, solutions) mixed factorial ANOVA indicated that there was no interaction, $F(1, 66)=.81$, $p=.53$, and no main effect of Method on the number of reasons, $F(1, 66)=2.01$, $p=.16$, or solutions, $F(1, 66)=.646$, $p=.43$. However, there was a statistically significant main effect of knowledge on the number of reasons and solutions generated, $F(2, 63)=3.42$, $p=.039$, indicating that those with greater knowledge of the plans generated more of them. When we compared the distribution of reasons generated, an interesting pattern emerged. A chi-square indicated that the distribution across reasons for the two methods was different, X2(3, N=199)=8.88, p=.03. Participants in the Worst -case scenario conditions generated proportionately more reasons for plan failure that were outside their control (M=50.6%) compared to the Premortem (M=32.7%), while those in the Premortem condition generate more that were within their control (M=34.5%) compared to those in the Worst-case condition (M=19.1%). As solutions were only generated individually and not discussed in a group, the distribution pattern was the same for solutions, X2(3, N=155)=13.4, p=.004.

CONCLUSIONS

This study adds to the growing body of research quantifying plan evaluations in several ways. While the number of reasons and solutions generated individually did not statistically differ, the type of reasons did in interesting ways. Because reasons from the Premortem were more in the teams' control, the solutions were more actionable than those generated in the Worst-case scenario. This pattern of explanations mirrors research on attribution theory in which people focus on situation or person attributes (Ross, 1977) . It suggests that the Premortem method differs from the Worst -case scenario method in important ways, extending the findings from Veinott, Klein, & Wiggin (2010). The Premortem method can be effective for organizations, including military small unit teams (Vane & Toguchi, 2010) and has recently been suggested as a formal step in a scenario planning method to ensure more flexible strategies (Meissner & Wulf, 2015).

ACKNOWLEDGEMENT.

We want to thank Elizabeth Kelliher and Kaitlyn Roose for help coding the data.

REFERENCES

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "Planning Fallacy": Why People Underestimate Their Task Completion Times. Journal of Personality and Social Psychology, 67(3), 366-381.

Klein, G. (2007, September). Performing a Project Premortem. Retrieved from Harvard Business Review: https://hbr.org/2007/09/performing-a-project-Premortem

Meissner, P., & Wulf, T. (2015). The Development of Strategy Scenarios Based on Prospective Hindsight. Journal of Strategy and Management, 8(2), 176-190.

Mitchell, D. J., Russo, J. E., & Pennington, N. (1989). Back to the Future: Temporal Perspective in the Explanation of Events. Journal of Behavioral Decision Making, 2, 25-38.

Peabody, M. (2017). Improving Planning: Quantitative evaluation of the Premortem Technique. Unpublished Masters of Science Thesis. Michigan Technological University.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. Advances in experimental social psychology, 10, 173-220.

Sunstein, C. R. (2009). Worst-case scenarios. Harvard University Press.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. Science, 211, 453-458.

Vane, M. A., & Toguchi, R. M. (2010). Achieving Excellence in Small-Unit Performance. Military Review, 73-81.

Veinott, B., Klein, G. A., & Wiggins, S. (2010). Evaluating the Effectiveness of the Premortem Technique on Plan Confidence. 7th International ISCRAM Conference. Seattle.

Veinott E. & Peabody, M. (2017) Planning in the wild: Effect of planning method on plan resilience. Manuscript in preparation.

Extending Working Life in the NHS – Opportunities, Challenges and Prospects

Work and Retirement Decision-making – The need for a broader perspective on choice architecture.

Andrew WEYMAN & Deborah ROY
Department of Psychology, University of Bath
Invited poster

ABSTRACT
Arguably the key finding on pension investment and retirement decision making is that most UK citizens are not active planners. Rather, they are more disposed to react to options they are presented with. Although there is almost universal acceptance of the need to make financial provision for retirement, there is significant inertia and reticence over active engagement. Most people have limited knowledge or understanding of options or their implications and those are motivated to engage tend to become bewildered by the complexity and unknowable elements. People are also being *asked* to make their choices against a background of unprecedented (in the post WW2 period) flux and fluidity in employer and State pension arrangements. Recent rises in State pension age, the prospect of further rises, combined with high profile media reporting of pension failures, tends to interpreted as a system in crisis. Potentially corrosively, there is a risk that the associated uncertainty may feed latent procrastination and inertia, to further inhibit already weak motivations to engage in *saving for tomorrow* (Pettigrew, et al, 2007; **Wicks and Horack, 2009**); rather than strengthen motivation to plan. The UK Government's auto-enrolment pension scheme goes some way towards addressing this, but there are questions over whether its voluntary nature and realised pension value will prove to be sufficient to be meaningful and the impact of planned rises in employee contributions on opt-out rates.

The key role played by employers in configuring the choice architecture that underpins 50yrs+ pension and employment options seems to be conspicuous by its absence within UK public policy perspectives on motivating extended working life. While the, largely G/government funded research in this area points to the attractiveness of mid-late career transitions to part-time hours and a desire for greater flexibility over hours worked (features that tend to be attractive to employees of all ages), there has been very little attention paid to the orientations of employers towards older workers; the impact of employer practices on employee retirement 'decisions', or ways in which to motivate employers to adopt more older worker friendly policies. Contemporary UK policy perspectives and commissioned research on extending working life present as dominated by a theory of change focused on individual motivation (for a review of evidence see Weyman et al, 2012).

It may be tempting to conclude that this partial and myopic perspective in this area of public policy reflects some overt attempt at alignment between influential individual psychology models[5] that have tended to dominate public policy perspectives on behaviour change over the last two decades (see Solesbury, 2001, Healy, 2002) and the Government's personal responsibility agenda. However, on balance any alignment seems more likely to be the product of serendipity than purposeful intent. It is also presents as being out of step with the choice architecture perspective that has been influential elsewhere, Nevertheless, the post 2010 period has witnessed a strong steer from Government for its Departments and Agencies to reduce the regulatory burden on employers (House of Commons, 2005; Cable, 2011; Department for Trade, 2012). Whether the product of philosophical underpinnings or happen-stance, limiting the public policy perspective on intervention to approaches that seek to change individual behaviour, within an otherwise unchanged world, presents as partial.

[1]Contemporary UK perspectives on behaviour change are dominated by health-psychology derived vale expectancy individual motivation models - see National Institute for Clinical Excellence 2007; National Institute for Care Excellence 2014.

A focus on individuals as the primary focus for intervention embodies the assumptions that: employee volition represents the principal barrier to extending working life; that individuals possess a strong, or at least sufficient, sense of agency and knowledge of their options and arising implications, and that employers are motivated to retain older staff.

The overwhelming evidence is that the majority of established employees are not active decision makers in the extending working life domain. The dominant picture is one of rather vague and impressionistic understandings of the options available to them, and their implications, e.g. in relation to tax liabilities and impacts on pension values, and a tendency to be inhibited from initiating discussions with their employer over late career working arrangements, e.g. due to limited precedent / norming, or worries over perceptions of commitment (Hedges and Sykes, 2009; Wicks and Horack, 2009; Weyman et al, 2012).

In sum, mirroring evidence from UK pensions research, findings indicate that most individuals are not options seekers, they are more accurately characterised as the reactive to options presented to them (Phillipson and Smith, 2005; Vickerstaff et al, 2008; **Hedges and Sykes, 2009;** Weyman, et al 2012). Employers play a key role in defining the choice architecture of late working life employment options and, potentially, as initiators of discourse with employees over their needs, aspirations and intentions. In many organisations it is apparent that features recognised as attractive to older employees, such as a transition to reduced or flexible hours, tend to be weakly normed, with limited precedence and restricted access. At a fundamental level, in the absence of legislative change, employer motivations to retain and engage older workers rest upon the value they perceive from doing so. This can be predicted to be variable across different employment sectors, with some evidence that it is strongest amongst those experiencing skills shortages, e.g. sectors such as engineering, and former nationalised industries, where changes in the industrial base over recent decades have resulted a contraction in the availability of traditional apprenticeship opportunities (Vickerstaff et al - in progress).

For older individuals in stable employment, the opportunity to remain with their established employer has been found to produce a more favourable result, in terms of rates of pay, quality of work and quality of working life, than individuals who have a need to seek alternative employment (Phillipson and Smith, 2005). In the case of the former, an extension to working life represents an opportunity to capitalise on their established experience and skill set. However, for over 50's who become challenged by immutable job demands, relative to their capacity to meet them; experience imbalance / dislocation between their work and non-work aspirations / commitments; are victims of employer age discrimination; or otherwise find themselves out of work, there choice is limited to seeking alternative employment, or withdrawal from the labour market. Impacts can be predicted to vary structurally, with respect to employment sector and skill-set. Those who work in physically demanding occupations, e.g. construction, or are exposed to high job-stress / high job-demands, e.g. emergency services; are least skilled, or suffering from long term ill health / disability present as the most vulnerable to forced late career transitions (Vickerstaff et al, 2008).

Changes to NHS pension arrangements, together with the raising of the State pension age, are widely assumed to result in people working longer. However, the extent of the increase, the extent to which it will embrace all professions and staff grades, and how it will affect specific employers are as yet unclear. Variables affecting the retirement decision include individual factors (physical and mental health status, psychological dispositions, attitudes and beliefs, and knowledge of pension and retirement options); financial factors (pension value, other savings and financial obligations, e.g. dependants, mortgage status); family and social network factors (caring responsibilities, retirement date of partner and reference group retirement norms); and workplace factors (job quality, job design, job demands, employer policies and practices, e.g. opportunities for flexible working, retention of older workers) (see Shacklock *et al.*, 2007; Armstrong-Stassen and Ursel 2009; Maitland, 2010). How these factors will influence extended working life in the NHS is unclear.

A component of our Medical Research Funded research on Extending Working Life uses a mental models approach to explore and characterise employee decision making. Mental models relate to how people make sense of their world, and make decisions on the basis of this often impressionistic conceptualisations and partial understandings. Understanding extant mental models of employees and their managers is important as it provides insight into the criteria that people are applying in their decision making. e.g. there is strong evidence that people exhibit an array of known decision biases and heuristic judgements in the area of extending working life and pension choices (with the result that they sometimes make poor choices), and that stereotypes impact on perspectives on older employees (Weyman, 2010; Weyman et al, 2012; Weyman et al, 2013).

Mapping mental models is therefore important from the perspective of communication with employers, managers and employees, and in informing the design interventions that address older worker issues, not least with respect to aspects relating to cultural change. Insight into employee mental models is of particular salience in the context of exploring how aspects of the established (and potentially malleable) decision architecture (options available to employees e.g. part-time or flexible work - see Thaler and Sunstein, 2008; Weyman et al, 2012) impact on employee retention / retirement behaviour and pension choices (e.g. choices over alternative pension schemes, early draw-down). Their impact on employee orientations to extended working life and decision behaviour at key transition points are of central interest.

References

Armstrong-Stassen. M. and Ursel, N. (2009) Perceived organizational support, career satisfaction, and the retention of older workers. *Journal of Occupational and Organisational Psychology*; 82(1); pp. 201-220.

Hedges. A.; Sykes, W. and Groom C. (2009) *Extending working life: Changing the culture, Qualitative research into effective messages* Department for Work and Pensions Research Report No 557

Maitland, A. 2010. *Working Better: The over 50s, the new work generation*. Manchester: Equality and Human Rights Commission

National Institute for Care Excellence (2014) *Evidence reviews and economic analysis for NICE guidance on workplace health of older workers*. http://www.employment-studies.co.uk/project/evidence-reviews-and-economic-analysis-nice-guidance-workplace-health-older-workers

National Institute for Clinical Excellence (NICE) (2007) *Behaviour change at populations, community and individual levels.* London. National Institute for Clinical Excellence.

Pettigrew, N., Taylor, J., Lancaster, J. and Madden, R. (2007), ‘‘Live now, save later? Young people, saving and pensions’’, DWP Research Report No. 438, HSMO, Norwich.

Phillipson, N, C. & Smith, A. (2005) *Extending working life: A review of the research literature*. London: Department for Work and Pensions.

Shacklock. K. and Fulop, L. (2007) Managing older worker exit and re-entry practices: A ‘revolving door’? *Asia Pacific Journal of Human Resources. 45(2); pp. 151-166*

Thaler R. H. and Sunstein C. R. (2008) *Nudge: Improving Decisions About Health, Wealth and Happiness* Yale. University Press: New Haven and London.

Vickerstaff, S. *et al.* (2008). *Encouraging labour market activity among 60-64 year olds*. [Online]. Department for Work and Pensions. Available at
http://webarchive.nationalarchives.gov.uk/+/http://www.dwp.gov.uk/asd/asd5/rports2007-2008/rrep531.pdf.

Weyman, A.K. (2010) What about older and debilitated workers? in *Well-being and working life: towards an evidence-based policy agenda.* ESRC Seminar Series. Mapping the public policy landscape.
http://www.esrc.ac.uk/funding-and-guidance/collaboration/seminars/archive/well-being.aspx

Weyman, A.K.; Wainwright, D.; O'Hara, R.; Jones, P. and Buckingham, A. (2012) *Extending working life behavior change interventions.* Department for Work and Pensions, Research report 809.

Weyman A.K.; Meadows, P. and Buckingham, A. (2013) *Extending working life: Audit of research relating to impacts on NHS employees*. Report to the NHS Working Longer Review Group, NHS Staff Council.

Wicks, R. & Horack, S. 2009. I*ncentives to save for retirement: understanding, perceptions and behaviour A literature review* London: Department for Work and Pensions.

Understanding Employee Susceptibility to Phishing: A Systematic Approach to Phishing Simulations

Emma J. WILLIAMS[a] and Adam JOINSON[a]
[a]School of Management, University of Bath, BA2 7AY

ABSTRACT
Employees are increasingly targeted by fraudulent emails sent by malicious individuals attempting to circumvent secure technical controls. However, current understanding of the factors that impact employee decision-making regarding phishing emails within the workplace remains limited, with experimental work often conducted outside of the naturalistic settings in which phishing emails are encountered. The current paper proposes an initial framework for exploring naturalistic decision-making in this area using data from routine phishing simulations. These simulations can be harnessed to further understand the range of factors that may impact employee decision-making within the workplace, including how potential susceptibilities may be more effectively reduced.

KEYWORDS
Decision Making; Security; Cyber Security; Social Engineering; Phishing.

INTRODUCTION

Malicious outsiders commonly use phishing emails as a means to gain access to secure technical systems by targeting employees and persuading them to click on malicious links or attachments (Workman, 2008). These emails often target particular individuals or groups (a process commonly known as *spear phishing*) and use a range of well-documented influence techniques to persuade users that (a) the communication is genuine (and hence should be trusted) and (b) that they should undertake the desired behavior with minimal consideration (e.g., click on a link, open an attachment, respond to the sender, provide sensitive information or forward the communication on to colleagues). These techniques often include:

- **Instilling a sense of urgency**, such as requiring a response within 24 hours to prevent account closure or providing time-limited or time-relevant information (Stajano & Wilson, 2011).
- **Providing information of interest or use to the user**, whether through professing information that will be perceived as important or required to complete a work or personal task, or referring to information that is likely to 'grab' attention and induce curiosity or credulity (Loewenstein, 1994).
- **Encouraging emotional responses**, usually through anxiety or panic relating to a potential threat or loss (e.g., freezing an account, removing or restricting access or availability, or claimed identity theft) or by inducing positive emotions, such as excitement, desire, pride or hope relating to excessively large prizes, 'too good to be true' offers, limited opportunities or miracle cures (Finucane, Alhakami, Slovic & Johnson, 2000; Zajonc, 1980).
- **Exploiting compliance with authority**, whereby individuals are instructed to complete a task (such as processing an invoice or reading a policy document) by someone impersonating a relatively high status individual within the organisation (Cialdini, 2007).
- **Focusing on contextual or work-related communication norms** (Kahneman, 2011), including cultural holidays or events (e.g., Christmas, Easter, World Cup), activities (e.g., parcel delivery updates), and common or targeted work topics (e.g., policy updates, invoices to finance personnel, personal details update forms to HR personnel).

A number of high-profile cases have highlighted the potential impact of phishing at the organisational level, which has contributed to an increased interest in mitigating the risk of these influence attempts (e.g., Gmail attack, 2011; IMF attack, 2011; RSA attack, 2011). Specific details regarding the scenarios of successful phishing attacks are not generally openly accessible however, making it difficult to determine the various factors that may impact naturalistic decision-making in this area, and crucially, how such susceptibilities can be reduced in the future.

Recent theoretical models suggest that the influence techniques used within phishing emails monopolise attentional resources and encourage recipients to engage in relatively automatic, heuristic forms of processing when deciding how to respond. This can lead to authenticity cues within the email, such as discrepancies in the sender address, being overlooked (Vishwanath, Harrison & Ng, 2016; Vishwanath, Herath, Chen, Wang & Rao, 2011).

In addition to aspects of the message itself, the extent that individuals rely on such forms of processing is also likely to be influenced by the specific characteristics of the user, including their online communication habits, the degree of knowledge that they have, and their beliefs regarding the risks of operating online (Vishwanath et al, 2016). However, current understanding regarding the primary factors that may impact decision-making within the workplace remains limited and likely reflects a complex interplay between the context that an employee is operating within (e.g., job-specific factors, expectations and norms regarding incoming communications, organisational security culture), the particular phishing email that they are targeted by (e.g., the particular influence techniques that it uses) and their own characteristics (e.g., knowledge, beliefs regarding online risk).

USING PHISHING SIMULATIONS TO UNDERSTAND SUSCEPTIBILITIES

In order to raise employee awareness of the risk of phishing emails at work, many organisations run routine phishing simulation exercises that target their own employees. This not only provides a means to educate users, but also allows the organisation to understand its relative risk from phishing by monitoring the proportions of staff that are likely to respond to phishing emails (known as the 'click-rate').

Substantial data can be captured from in-house phishing simulation exercises, including the proportion of people in different staff groups or divisions who click on links contained within emails (click-rates) or provide their user credentials (disclosure-rates), the proportion of people who report the email as suspicious to IT security personnel (reporting-rates), and the time of day, week or year that the e-mail is viewed and responded to. The extent that this data is currently utilized varies according to the resources and goals of the particular organisation, however, it provides a currently under-used resource to further examine the potential factors that may impact employee decision-making.

Following initial discussions with security professionals from a range of organsiations within the UK Critical National Infrastructure (CNI), potential opportunities to harness this existing source of 'in-house' data were identified by the authors. This focused on identifying the particular influence techniques used within historic phishing simulation emails and examining how differences in these techniques may impact employee response behavior. In order to do this, an initial screening of the influence techniques used within phishing simulation emails was undertaken to explore the potential of this data source further (for example, see Table 1).

Table 1. Example method for initial screening of simulation phishing emails

Subject of E-mail	Primary Influence Techniques	Click Rate
Secure e-mail delivery	- Focuses on urgency with time-limited availability to access mail ("*link will expire 24 hours after this notification*") - Suggests a loss through the threat of withdrawal of access to information that may be important in job context - Invokes curiosity as to the content and sender of the secure e-mail ("*someone has sent you an email*")	32%

Unfortunately, the extent that historic data can be exploited is currently limited by considerable variation in the type of phishing simulations that are sent to users, with different employees potentially receiving different e-mails, and little standardisation in relation to the language, content or length of e-mails sent at different time-points. This can make it difficult to compare response behavior over time, across different staff groups or across message types due to the different motivating factors that are likely to be involved (e.g., the degree of relevance, interest or perceived authenticity). By developing a more systematic approach to phishing simulations, it is hoped that such data can be harnessed to address the following key questions:

- Are some phishing e-mails more effective in persuading employees to respond in a naturalistic work context? If so, why? What influence techniques do they use? And how best can we improve employee resistance to these techniques?
- Does the impact of particular e-mails on decision-making vary according to situational or individual factors, such as staff grade, job role, time of day, degree of training, or security culture? If so, can differences in individual or contextual susceptibility be targeted more appropriately when designing interventions and mitigations?

A SYSTEMATIC CLASSIFICATION APPROACH

A standardised classification framework was developed based on the influence techniques that may be present within a particular phishing email, which allowed simulation emails to be scored according to the presence or absence of a range of influence techniques and authenticity cues by multiple independent raters. Further detail regarding this scoring framework is shown in Figure 1 and Figure 2.

Definitions	
Timely	The content of the e-mail is relevant only at a particular time of year or in relation to particular events. For example, cultural holiday periods, referencing recent or upcoming events or news items that the individual is likely to be familiar with due to general societal awareness or features in the media.
Personal	The content of the e-mail is designed to relate directly to the receiver and any online accounts that they may have, such as account notifications, receipts or deliveries.
Work	The content of the e-mail relates directly to the work context of the receiver, such as organisational communications and work-related communications.
Reward	The e-mail provides a tangible reward if the receiver responds, such as discounts, special offers and prizes.
Loss	The e-mail states that the receiver will lose something if they do not respond, such as account access, money, security, or information.
Information	The e-mail states that the receiver will be provided with useful or relevant information if they respond, such as policy documents, news stories, or alerts.
Curiosity	The e-mail provides information that will provoke the interest of the target, but not due to relevance or need, such as funny videos, images, or news stories that are not of direct relevance to the receiver.
Social	The e-mail exploits social norms, rules or processes to persuade the target to respond, such as the tendency to feel obliged to repay gifts or favours, using peer pressure or conformity, making polite or friendly requests and helping others in need.
Urgency	The e-mail states that the receiver has a limited amount of time to respond if they wish to engage with the e-mail content, such as being time-limited, urgent or scarce.

Figure 1. Definition list used for independent rating of phishing simulation emails

Influence Technique					
To what extent does it use reward-based influence techniques?	To what extent does it use loss-based influence techniques?	To what extent does it use urgency-based influence techniques?	To what extent does it use informational-based influence techniques?	To what extent does it use curiosity-based influence techniques?	To what extent does it use socially-based influence techniques?
3 = very much, 2 = a bit, 1 = not at all	3 = very much, 2 = a bit, 1 = not at all	3 = very much, 2 = a bit, 1 = not at all	3 = very much, 2 = a bit, 1 = not at all	3 = very much, 2 = a bit, 1 = not at all	3 = very much, 2 = a bit, 1 = not at all

Figure 2. Example section of scoring framework for rating of phishing simulation emails

This scoring framework is currently being applied to a corpus of historic phishing simulation data from organisations that will allow further understanding of the extent that particular message factors impact employee decision-making within the workplace. By using this framework in combination with a more rigorous simulation methodology, it is hoped that such data can be used to address a range of questions that will be of interest to both academic and practitioner communities.

For example:

- **Are some groups of employees more likely to respond to phishing emails than others?** (e.g., different length of time in the organisation, different job roles or clearance levels, different geographical location). This may be used to develop targeted initiatives for particular employee groups.
- **Has employee susceptibility to phishing emails changed over time?** This may be used to assess the impact of an intervention, such as the role-out of new training between specific time-points, or responses at different times of the day or week.
- **Are employees more susceptible to certain types of phishing email?** This may relate to the types of phishing emails that employees typically receive and will help identify particular vulnerabilities that an organisation may have. For example, internal versus external emails, or those which use different influence techniques (e.g., claiming to offer a reward versus a potential loss of access to an account). This can then be used to prioritise training and awareness campaigns in relation to these types of attack.

CONCLUSION

The current paper provides an overview of on-going work to exploit phishing simulation data in order to further understand why employees respond to phishing emails within workplace settings. By understanding the primary mechanisms that impact decision-making in this context, it is hoped that more effective, targeted interventions can be developed to reduce susceptibility in the future, including training and awareness initiatives, decision support systems and improvements to organisational policies, processes and procedures.

The proposed framework represents an initial process that can be used to aid in the examination of phishing simulation data by focusing on the influence techniques used within phishing emails. It is currently undergoing iterative development and testing using historic organisational phishing simulation data. By combining this framework with particular simulation methodologies in the future, it is hoped that researchers and organisations can systematically examine potential relationships between factors related to the message itself (e.g., the influence techniques that it uses and the scenario that it creates) and wider situational factors (e.g., organisational culture, job role, staff grade, exposure to training) when considering naturalistic decision-making in relation to phishing attacks. Only by understanding why and when phishing emails may be successful, can truly effective means of countering them be developed.

ACKNOWLEDGMENTS

We thank all individuals and organisations that have provided their time and comments in relation to the development of our approach.

REFERENCES

Cialdini, R. (2007). Influence: The psychology of persuasion. New York: HarperCollins.

Finucane, M.L., Alhakami, A., Slovic, P. & Johnson, S.M. (2000). The affect heuristic in judgment of risks and benefits. Journal of Behavioral Decision Making, 13, 1–17.

Kahneman, D. (2011). Thinking, fast and slow. London, UK: Penguin.

Stajano, F., & Wilson, P. (2011). Understanding scam victims: Seven principles for systems security. Communications of the ACM, 54(3),70-75.

Vishwanath, A., Harrison, B., & Ng, Y.J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. Communication Research, online pre-print, 1-21.

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H.R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. Decision Support Systems, 51, 576-586.

Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. Journal of the American Society for Information Science and Technology, 59, 662–674.

Zajonc, R.B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35, 151–175.

PANELS

Decision Making Under Uncertainty (DMU)

Panel Chair: Gareth CONWAY a,
Panellists: Peter CHALLENORb, Simon FRENCH c,
Harold SIMPSON d, and David TUCKETT e
a Dstl
b RCUK Decision-Making under Uncertainty Network (M2D)
c University of Warwick
d UK Ministry of Defence
e RCUK Decision-Making under Uncertainty Network (CRUISSE)

OVERVIEW

Developing and sustaining understanding and making decisions under uncertainty is an area that continues to be relevant for practitioners and academics alike, and has always been a central feature of NDM research. This panel is made up a range of professionals who work with or under these conditions, in different roles, and using different perspectives. Each of the speakers will provide their views and experiences on working with or under uncertainty, and lay out the challenges they see from their perspective. The Panel Chair will pose questions to the panel members, and facilitate questions from the audience. We hope to explore the potential for inter- or multi-disciplinary research, with particular emphasis on proficient sense making and decision making, supporting decision makers, and building the expertise of decision makers.

PANEL CHAIR

Dr Gareth Conway is a Chartered Psychologist at the Defence Science & Technology Laboratory (Dstl), part of the UK Ministry of Defence; he also holds a Visiting Professorship at the University of Huddersfield. His main interest is in how professionals understand and solve problems via the use of professional expertise and processes of inquiry, particularly in the government policy area. As a researcher-practitioner, he has undertaken research on a broad range of defence and security issues, and as a practitioner, he has been attached directly to MOD Head Office, the British Army, and the UK HQ in Afghanistan (Task Force Helmand) to undertake policy reviews and provide scientific advice and analysis directly to senior decision makers. Gareth's current work includes leading MOD's Government Social Research profession; leading the scientific support to the MOD's Policy, Strategy and Parliamentary (PSP) profession; bringing together diverse disciplines who consider decision making under uncertainty; and acting as a consultant and reviewer for research on the development of adaptability in the defence workforce. Gareth has just completed a departmental policy review on ethical and scientific assurance for human participant research.

PARTICIPANTS

Peter Challenor, University of Exeter

Professor Peter Challenor has broad interests; mainly about uncertainty in the natural world. These range from the statistical analysis of complex numerical models (such as those used to simulate climate) to the interpolation of noisy data and the estimation of the amount of renewable energy in the ocean. Peter is the principal investigator on the NERC RAPID-WATCH project RAPIT looking at the risk of the shut down or significant slowing of the Atlantic Meridional Current. This project relies on thousands of climate simulations carried out by members of the public via climate prediction.net.. He is also project lead for M2D. The M2D Network will focus on model-informed decision making and defines itself as broad, inclusive and multidisciplinary. It welcomes the active participation of members from a diverse range of academic research and public / sector working backgrounds.

Models to Decisions (M2D) Network goals

The M2D Network will focus on model-informed decision making and defines itself as broad, inclusive and multidisciplinary. It welcomes the active participation of members from a diverse range of academic research and public / private sector working backgrounds.

Goals:

1. Establish a decision making under uncertainty community in the UK via networking events (both online and real-world).

2. Defining a research agenda that will address how to improve decision making under uncertainty in real-world problems.
3. To actively engage with decision makers, develop and deliver 'best practices' that are salient, useable and credible to the user community.
4. Pump priming research projects via short-term feasibility studies to tackle substantive interdisciplinary research problems involving models, uncertainty and decisions.

Simon French, University of Warwick

Simon is a member of the Applied Statistics and Risk Initiative in the Department of Statistics at the University of Warwick. He is well known for his work on risk and decision analysis, both theory and practice. He has collaborations with psychologists aimed at helping to understand how people behave and how they can be supported in complex tasks and decisions. His book with Nadia Papamichail and John Maule surveys all aspects of decision analysis and support. In applications, Simon has worked with many regulators and organisations in the public sector including the Department of Health, Public Health England, the UK Food Standards Industry, the European Food Safety Authority, the European Space Agency, and many parts of the nuclear industry at national and European levels. He took part in the International Chernobyl Project. In his work, the emphasis is on multi-disciplinary approaches to solving real problems. He has a particular interest in supporting emergency management processes.

Harold Simpson, UK Ministry of Defence

Lieutenant Colonel Harold Simpson MBE read History at Selwyn College, Cambridge, and joined the Royal Army Educational Corps in 1988, after completing officer training at the Royal Military Academy, Sandhurst. He has fulfilled a variety of roles in the UK and overseas including Hong Kong, Cyprus, Sierra Leone, the Sudan and the Balkans. From 2012-2014 he served as a member of the academic staff at the Defence & International Relations Department at Sandhurst. Of recent note, Lieutenant Colonel Simpson has worked in NATO's Joint Force Command (in Brunssum, the Netherlands) in a role focussing on the Allied response to the Ukraine Crisis. He is currently employed in the Ministry of Defence in Whitehall in a military education policy role.

Training and education interventions for decision making under uncertainty

During this first decade of this century the Army became increasingly aware that traditional models of military training and education were no longer working in the harsh realities of operations. In particular the contemporary operating environment was (and indeed still is) dominated by VUCA (Volatile, Uncertain, Complex, Ambiguous) characteristics. Adaptive expertise in VUCA environments demands agile thinking and the ability to 'Learn-adapt exploit-influence'; leaders must be open-minded, who tolerate and learn from their mistakes and those of their soldiers.

The Royal Military Academy was asked to introduce *critical thinking* and *decision making under uncertainty* to the Officer Cadets' curriculum, and based on his experience of operations in West Africa and Eastern Europe, Harold developed a series of complex and challenging exercise scenarios. The scenarios present a number of constantly changing "wicked" problems, reflecting the realities of VUCA, for which the Officer Cadets have received no prior training or experience. The Officer Cadets are therefore obliged to be adaptive in their approach. This introduction to decision making under uncertainty is a foundation to future training at both Sandhurst, and in their subsequent careers.

David Tuckett, UCL

David Tuckett is Director of the University College London (UCL) Centre for the Study of Decision-Making Uncertainty and a Senior Research Fellow at the Kiel Institute for the World Economy in Germany. Trained in Economics, Medical Sociology and Psychoanalysis he is the author of *Minding the Markets: An Emotional Finance View of Financial Instability.* In the last few years he has developed a theory of decision-making under radical uncertainty, Conviction Narrative Theory, which focuses on how agents find the confidence to act when there is no reliable evidence to suppose the outcome will be gain rather than loss. He is leading Research Council UK's Network 'Challenging Radical Uncertainty in Science, Society and the Environment (CRUISSE)' and also a Co-Investigator on the ESRC-NIESR Re-Building Macroeconomics Network.

CRUISSE (Challenging Radical Uncertainty in Science, Society and the Environment) network

The CRUISSE network aims to bring academics from disciplines in mathematical, physical, psychological, social and other sciences together to better understand and help practitioners who are making difficult decisions. Support stems from understanding the actual question, so the first step for the network is to invite decision-makers in industry, government and civil society to discuss their challenges. We will be looking to identify what aspects of the real-world challenges that practitioners face (a) fall into traditional statistical/decision theoretic methods (Box 1), or (b) require a broader approach as mature probabilities are not available (Box 2), or (c) are dominated by radical uncertainty and require thoughtful case-by case reflection (Box 3). The second step is to explore the way problems are being set up and framed in the light of the type of uncertainties identified. We aim to develop useful guidelines including steps towards a guide to good practice, which will broaden the awareness of alternative approaches. Practitioners aware of the kinds of problems and solutions in each “Box”, and where the problems of interest to them lie, will be a step closer to resilient decision-making, informed by seeing approaches to other hazards facing real-world decision making.

Training and Developing Decision Making and Thinking Skills for High Risk Work: A View from the Trenches

Robert HUTTON[a], Rob DAVIS[b], Garin UNDERWOOD[c], Richard O'HARE[b] Phil BUTLER[d], and Sabrina COHEN-HATTON[e]

[a] *Trimetis Ltd*

[b] *Avon Fire & Rescue Service*

[c] *Virgin Atlantic Airways*

[d] *Cardiff University*

[e] *London Fire Brigade*

ABSTRACT

The purpose of this panel session is to provide a snap-shot on several efforts to improve cognitive skills in high risk work environments and to identify challenges from these efforts to the research community. The panel is made up primarily of practitioners representing diverse domains. Each practitioner will present a short description of their operational training problem. The first example is a Fire & Rescue Service command development programme which uses tactical decision making exercises to accelerate the acquisition of decision skills. The second is a nascent aviation 'human factors' development programme which is questioning the value of current Crew Resource Management approaches to training decision making skills. The third is an attempt to develop training and development interventions to improve adaptive thinking performance in a large military headquarters. Following the short presentations, applied training and macrocognitive research challenges will be highlighted as stimuli for attendee discussion.

KEYWORDS

Learning and Training; Emergency Response; Military; Cognitive Skills; Thinking Skills; Aviation; Research-Practice Gap.

INTRODUCTION

This panel session will provide an opportunity for researchers and practitioners to participate in a discussion about the challenges of supporting the development of thinking and decision making skills for operators in high risk, complex environments. The panel is made up of a mix of operational commanders, training and development professionals, and researchers. Each panel member will provide a 5-10 minute introductory description of their training challenge. Following each individual presentation there will be 5 minutes for follow up questions. Once all the presenters have spoken, we will provide a number of applied and research questions and challenges that have emerged from the presentations. The questions will be opened up for discussion with the attendees at the session (assuming a 90 minute session). If required, each participant will also have prepared an answer in terms of their perspective on the problem, for further discussion with attendees. The next section describes each of the panel members and a brief description of the topic for their introductory presentation.

A nascent Community of Interest has emerged in the last 12 months with respect to an interest in training complex, non-technical skills. The presenters here represent the core group but our interest is in growing the group, particularly in the UK and Europe. We intend to collect the names of others interested in participating in this effort following the presentation. The purpose of the group will be to provide a resource for sharing knowledge, skills and ideas around the improvement of training nontechnical skills in the high risk work environments.

FIRE & RESCUE SERVICE TACTICAL DECISION MAKING EXERCISES

Presenter: Rob Davis

Rob Davis is the Assistant Chief Fire Officer for Avon Fire and Rescue, Director, responsible for Learning and Development and Risk Reduction within the service. He has been a Firefighter for 27 years, undertaking most FRS roles; including Operational Response, Hazmat advisor, National Inter-Agency Liaison (NILO) Officer for counter terrorist incidents, flood advisor, trainer/instructor, Urban Search and Rescue (USAR) technician, technical fire safety and unitary manager. Rob has attended numerous emergency response incidents within his career. In addition, Rob is one of the Team Leaders for Search and Rescue Assistance in Disasters (SARAID) a UK NGO

Contact Information of First Author: rob.hutton@trimetis.co.uk

and recognised United Nations organisation, involved in USAR response and Disaster Risk Reduction activities around the world. Rob has responded to the Mozambique floods (2000), Sri Lanka tsunami (2004), Pakistan earthquake (2006) and Haiti earthquake (2010) as a Team Leader. Rob holds a diploma in Emergency Planning and an MSc in Disaster Management from Coventry University, England.

Topic

The relationship between the professional academic rigour of Naturalistic Decision Making (NDM) and the role of Non-Technical Skills (NTS) and its application within the vocational setting of risk critical industries is an interesting one. Within the Fire and Rescue service this research is starting to shape the training required for Incident Commanders to operate in a safe way, the academic knowledge of NDM is more widely understood and training is being designed to embed this academic research into the role of a Firefighter.

Incident Command knowledge has previously centred on technical and procedural knowledge, rather than the tacit knowledge and skills acquired through experience. Following a number of tragic national and international Firefighter fatalities, the role of NDM is the missing piece of the jigsaw puzzle when it comes to safety at the sharp end and combining the technical competence with the non-technical to create a future holistic Incident Commander, making the incident ground a safer place to operate for Fire-fighters and the community they serve.

The question is how do we improve the link between academia and risk critical industries with designed 'awareness and decision making' training packages that can improve the understanding and practical application of NDM and NTS to make our communities and risk critical industries safer?

Additionally, how do we improve the learning environment across risk critical organisations to share best practice in training and simulation and provide greater coherence in learning?

Presenter: Richard O'Hare

Richard O'Hare is a Watch Manager in the Incident Command Training Team of Avon Fire & Rescue Service. Having spent 25 years in the Fire Service, the majority of which serving on operational fire stations, he is now involved in the teaching and assessment of emergency services Incident Command and non-technical skills.

Topics

1. There has been a position within the Fire Service for several years that recognition-primed decision-making is inherently bad. Even when we are trained to use handrails like the Joint Decision Model, research, by the likes of Dr Cohen-Hatton, shows we still use RPD at least 50% of the time. This seems to indicate that use of RPD is inevitable when using humans to make decisions in fast moving, high stress environments?
2. There is a critical challenge with respect to methods of improving training to 'skill-up' our decision-makers in an environment where they get fewer real incidents. How do we turn inexperienced, less confident commanders into capable, expert decision-makers? Computer simulations/exercises are the norm at the moment but often lessons learned fail to be implemented back in the real world.
3. How do we avoid doing this in a judged, assessed environment? Expanding the concept of 'safe training' where 'safe' relates to reputational rather than physical well-being.
4. How do we measure the quality of decisions taken? Is it even possible to empirically measure something which can be so subjective? Are decisions made using handrails 'better' than those made intuitively by an experienced decision-maker?
5. What barriers exist that prevent this training being adopted at the sharp end. Does organisational culture or structure impede us from accepting new practice in decision-making? Does group mentality overcome individual desire to change?

Are we trying to fit a square peg in a round hole? My gut feel is that the Fire & Rescue Services are slightly off on our focus, especially for our initial incident commanders. At the moment we give some input to decision-makers, expect them to 'get it' and then have the weight of focus on summative assessment. This can be either in formal assessments or as monitoring at actual incidents. We have an expectation that decision makers will use handrails like the Joint Decision Model all the time but research shows the best we can hope for is a 50% uptake on that. Are we trying to achieve the impossible by expecting commanders (especially junior levels in the initial phases of incidents) to adopt decision-making practices that human beings simply aren't 'wired' to do?

Research Challenge: If at least half our decisions are intuitive, should we accept this and make sure that people are as fully skilled as possible for that phase? Is the answer more training? Exposure to command situations without the threat of being taken off of command duties or made to look foolish in front of crews. Giving commanders the opportunity to add to their library of experience should improve their RPD phase, allowing them to get swiftly through it and into more analytical thinking, making full use of handrails, which will then raise their situational awareness.

Presenter: Phil Butler

Phil served in the UK Fire and Rescue Service (FRS) as a firefighter and officer for 31 years. After joining the London Fire Brigade in 1990 from the Dorset Fire Brigade he became a middle manager and served as a Borough Commander. In 2013 Phil participated in the UK FRS National Operational Guidance Group's Incident Command project. This included a decision-making focused research project with Dr Sabrina Cohen-Hatton that led to the introduction of 'decision controls' to the UK FRS decision-making model. In 2015 he gained a BSc Psychology degree from Birkbeck College, London University and in 2016 an MSc in Social Science Research Methods (Distinction) from Cardiff University. He is currently a PhD research student at Cardiff University in the School of Psychology developing a behavioural marker system for UK FRS incident commanders.

Topics

How to improve incident command? The UK FRS faces a number of challenges in relation to incident command (CFOA, 2015) such as how does the UK FRS improve incident command despite the reducing opportunities for its incident commanders to gain experience? For example, the number of fires attended by English Local Authority FRSS in 2013-14 was the second lowest ever recorded and represented a 64% reduction compared to 2003-04 (DCLG, 2015).

Part of the answer for the first question could be to increase the number of opportunities to practice incident command, but the fidelity of any exercise can never reach that of a genuine emergency response (Orasanu & Connolly, 1993). Yet improvements to incident command may not be made just by individuals practicing as that may reinforce poor performance. The nature and type of training required must be the equal of the learning required by incident commanders. This leads to a second, more fundamental question for the UK FRS.

How to be a good incident commander? The UK FRS has a set a national occupational standards, within which are some that relate to incident command: WM7, EFSM2 & EFSM1 (Skills for Justice, 2014; 2013; 2013). Each one of these lists a set of performance criteria and knowledge and understanding statements, but they are outcome-based, e.g. incident commanders must be able to '…make decisions that continue to minimise risk and maximise progress towards… [the incident commander's]…objectives..' (Skills for Justice, 2014) and know and understand '…how to make and apply decisions based on risk assessment…' (Skills for Justice, 2014, p). As such they do not detail how the incident commander may make a decision.

However, the UK FRS has published a foundation document for incident command, which underpins its national incident command policy guidance (National Operational Guidance Programme, 2016). The decision making section within the foundation document is based upon the empirical research of Cohen-Hatton, et al, (2015) and Cohen-Hatton and Honey (2015). It provides guidance on intuitive and analytical decision making and introduced the concept of 'Decision Controls' (Cohen-Hatton & Honey, 2015). These controls addressed an issue identified by other studies, e.g. Groenendaal and Helsloot (2016) who identified the need for incident commanders to reflect upon the impact of their decisions. This section also refers to the Joint Decision-making Model used by the emergency services when dealing with multi-agency, complex and major incidents. Making decisions within this context is yet another challenge for UK FRS incident commanders.

Overall, the foundation document highlights the technical and non-technical skills (termed command skills) required of an incident commander. At present the command skills are based upon a core set of non-technical skills (Flin, O'Connor, & Crichton, 2008), but recent research has identified a prototype set of UK FRS Command Skills (Butler, 2016). Ongoing research is developing the prototype set of skills and designing a corresponding behavioural marker system. However, Rake and Njå (2009) and Groenendaal and Helsloot (2016) have highlighted other command skill-based matters in need of investigation, e.g. the need to ensure decisions are implemented and why incident commanders are reluctant to change an existing plan.

Conclusion

The research to date has added to the scientific knowledge of what incident commanders do and how it may be done well. In doing so it has extended the learning needs of UK FRS incident commanders and, notwithstanding that more research is required, these needs must be met. This ultimately leads back to the first question about improving incident command and the provision of training as one means to achieve that. It is not just about practice, but new training has to be developed to meet the evolving learning needs of incident commanders as a result of research, in particular those associated with UK FRS command skills, including decision making.

AVIATION COMMAND DEVELOPMENT CHALLENGES

Presenter: Garin Underwood

Garin is a commercial airline pilot with an academic background in Human Performance and Sports Science, and 6 years in the teaching profession. For the last 3 years he has been delivering professional development courses into aviation, with a focus on high performance training strategies and technology solutions to support evidence-based training initiatives. Close connections into elite sport, education, the fire service and medical profession has led to involvement in working groups and conference speaking to support cross industry safety initiatives.

Garin was educated at Loughborough University and holds a BSc (Hons) degree in Sports Science, and PGCE (Cert Ed). His professional life started as a lecturer in Physical Education at Exeter College. He joined British Midland (bmi) in 1995 as a sponsored cadet, gaining his ATPL/IR from Oxford Air Training School in 1996. He now has 14000 flying hours and ratings on Boeing 737, 777, 787 and Airbus 320, 330, 340, 350 aircraft. He is currently a Senior Training FO and SME at Virgin Atlantic, based at London Heathrow on the Dreamliner (Boeing 787) fleet.

Topics

Pilot training has changed little in decades and there is a growing acceptance that it is out of date. From a human performance perspective, significant mismatches exist between research-based recommendation and actual practice. In recent years the airline industry has undertaken an extensive review of current practice and recognizes the need to change the way pilots are trained to reflect the modern operation.

Changes are proposed on the understanding that the training of non-technical skills need to become fully integrated in to the performance environment. This presents a significant challenge to the training communities and questions traditional 'standalone' approaches to CRM and Human factors. It demands a broader repertoire of training skills and high levels of subject matter expertise in the competencies and behavior markers that equate to optimum performance. In the meantime, whilst aviation questions its value, other safety critical industries look to aviation style CRM as the model to follow.

Evidence Based Training (EBT) has emerged as the best solution. A significant amount of work has gone into the research and development of EBT. From a pedagogical perspective, it represents a far superior approach to current regulatory driven training practice, yet progress is stalling. Some are progressing with the organizational and administrative changes required, but few if any, are facing up to the challenges it presents to the training community. Significant cultural and organizational barriers add to the challenge.

Whatever the eventual solution, cross industry collaboration will surely speed the journey and improve potential outcomes. Since EBT is not a new concept, plenty of expertise and experience exists outside the industry. The Fire Service presents an interesting case in point. Ahead of most in facing up to cultural issues, supporting trainer development and involving the academic community, it is addressing key issues that holds others back. The focus on NDM provides real structure and purpose. Its academic rigour and proven effects on outcomes in risk environments makes sense of NTS and their dynamic interactions and is paving the way for more effective training strategies.

NDM is evidently a key point of common ground that could connect all risk professions, but currently in aviation its profile is low. Its growing focus on 'startle factor' is a response to current biggest cause of accidents. The sudden breakdown of mental processes 'in the moment'. If only the industry looked outside itself it would find extensive research in the Performance Sciences. Perhaps a training input that addressed the dynamic interaction between stress and NDM would provide the ultimate performance model for risk environments? Specifically, how does NDM change when stress reactions start to shut down the DM process? What coping strategies would be most effective to recover DM capacity? And crucially, how best to train for it?

In the meantime, it would be interesting to know the extent to which decision-making training interventions are changing front line behaviour and performance in the real world. Empirical evidence of positive change would present a powerful example to other industries.

MILITARY JOINT HEADQUARTERS AND DECISION MAKING IN COMPLEX OPERATIONS

Presenter: Rob Hutton

Rob Hutton is a decision researcher and cognitive systems consultant with over 20 years' experience. Rob started Trimetis Ltd with two former BAE Systems colleagues who now conduct cognitive systems research and consultancy for the MOD and other organisations trying to improve cognitive work in complex, high risk environments.

Topics

Command and control in military organisations is a complex endeavour. Headquarters and command structures exist to enable coordinated actions at the sharp end of the organisation by providing orders and guidance based on analysis and planning against the mission and intent from 'above', and monitoring and adjusting actions based on information about how the situation is changing, with or without intervention. From a small unit leader operating within an area that he she can directly see to the commander of a joint (tri-service), multinational, and multi-agency organisation, understanding and decision making present different challenges. Expertise in small unit leaders can be honed by paper-based tactical decision games as well as relatively low cost small unit field exercises. However expertise at the art and science of operational level command is arguably an order of magnitude more challenging. Situational complexity presents emergent, unpredictable, multi-faceted problems, and internal organisational dynamics present all manner of 'social' challenges. When a HQ exercise takes a year to organise and tens of millions of pounds to put on, what are the alternatives for problem-based, experiential learning and development to improve adaptive thinking and decision making in senior leaders?

We have spent a year exploring the implications of NDM-based training and development approaches to the development of "adaptive expertise" (Bohle Carbonel et al., 2014). Having identified six fundamental training principles for developing the cognitive skills that underpin adaptive performance, we are currently evaluating them in two contexts: 1) a "quick win" effort to explore the impact of interventions in a joint operational headquarters exercise, and 2) a longer term effort to evaluate the six training principles and their effectiveness at improving adaptive performance in a more controlled environment, but demonstrating generalisation to the military operational challenge of preparing decision makers or complex and chaotic situations.

The key training challenge for this military works relates to the challenge of developing decision makers who are embedded in an organisational structure developed to manage uncertainty and to be adaptive, but where individual expertise and experience is not enough to make effective decisions in multi-faceted, emergent problem situations that are, arguably, beyond the wit of one "heroic commander". Can scenario/problem-based solutions be applied to contexts where operational complexity emerges over days and weeks, rather than minutes and hours? What aspects of expertise should be the target for training and development for these "expert-generalists"?

CONCLUSION

The topics described above will be presented in brief presentations by the presenters identified. Following each presentation, audience members will be asked to respond to particular issues or questions relating to the presentations, including sharing their experiences, approaches, and related research with respect to the issues raised in each topic. Following the presentations, we have identified some specific challenge questions that we would like to discuss with the audience's input and participation, based on mutual interests and experiences.

Applied Training & Development Questions

- How do we prepare decision makers for when their cognitive capabilities are impaired by environmental conditions (e.g. hypoxia, heat stress, depressurisation, etc.)?
- How do we evaluate training impact on cognitive performance in complex operational decision making contexts, particularly process versus outcome measures?
- Does training expert perceptual and sensemaking skills in tactical scenarios scale up to being an adaptive expert in larger, complex, multifaceted problem scenarios at the operational level of performance?

Macrocognitive Questions

- Is RPD an appropriate model for decision making and problem solving in the variety of novel scenarios experienced by the Fire & Rescue Services (given that fire is one of only a multitude of hazardous situations in which the FRS are required to act)?
- Is developing individual expertise enough for decision making in the complex scenarios of joint operational level headquarters? Is the notion of an "expert generalist" a reasonable target for training and development?
- If accelerated expertise is an objective of training/learning/development, how far does individual expertise take us when we become a decision maker in a larger organisation where the complexity of operational and strategic level challenges exceed the capacity for individual expertise?

We expect to spend over half the session in panel and audience discussions. The topic presenters will all be prepared to discuss the specific questions, ideas and potential solutions for questions that fall directly out of their own topics, but it is expected that there will be audience members who are struggling with similar or related issues which we will attempt t surface and discuss.

A CALL TO ARMS

In order to wrap up the panel discussion, the panel would like to invite participants to share their contact information in order to participate in a UK-led Community of Interest for developing cognitive skills for high risk work, particularly for colleagues in the UK and Europe. If there is enough initial interest, we will organise an ad hoc meeting during the conference to discuss the nature of future interactions.

REFERENCES

Bohle Carbonell, K., Stalmeijer, R. E., Könings, K. D., Segers, M., and Van Merriënboer, J. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational Research Review* 12, 14-29.

Butler, P.C. (2016). *What are the non-technical skills of UK Fire and Rescue Service Incident Commanders?* (Unpublished masters dissertation). Cardiff University, Cardiff.

Chief Fire Officers' Association. (2015). *The future of incident command*. Chief Fire Officers' Association.

Cohen-Hatton, S.R., Butler, P.C., & Honey, R.C. (2015). An investigation of operational decision making in situ: Incident command in the U.K. Fire and Rescue Service. *Human Factors*, *57*, 793 – 804

Cohen-Hatton, S.R., & Honey, R.C. (2015), Goal-orientated training affects decision-making processes in virtual and simulated fire and rescue environments. *Journal of Experimental Psychology: Applied*, *21*, 395 - 406

Department for Communities and Local Government. (2015). *Fire statistics monitor: England April 2013 to March 2014.* Retrieved July 15 2016 from Gov.UK website: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325696/Fire_Statistics_Monitor_April_2013_to_March_2014_final__3_.pdf

Flin, R., O'Connor, P., & Crichton, M. (2008) *Safety at the Sharp End a guide to non-technical skills*. Ashgate: Aldershot

Groenendaal, J., & Helsloot, I. (2016). A preliminary examination of command and control by incident commanders of Dutch fire services during real incidents. *Journal of Contingencies and Crisis Management*, *24*, 2 – 13

National Operational Guidance Programme. (2015). *The foundation for incident command* (2nd ed.). Retrieved June 7 2016 from National Operational Guidance Programme website: http://www.ukfrs.com/Pages/Guidance-Catalogue.aspx?guidanceid=141

Orasanu, J., & Connolly, T. 1993. The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3-10). Norwood: Ablex.

Rake, E.L., & Njå. (2009). Perceptions and performances of experienced incident commanders. *Journal of Risk Research*, *12*, 665- 685

Skills for Justice. (2013). SFJ EFSM1 Provide strategic advice and support to resolve operational incidents. Retrieved June 25 2016 from the National Occupational Standards website: http://nos.ukces.org.uk/PublishedNos/SFJEFSM1.pdf#search=EFSM1

Skills for Justice. (2013). SFJ EFSM2 Lead, monitor and support people to resolve operational incidents. Retrieved June 25 2016 from the National Occupational Standards website: http://nos.ukces.org.uk/PublishedNos/SFJEFSM2.pdf#search=EFSM2

Skills for Justice. (2014). SFJ FRS WM7 Lead and support people to resolve operational incidents. Retrieved June 25 2016 from the National Occupational Standards website: http://nos.ukces.org.uk/PublishedNos/SFJWM7.pdf#search=WM7

Applying NDM to Change Behaviours and Persuade

Panel Chair: Laura MILITELLO[a],
Panelists: Emilie M. ROTH[b], Devorah KLEIN[c], B.L. William WONG[d,] and Gary KLEIN[e]
[a]*Applied Decision Science, LLC*
[b]*Roth Cognitive Engineering*
[c]*Marimo Consulting, LLC*
[d] *Middlesex University*
[e]*ShadowBox, LLC*

OVERVIEW

Naturalistic decision making (NDM) methods and models have been applied to a range of products and interventions across many domains. This panel will reflect on how the applications we create change behaviors and persuade, both intentionally and unintentionally. We have assembled a diverse set of panelists. Emilie Roth is a cognitive engineer with extensive experience studying and designing for large-scale systems such as nuclear power plants, railways, and military command and control. Dr. Roth will discuss how technology shapes behavior and persuades. Devorah Klein is a cognitive psychologist and designer who leads Marimo Consulting, LLC, a firm focused on the design of healthcare products, devices, and education to support both patients and clinical staff. Dr. D. Klein will share her perspective on how to support patients in making changes in order to realize long-term benefits. William Wong is a professor of human-computer interaction and leads the Interaction Design Centre at Middlesex University. Dr. Wong will share lessons learned from a series of studies exploring how technology design changes behavior. Gary Klein is a founder of the NDM community and has described many of the models that inform NDM research. Dr. G. Klein will share his insights into how NDM can complement and extend the choice architecture approach popularized by behavioral economics. Note, we also have contacted a fifth potential panelist, David Halpurn from The Behavioral Insights Team. Dr. Halpern was not able to commit to participation at the time of this submission, but we are hopeful that he or someone from his firm will be able to join the panel and discuss how the UK Behavioral Insights Team is applying choice architectures.

EMILIE ROTH, ROTH COGNITIVE ENGINEERING

Biosketch

Roth is owner and principal scientist of Roth Cognitive Engineering. She is a cognitive psychologist by training (Ph.D. University of Illinois, Urbana-Champaign), and has participated in cognitive analysis and design in a variety of domains including nuclear power plant operations, railroad operations, military command and control and healthcare. She is a fellow of the Human Factors and Ergonomics Society, is an Associate Editor of the Journal of Cognitive Engineering and Decision Making, and serves on the editorial board of the journal Human Factors.

Using Technology to Shape Behavior: Examples of Intentional and Unintentional Consequences

There are a variety of ways to guide the choices that people make and how they respond to situations. For example, the book ‘Nudge’ (Thaler and Sustein, 2008) argued for explicit design of the choice context to influence the choices that people make. A simple example is to place healthier food so that it is easier to see and reach for than less healthy options.

Another way to influence people’s choices and behaviors is through technology. Technology can be used to influence choices, afford new opportunities, or preclude certain behaviors. The impact of technology on behavior can be intentional – with designers explicitly intending to change behavior. Technology can also inadvertently alter user behavior. In this talk, I will provide some examples of these different cases drawn from my own work and those of others.

One example of using technology to explicitly influence behavior comes from the work of Chen and colleagues (21015). They were interested in how to present summary risk information to influence users to make less risky decisions with respect to cyber security when downloading phone apps. The study showed that people’s app-installation decisions can be affected by how a risk summary score is presented. Less risky apps tended to be chosen over more risky ones when the score was framed in terms of amount of safety rather than amount of risk. Thus, traditional concepts of framing can be applied in the design of user interfaces to persuade individuals to make less risky decisions with respect to cyber security.

Technology can also create new affordances that encourage individuals to alter the cognitive strategies they use. For example, in one nuclear power plant application, the introduction of computerized procedures allowed a team to simultaneously use multiple diverse cognitive strategies for responding to an emergency – making it possible to more rapidly identify situations where the procedures were 'off track' (O'Hara and Roth, 2005). While this was a positive outcome, it was not explicitly anticipated by the system designers.

Finally, technology can alter or preclude certain behaviors. One example comes from the railroad domain. Interviews and observations conducted at sites where new positive train control (PTC) were being installed, revealed that PTC caused locomotive engineers to alter both their monitoring and train handling strategies. For example, locomotive engineers modified how they operated the train to conform to the automatic braking profile of the PTC system. This led to more conservative and slower train handling. This was not an explicit intent of introducing PTC.

References

Chen, J., Gates, C. S., Li, N. and Proctor, R. W. (2015). Influence of Risk/Safety Information Framing on Android App-Installation Decision. Journal of Cognitive Engineering and Decision Making, 9, 149-168.

O'Hara, J. M. and Roth, E. M. (2005). Operational concepts, teamwork, and technology in commercial nuclear power stations. In Clint Bowers, Eduardo Salas and Florian Jentsch (Eds) *Creating High-Tech Teams: Practical guidance on work performance and technology*. (pp. 139-159). Washington, D. C.: American Psychological Association.

Roth, E. M, Rosenhand, H., and Multer, J. (2013). *Using Cognitive Task Analysis to Inform Issues in Human Systems Integration in Railroad Operation.* U. S. Department of Transportation, Federal Railroad Administration, Office of Research and Development: Washington, D. C. (DOT/FRA/ORD-13/31)

Thaler,R. H., and Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. New Haven: Yale University Press.

DEVORAH KLEIN, MARIMO CONSULTING, LLC

Biosketch

As a cognitive psychologist and designer, Devorah Klein focuses on turning observations into design opportunities, particularly in healthcare. She has worked on projects in medication adherence, diagnostics, drug discovery, new business models of healthcare, hospital workflow, new care models for delivery, environments design, digital health, device design, drug delivery, and health insurance. Adherence programs include therapies for treating weight loss, diabetes, HIV, asthma, rheumatoid arthritis, multiple sclerosis, oncology, and erectile dysfunction. She has worked at several of the leading design and innovation consultancies, including IDEO, where she created the Adherence Design offering.

Thoughts on persuasion

I'm the worst person to have on a panel about persuasion. Why? My philosophy of behavior change is, "I don't want to change people, I want to help people change themselves." I think the key to behavior change is to deeply understand the needs, wants, perspectives and goals of the people in question, and then, when appropriate, help them figure out how the proposed change meets those needs and goals and give them the tools to change. Persuasion at its worst is convincing people to engage in a behavior that is not aligned with their deepest goals and needs. Persuasion at its best helps people make decisions that are most in tune with them. That's the persuasion I'm interested in, which can get people to take medications that are unpleasant and engage in other activities with immediate pain for a long-term benefit.

So, what's the role of NDM in this? When most people think about persuasion, they think about marketing, trying to make us feel inadequate to get us to change. The best persuasion aligns people's goals with solutions to help attain them. That relies on a deep understanding of how they think, how they make decisions, their mental models, and their interactions with others. That's where NDM can play a critical role.

B.L. WILLIAM WONG, MIDDLESEX UNIVERSITY

Biosketch

Dr B.L. William Wong is Professor of Human-Computer Interaction and Head, Interaction Design Centre, at Middlesex University's School of Science and Technology. His research interest is in the representation design of information and the interaction of user interfaces to support decision-making in complex dynamic environments such as emergency ambulance command and control, and air traffic control. His current research investigates the problems of using visual analytics in sense-making domains such as intelligence analysis. He is recipient of over US$25 million in grants, and is or has been the project coordinator for several multi-national, research consortiums EU project VALCRI (FP7-IP-608142), CRISIS (FP7-242474), UKVAC jointly funded by HM Government and US DHS, EPSRC and others. Together with his colleagues and students, he has published over 100 scientific papers.

Technology can be a game changer

Technology can be a game changer - but it has to be thoughtfully designed. Leaving usability issues aside, the design of the technology need to address fundamental issues of how people carry out tasks. In a couple of studies we ran:

Study 1: Interactive Visualization for Low Literacy Users

In this study, we compared the performance of Low Literacy users against High Literacy users in their use of a standard design website vs a new and novel visualisation and interaction design called INVISQUE. This design is based on representing each piece of information as individual electronic cards that can be freely moved about on a display. Several cards can be freely moved about and laid out in spatial arrangements that are meaningful to the user, such as this card is important and will place it here, vs. this is not important, and we can leave it over there. This design was compared against a website designed by a national social welfare organisation. For our study we extracted the exact data and incorporated them into the INVISQUE system. Both LL and HL users were given the same task to find information about various social benefits available using the website, or using INVISQUE. We found that by making information tangible and tactile by using the INVISQUE interface, low literacy users' performance had significantly improved against high literacy users. We postulated that this improvement by low literacy users when using the INVISQUE interface is possibly due to having technology that is more compatible with the way they thought and reasoned. For example, in other studies that compared IQs between agrarian communities and urban communities show that urban communities do better on IQ tests that require a greater level of abstract thinking skills - something which agrarian communities are less practised at. Hence, with INVISQUE which enables a very tangible way of interacting with information, it enabled LL users to consider information in a much more tangible manner - perhaps LL users have less advanced abstraction skills and therefore INVISQUE provided a more compatible form of support. [Kodagoda, Neesha, B. L.William Wong, Chris Rooney, and Nawaz Khan. (2012). "Interactive Visualization for Low Literacy Users: From Lessons Learnt to Design." Pp. 1159–68 in CHI '12 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.]

Study 2: Using Interactive Visual Reasoning to Support Sense-Making

This study focused around identifying 'influencers' within a community of practice, from the perspective of identifying implications for design to support investigators data exploration, reasoning process, and the assembly of evidence. For this study again we used a system called INVISQUE (INteractive Visual Search and QUery Environment) loaded with publications from the ACM Digital Library. Here the understanding how high-level problem of identifying key players within a domain can translate into lower-level questions and interactions. This, in turn, has informed our understanding of representation and functionality needs at a level of description which abstracts away from the specifics of the problem at hand to the class of problems of interest. [Kodagoda, Neesha, Simon Attfield, B. L.William Wong, Chris Rooney, and Sharmin Tinni Choudhury. (2013). "Using Interactive Visual Reasoning to Support Sense-Making: Implications for Design." *IEEE transactions on visualization and computer graphics 19*(12):2217–26.]

Study 3: Tactile reasoning: Hands-on vs. Hands-off – What's the difference?

In another study, we gave two groups of students a set of physical cards, where each card was printed with one piece of information - e.g. an animal, food it eats, its habitat, etc. When assembled, the cards can reveal a story about food chains in a given type of habitat. The first group was allowed to use their hands to freely move the cards around the desk to find the most suitable arrangement that could tell an interesting story, while the second group were not allowed to use their hands to touch or move the cards. Using a think aloud procedure, we asked participants to tell us what they were thinking as they thought about the cards and reasoned through to an explanation presented by the cards. Using verbal protocol analysis, we counted the number of utterances made by each participant that related to some form of insight or understanding. We observed that those who were free to touch and move their cards, made 99 utterances of insights vs 50 utterances by those who were not allowed to touch or move the cards. [Takken, S., & Wong, B. L. W. (2015). Tactile reasoning: Hands-on vs. Hands-off - what's the difference? *Cognition, Technology & Work, 17*(3), 381-390.]

Study 4: The Effect of Layout on Dispatch Planning and Decision Making

In earlier work, we found that if a display was designed in a manner that was compatible with the cognitive strategies of the task, performance can significantly improve. In Wong et al 1998 (a somewhat historical paper) we re-organised the order of the information showing ambulance stations from a semi-alphabetical order to a display that more closely corresponds with the geographical disposition of the stations. This led to a 40% improvement in response times to allocate ambulances between neighbouring stations. [Wong, W. B. L., O'Hare, D., & Sallis, P. J. (1998). The Effect of Layout on Dispatch Planning and Decision Making. In H. Johnson, L. Nigay, & C. Roast (Eds.), People and Computers XIII, HCI '98 Conference (pp. 221-238). Sheffield, UK: Springer, in collaboration with the British Computer Society.]

What we learnt?

(1) understand the fundamental ways in which the tasks we hope to support works. Identify the cognitive strategies that underlie the tasks, rather than the processes or the procedures that they have been trained to perform the task. (2) to develop technologies that will enable or persuade those behavioural changes, we need to develop a design concept that enables one to interact and manipulate information with fluidity, i.e. the ease by which a system can be used to express the variability of our thinking processes, yet provide support to enable rigorous preparation and treatment of our conclusions. By rigour we mean the ability of the system processes and results to withstand interrogation that ensures the validity of our conclusions (Wong, B. L. W. (2016). Fluidity and Rigour: Designing for Intelligence Analysis. In B. Akhgar, P. S. Bayerl, & F. Sampson (Eds.), Open Source Intelligence Investigation: From Strategy to Implementation (pp. To be published): Springer.)

GARY KLEIN, SHADOWBOX, LLC

Biosketch

Gary Klein, Ph.D., is known for (a) the cognitive models he described, such as the Recognition-Primed Decision (RPD) model, the Data/Frame model of sensemaking, the Management By Discovery model of planning in complex settings, and the Triple Path model of insight, (b) the methods he developed, including techniques for Cognitive Task Analysis, the PreMortem method of risk assessment, and the ShadowBox training approach, and (c) the movement he helped to found in 1989 — Naturalistic Decision Making. The company he started in 1978, Klein Associates, grew to 37 employees by the time he sold it in 2005. He formed his new company, ShadowBox LLC, in 2014. The five books he has written, including S*ources of Power: How people make decisions*, and *Seeing What Others Don't: The remarkable ways we gain insights*, have been translated into 12 languages and have collectively sold more than 100,000 copies.

Using NDM to change minds and change behavior

NDM practitioners act as Human Performance Specialists, applying what we know to influence and improve performance. We can use several levers, including training, organizational design, device design, incentives, selection, behavioral engineering, checklists, decision aids, and Choice Architectures.

What have we learned from NDM research that strengthens our capability for using these levers in order to influence and improve performance?

The field of Behavioral Economics has drawn from research on the Heuristics and Biases paradigm to design nudges that take advantage of peoples' biases.

But Behavioral Economics is fairly limited. It seems to have evolved and expanded into the use of Choice Architectures, essentially becoming applied social psychology. In its current form it seeks to influence behavior through strategies such as social proof, as well as through manipulating biases. The work of Cialdini and other social psychologists is now highlighted in the Choice Architectures being designed.

The NDM community should be able to further enrich our tools for persuasion and influence by adding applied cognition to the mix.

My own research and practice has centered on several avenues for taking advantage of applied cognition — what we have learned about the way people think and decide — in contrast to trying to exploit weaknesses such as biases.

a. The RPD model suggests that we can improve decision quality, not through checklists and procedures, but through regimens for enhancing expertise.
b. The PreMortem method can change team dynamics to allow authentic dissent to be voiced rather than hidden.
c. The ProMortem method is a visioning technique that helps to motivate and guide teams.
d. The Causal Landscape tries to balance the tendency to oversimplify and the tendency to over-analyze our assessments of what has happened and how to proceed.
e. More and more our training efforts center around creating mindset shifts in addition to helping trainees add to their playbooks.
f. Ben Schneiderman and I have been exploring system design concepts to foster and support expertise rather than seeking to supplant it with Artificial Intelligence.

John Schmitt and I have identified mental pressure points to allow police and military to gain voluntary compliance as opposed to coercive compliance.

DOCTORAL CONSORTIUM PRESENTATIONS

Tornado Risk Literacy: Beliefs, Biases, and Vulnerability

Jinan N. ALLAN[a, b], Joseph T. RIPBERGER[a], Vincent T. YBARRA[a, b] and Edward T. COKELY[a, b]

[a]*National Institute for Risk and Resilience, University of Oklahoma*

[b]*Department of Psychology, University of Oklahoma*

ABSTRACT

Risks and crises that are relatively rare, severe, and hard to anticipate present unique high-stakes communicaiton challenges (e.g., disease outbreaks, terrorist attacks, natural disasters). We present a study designed to detect and monitor key factors likley to shape responses of individuals during natural hazard events. Working in the domain of tornado risk literacy, we investigated over two thousand Oklahomans via a longitudinal risk perception survey. Preliminary findings documented connections between dangerous tornado beliefs and decision vulnerabilities. Psychometric and structural process models estimated the extent to which common misunderstandings (i.e., belief in tornado myths) were predictable products of domain general decision making skill (e.g., risk literacy as measured by the Berlin Numeracy Test; see RiskLiteracy.org) and two types of domain specific knowledge. Results indicate that these new instruments are likely to provide rough, yet robust indicies of tornado decision vulnerability, providing a method of monitoring changes and preparedness of diverse individuals in at-risk communities, which can be integrated with well-established naturalistic decision making approaches.

KEYWORDS

Decision making; Earth and atmospheric sciences; Judgment and decision making

INTRODUCTION

Previous literature indicates that individuals often hold misconceptions about extreme natural hazard events, which in part reflects the complexity inherent in real world risk phenomena and high-stakes risk analysis. For example, many people struggle to understand what appears to be very basic weather risk communications (Klockow, Peppler, & McPherson, 2014; Edwards, 2013). To illustrate, consider the statement “There is a 30% chance of rain.” Past studies have found participant interpretations of this statement that ranged in both scope and complexity, such that many well-educated individuals and professionals failed basic comprehension questions (Gigerenzer, Hertwig, Van Den Broek, Fasolo, & Katsikopoulos, 2005). Some participants believed it would rain for 30% of the day (eight hours), others that it would rain in 30% of the area (i.e., across 15 miles in the 45 square miles of San Francisco), and others still that 30% of weather forecasters think it will rain (i.e., 3 of 10 meteorologists predict rain). Unfortunately, none of these all-too-common misinterpretations are correct. In fact, to say there is a 30% chance of rain could indicate that the weather forecasters are 50% confident that *if* rain occurs, it will cover roughly 60% of the area, however it could also indicate that forecasters are 100% confident that *if* rain occurs, it will cover 30% of the area. This calculation is more formally called the Probability of Precipitation (PoP; See NWS, n.d.).

The National Weather Service provides forecasts for standard daily weather alerts (e.g., probabilities of precipitation), as well as predictions for extreme weather events (e.g., tornados, severe thunderstorms, etc.). Unfortunately, the general public is often unaware of how these extreme weather risks are communicated. Consider for example the many tornados and related extreme weather events (e.g., thunderstorms, hail, strong winds, etc.) that are common in the spring (i.e., April, May, June) across the southern plains of the United States. To help protect people and property from these events, the National Weather Service is charged with issuing tornado *watches* when the conditions are likely to support tornado formation, and there is some chance that a tornado could occur in the next few hours. Conversely, a tornado *warning* is issued when a tornado is *imminent*.

In the face of deadly natural hazards and extreme weather events like these, individuals leverage information from many different sources (e.g. news broadcasts, conversations with friends, applications on their smartphone, etc.) with varying degrees of reliability, and integrate these various sources of information in the service of decision making. Ideally, individuals would prioritize the information they received from reliable sources (e.g. the National Weather Service), and would accurately integrate that information (e.g., Skilled Decision Making; see Cokely et al., in press). Unfortunately, research indicates that many people rely primarily on past experiences and “folk science” information (Klockow et al., 2014). There are many reasons this is unsettling, but chief among them is the fact that these “folk science” beliefs are often rooted in myths or misconceptions about weather trends, which may lead to dangerous decisions (e.g., deciding to try to drive away from a tornado instead of sheltering in place).

Clearly, there are many factors that shape people's ability to adequately respond to emerging risks, like natural hazards and tornados (Klein, 2008). As such, public and private institutions charged with protecting people have called for increasing support from multidisciplinary scientists to be able to better estimate and understand the factors that shape decision vulnerability among diverse publics. What are the factors that inoculate individuals from these decision vulnerabilities and how prevalent are those factors among sub-poplutions who are at-risk and under threat?

The aim of this paper is to begin to map some of the connections between knowledge, skills, beliefs, and biases that shape individual decision making in the face of natural hazard dread risks—i.e., deadly tornados. One factor that has been identified to impact natural hazard decision making are beliefs, especially as they pertain to various dangerous folk science beliefs, or *myths.* For example, Klockow et al. (2014) studied "folk science" beliefs about tornados via interviews collected after the 2011 tornado outbreak in Alabama and Mississippi. From these interviews, they found that there are commonly held myths or misconceptions that people believe about tornados that are not in accordance with the current research on meteorological sciences. An example of the most prominent myths include: (i) tornadoes are attracted to waterways, (ii) roadways can make a tornado stronger, (iii) hills and mountains can protect from tornados, such that cutting down trees and narrowing the landscape can give the tornado more room to move, and (iv) tornados will happen "there" but not "here." People often hold a place-based optimism bias, such that they believe the tornado might hit near them, but will not hit their home exactly (Klockow et al., 2014; Edwards, 2013).

An individual who is in danger from a tornado must make numerous complex decisions. In accord with the main features of naturalistic decision making, protecting oneself and family from loss of life and property involves ill-defined goals and tasks, uncertain and ambiguous information, competing goals, dynamic conditions, and high stake, stressful situations for multiple individuals (Klein, 2008). To thoroughly investigate how individuals respond to these types of naturalistic decisions, researchers would have to wait for a tornado, and then follow their participants through their decision making process. However, armed with some valuable insights into human decision making in the face of a crisis, we can begin to unpack how psychometric assessments might detect vulnerable populations.

The aim of the paper is to begin to assess and detect tornado decision vulnerability, so that we can begin to map and estimate the extent to which extreme weather misunderstandings predicts an individual's propensity to believe in dangerous myths. Theoretically, by understanding who is at risk for decision vulnerabilities, we will better understand what factors might inoculate them and we can also estimate overall decision vulnerability at a community and/or population level, further informing public risk communication policies. Natural hazard risk information could then be communicated more efficiently to empower individuals to make informed decisions consistent with their own goals and values. To take the next step, in accord with Skilled Decision Theory (Cokely et al., in press), we hypothesize that both domain general decision making skill (e.g., statistical numeracy) and domain specific skills (e.g., tornado knowledge) are likely to predict overall decision vulnerability (e.g., belief in tornado myths), such that general skills will have both direct and indirect effects on vulnerability (e.g., more generally skilled individuals will also acquire more accurate knowledge and understanding of weather phenomena, further reducing their vulnerability).

METHOD

Participants

The Oklahoma Meso-scale Integrated Socio-geographic Network (M-SISNet) project is a panel survey that measures perceptions of weather, as well as views on government policies and societal issues. This data is used to understand how perceptions and views might shape Oklahomans' use of water and energy, and extreme weather reactions. Over two thousand residents from around the state of Oklahoma are surveyed quarterly (N = 2528). Participants in this program are compensated for their time, and are asked roughly one hundred questions each quarter.

Procedure

Within the assessment battery, participants first answered basic demographic questions, including age, gender, race, employment, and family status. Following this, participants were given the Berlin Numeracy Test (Cokely, Galesic, Schulz, & Garcia-Retamero, 2012) and Schwartz (1997) numeracy items. These scales are the strongest predictors of risk literacy (i.e., one's ability to understand and interpret information about risk) and general decision making skill, far outperforming other measures of cognitive abilities, skills, and personality traits (Cokely et al., 2012). Participants were then assessed on their tornado knowledge and folk science beliefs (Klockow et al., 2014). Participants first answered four hazard knowledge questions, followed by five questions regarding belief in common tornado myths, and another four questions regarding knowledge of the tornado warning system.

RESULTS

Myth Vulnerability

Leveraging expertise from the National Weather Service, we created five true or false items that assess belief in tornado myths and are used to approximate tornado decision vulnerability (Edwards, 2013). With exploratory factor analysis, we found that the five myth items loaded significantly onto one factor, and chose to use an orthogonal (varimax) rotation, because the data is not highly correlated (Osborne & Costello, 2009). Then, using a two-parameter logistic model (2PLM) in Item Response Theory (IRT) which accounts for both item difficulty and discriminability, we removed one question that had poor psychometric properties. For this item ("You must be underground or in an aboveground tornado shelter to survive a violent tornado"), the probability of a participant answering the question correctly decreased, as the latent IRT ability factor increased (i.e., had negative discrimination).

The resulting four item unidimensional measure of belief in tornado myths was used as our criterion variable, as it is a rough approximation of tornado decision vulnerability. However, the items selected were not very difficult for this sample of Oklahomans' and this scale demonstrated large ceiling effects (many participants got all or most of the items correct; Table 1). The Item Characteristic Curves (ICC) and Test Information Function (TIF) further demonstrate this property as most of the information (70.94%) that this test provides about a participant's ability is found within the lower half of the latent ability scale (Figure 1).

Table 1. Items and the proportion answered correct.

		P(correct)
Tornado Myths		
M1	Tall buildings protect large cities from tornadoes [false]	0.96
M2	Mountains, rivers, and lakes do not protect nearby areas from tornadoes [true]	0.76
M3	When driving, you should not take shelter from tornadoes under a bridge or overpass [true]	0.87
M4	When sheltering in a house from a tornado, you should open all the windows to equalize the pressure inside and outside to prevent the house from exploding [false]	0.80
Hazard Knowledge		
H1	In Oklahoma, most tornadoes occur… [April, May, and June]	0.95
H2	In Oklahoma, most tornadoes move… [in a diagonal line, from southwest to northeast]	0.89
H3	In Oklahoma, most tornadoes are rated… [EF0 or EF1]	0.29
Warning Knowledge		
W1	To the best of your knowledge, who issues tornado WARNINGS in your local area? [National Weather Service]	0.90
W2	Now consider those occasions when tornado WARNINGS are issued by the National Weather Service in your local area and tornadoes actually occur. On average, how much time is there between when tornado WARNINGS are issued and when tornadoes occur? [less than 1 hour]	0.88
W3	To the best of your knowledge, approximately what percentage of tornadoes that occur in the area around where you live are preceded by a tornado WARNING that is issued by the National Weather Service? [65-85%]	0.17

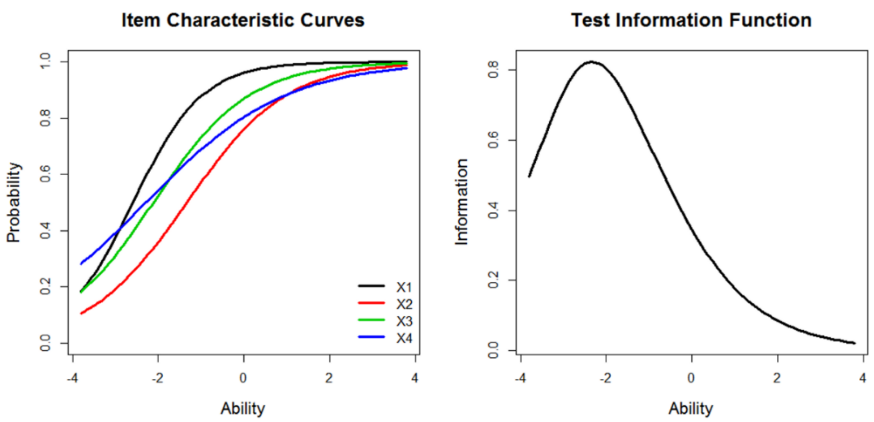


Figure 1. The Item Characteristic Curves (ICC) and Test Information Function (TIF) for Tornado Myth items gathered via Item Response Theory (IRT) Analysis.

Knowledge

Eight multiple choice items were included to assess domain specific knowledge about tornados. These questions were considered to be two separate latent factors: warning knowledge (four items) and hazard knowledge (four items). Roughly, warning knowledge measures knowledge of the tornado warning system (e.g., who issues tornado warnings, how often are they wrong, etc.), whereas hazard knowledge assessed knowledge of tornados themselves (e.g., direction of travel, typical tornado season, etc.).

Factor analysis demonstrated a unidimensional structure for both Warning Knowledge and Hazard Knowledge. IRT further demonstrated that the item "To the best of your knowledge, approximately what percentage of tornado WARNINGS that are issued by the National Weather Service in the area around where you live are followed by the actual occurrence of a tornado?" should be removed due to its negative discrimination. The correct answer to this question was roughly 40%, but any answer given between 30% and 50% was marked as correct (Iowa Environmental Mesonet, 2017).

Further, this test was low ability, as 85.47% of the Test Information Function was below the mean latent ability factor. It also had a broader range of test information, whereby Question W1 was easy for participants to answer, and Question W3 was difficult for participants to answer. Finally, Question W2 demonstrated high item discriminability (Figure 2). Taken together, these items can indicate some discrimination between participants' ability on tornado warning knowledge.

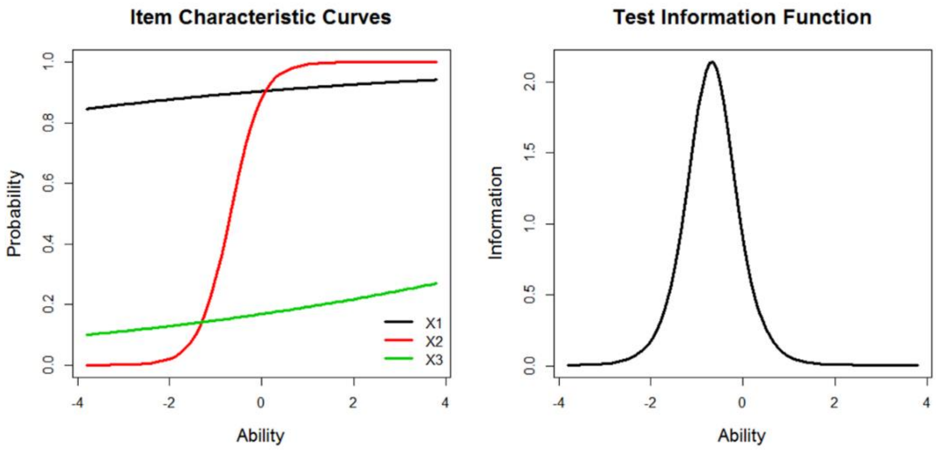


Figure 2. The ICC and TIF for Warning Knowledge.

Finally, among the Hazard Knowledge items, one item with negative discrimination was removed. The final three item test had low difficulty as 72.02% of the test information (TIF) was below the mean. Further, two of the items (Questions H1 and H2) demonstrated stronger discriminability, but low difficulty, whereas the final item (Question H3) had low discriminability, but high difficulty. This indicated that if a participant got this item correct, they are likely high-ability, and answered many other items correctly, too.

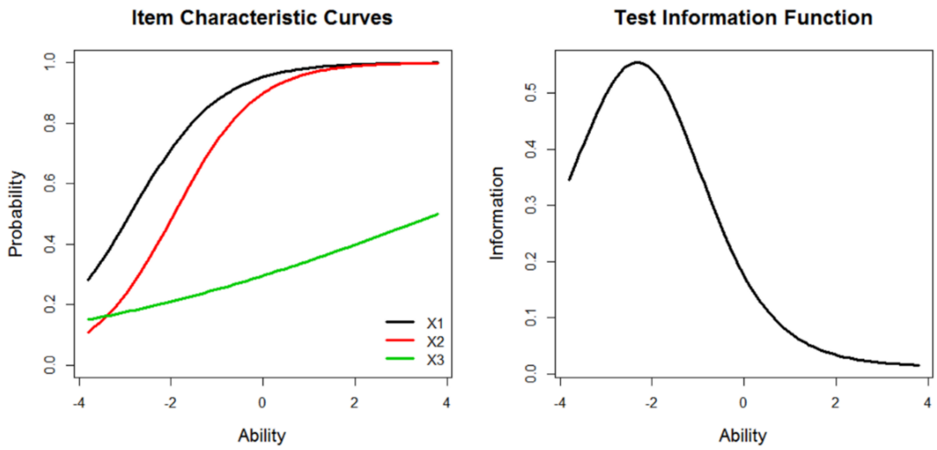


Figure 3. The ICC and TIF for Hazard Knowledge.

A Structural Process Model

Using sequential linear regression techniques, we next created a cognitive process model that demonstrated the causal implications of domain general and domain specific knowledge on tornado myth vulnerability. The model presented showed that both types of knowledge (warning and hazard) impact tornado myth vulnerability, such that having increased domain specific knowledge leads to decreased belief in myths. Similarly, domain general decision making skill (numeracy) both directly and indirectly impacts tornado myth vulnerability. Finally, demographics can impact tornado decisions. However, with the exception of age, this only occurred through the impact of numeracy. Thus, age was the only demographic variable that had a direct impact on decreased tornado myth beliefs (older participants were less likely to believe these myths).

The model demonstrated that our predictors significantly explained tornado myth vulnerability (R^2 =.032, $F(7, 1955) = 9.221$, $p < .001$). Both directly and indirectly, numeracy significantly predicted decreased tornado decision vulnerability (i.e., decreased myth beliefs; $\beta = .049$, $SE = .024$, $p < .001$). Warning Knowledge ($\beta = .085$, $SE = .022$, $p < .001$) and Hazard Knowledge ($\beta = .128$, $SE = .023$ $p < .001$) also significantly predicted decreased vulnerability. However, numeracy significantly predicted both Warning Knowledge ($\beta = .093$, $SE = .024$, $p < .001$) and Hazard Knowledge ($\beta = .124$, $SE = .023$, $p < .001$). As such, this is a partial mediation model. Finally, age was the only demographic variable to have a direct and significant impact on decision vulnerability ($\beta = .079$, $SE = .023$, $p < .001$; Figure 4). Though it is not perfect, the model presented here is a good proof of concept that general decision making skill and domain specific knowledge impacted beliefs that could lead to potentially hazardous decision making.

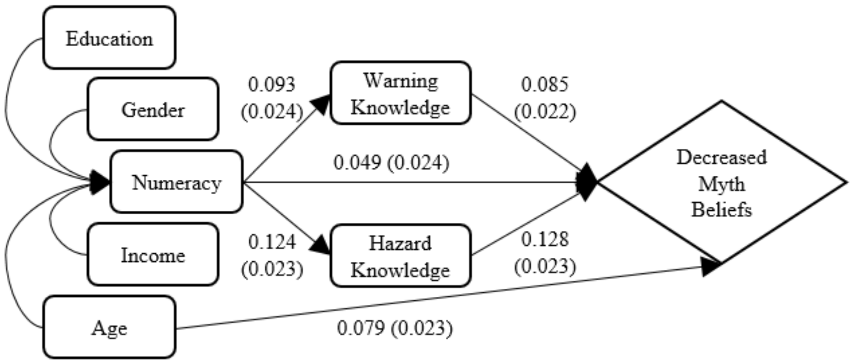


Figure 4. Process model of Tornado decision vulnerability.

DISCUSSION

The research presented here distilled and validated new measures of tornado understanding, documented a robust link between numeracy and tornado decision vulnerability, and estimated a quantitative structural process model demonstrating that both general decision making skill and domain specific knowledge (i.e., warning and hazard knowledge) predicts the likelihood of endorsing (and potentially spreading) deadly tornado myths. Specifically, results indicate that general risk literacy, as assessed by the Berlin Numeracy Test, partially inoculates against belief in tornado myths in part because more risk literate individuals tend to acquire a better understanding of two types of risk-relevant factors, namely (a) basic understanding of *the phenomenon itself* (i.e., hazard knowledge) and (b) functional understanding of *official warnings* (i.e., warning knowledge).

Taken together, the models and instruments presented here begin to demonstrate that (i) there are different types of tornado knowledge (i.e., warning and hazard), (ii) the participants in Oklahoma are largely experienced with tornados and did not fall trap to some of the myths provided (i.e., Oklahomans' know too much about tornados, so this test had large ceiling effects), and (iii) in order to assess tornado knowledge and vulnerability with higher fidelity, we will need to increase both the quantity and the quality of specific assessment items. Although robust, psychometric analyses suggest our sample of knowledge and myth items was too small, and probably also was not sufficiently representative of the different tornado myths that people believe. Results also suggest that the current assessment items may not be sufficiently representative of more influential decision vulnerability factors, namely the sources and types of information people actually use while making decisions about tornados (e.g., which sources do people trust and why?).

Moving forward, future research should consider other factors that might also predict tornado decision vulnerability. For example, what role do other values, beliefs, and goals have on the decisions people make when faced with a tornado warning (e.g., risk perceptions; see Slovic & Weber, 2002)? Further, when and why do at-risk individuals prioritize comfort or personal goals over taking shelter? Future research in this area should begin to estimate the extent to which people rely more on (i) analytic information (i.e., information provided by the National Weather Service and other news organizations) or (ii) past experiences, to inform their decisions about taking action and why (Hertwig, Barron, Weber, & Erev, 2004; Kahan, 2012). The model presented here focused on basic demographics and cognitive abilities, both of which are key factors known to shape vulnerably in many risk domains (e.g., health, wealth, well-being). While there are clearly some shortcomings of this model and the related instrumentation, analyses present some of the first evidence on the interplay of key factors that impact decision quality during actual tornado crisis situations.

Finally, the research here demonstrates that both general decision making skill (numeracy) and domain specific knowledge (warning and hazard knowledge) shape (mis)understanding of important and life-threatening factors in decision making. As we begin to learn more about how individuals (e.g., experts, emergency managers, and citizens) interpret tornado risk information, it will become increasingly important that we are able to provide information in such a way that promotes informed decision making. Theoretically, and given the extant research, there is good reason to think that rare tornado events may provide an efficient paradigmatic model for the investigation of broader dread risk literacy and decision-making resiliency during hard to anticipate crises. Although there are always notable limits, it seems likely that the same kinds of approaches, methods, and instrumentation that we will develop in the service of tornado risk literacy may be extrapolated to many other crises that are rare, severe and hard to predict (i.e., develop instruments that monitor and predict individual and community-level decision vulnerabilities in the face of disease outbreaks, terrorist attacks, or other dread events). In turn, psychometrically robust and validated measurements, can be used to facilitate the investigation of the cognitive processes involved in decision making under risk using more precise and detailed techniques (e.g., cognitive process tracing or cognitive task analysis; see Keller, Cokely, Katsikopoulos, & Wegwarth, 2010). Finally, we can use these investigated processes to create intervention strategies that have the propensity to inoculate individuals from believing false information associated with weather and weather safety. One intervention strategy is the use of simple decision aids (e.g., icon arrays). Overwhelming evidence shows that simple decision aids dramatically improve decision making outcomes because they circumvent the direct impacts of general decision making skills, by helping people develop more representative understanding of underlying risks and trade-offs (Cokely et al., in press; Garcia-Retamero & Cokely, 2013). As we continue to better understand how people make tornado decisions, we will be well-positioned to make recommendations as to how risk communications should be enhanced to benefit all individuals, especially those who are systematically disadvantaged (e.g., those living in mobile homes, those who are less numerate, etc.).

ACKNOWLEDGEMENTS

Partial financial support for this research was provided by grants from the U.S. National Science Foundation (SES–1253263), the US National Science Foundation EPSCoR program (IIA-1301789), and with substantial support from the National Institute for Risk & Resilience at the University of Oklahoma.

REFERENCES

Cokely, E.T., Feltz, A., Ghazal, S., Allan, J.N., Petrova, D., & Garcia-Retamero, R., (in press). Decision making skill: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (2nd Eds.), *Cambridge Handbook of Expertise and Expert Performance*. New York, NY: Cambridge University Press.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment & Decision Making*, *7(*1).

Cokely, E. T., Kelley, C. M., & Gilchrist, A. L. (2006). Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin & Review*, *13*(6), 991-997.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4*(1), 20-33.

Edwards, R. (2016). The Online Tornado FAQ. Retrieved January 28, 2017, from http://www.spc.noaa.gov/faq/tornado/

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, *22*(5), 392-399.

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making, 9*(1), 15-34.

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). “A 30% chance of rain tomorrow”: How does the public understand probabilistic weather forecasts?. *Risk analysis, 25*(3), 623-629.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science, 15*(8), 534-539.

Iowa Environmental Mesonet (2017). IEM Cow (NWS Storm Based Warning Verification). Retrieved January 29, 2017, from http://mesonet.agron.iastate.edu/cow/

Kahan, D. M. (2012). Cultural cognition as a conception of the cultural theory of risk. In *Handbook of risk theory* (pp. 725-759). Springer Netherlands.

Keller, N., Cokely, E. T., Katsikopoulos, K. V., & Wegwarth, O. (2010). Naturalistic heuristics for decision making. *Journal of Cognitive Engineering and Decision Making*, *4*(3), 256-274.

Klein, G. (2008). Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 456-460.

Klockow, Kimberly E., Randy A. Peppler, and Renee A. McPherson. (2015). Tornado folk science in Alabama and Mississippi in the 27 April 2011 tornado outbreak. *GeoJournal 79*(6), 791-804.

NOAA (2013). National Weather Service: Weather-ready nation roadmap. National Weather Service.

NWS (n.d.). FAQ – What is the Meaning of PoP. Retrieved January 28, 2017, from http://www.weather.gov/ffc/pop

Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, *12*(2), 131-146.

Richardson, D. S., Bidlot, J., Ferranti, L., Ghelli, A., Haiden, T., & Hewson, T., et al. (2011). Verification statistics and evaluations of ECMWF forecasts in 2010-2011 (ECMWF technical memorandum, Vol. 564). Reading, UK: European Centre for Medium-Range Weather Forecasts.

Ripberger, J. T., Silva, C. L., Jenkins-Smith, H. C., Carlson, D. E., James, M., & Herron, K. G. (2015). False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk analysis*, *35*(1), 44-56.

Ripberger, J. T., Silva, C. L., Jenkins-Smith, H. C., & James, M. (2015). The influence of consequence-based messages on public responses to tornado warnings. *Bulletin of the American Meteorological Society*, *96*(4), 577-590.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*, *127*(11), 966-972.

Slovic, P., & Weber, E. U. (2002). Perception of risk posed by extreme events. Presented at “Risk Management strategies in an Uncertain World,” Palisades, New York.

The Air Defence Task: Understanding the cognitions that underpin automation usage to support classification decisions in practice.

Chloë BARRETT-PINK[a], Laurence ALISON[a] and Simon MASKELL[b]
[a]*Critical and Major Incident Psychology, University of Liverpool, UK*
[b]*Electrical Engineering & Electronics, University of Liverpool, UK*

ABSTRACT

Introduction: The complexity of modern warfare is increasing requiring a synergy between man and machine to maintain effective decision making. This research aims to explore the influence of several cognitive traits upon automation usage decisions within a critical naval environment. **Method:** Firstly, a detailed literature review and small qualitative study to collect narratives from Royal Navy personnel was undertaken to understand the current decision making process that underpins the Air Defence Task. Following on from this, an experimental microworld will explore automation usage decisions in reality. A mixed methods approach to data analysis will be taken, using both qualitative and quantitative techniques. **Results:** We will extend the understanding into the influence certain cognitive traits can have upon automation usage decisions. We will then provide practical recommendations for how automated systems should be introduced into service to ensure they are utilised appropriately.

KEYWORDS

Decision making; General and miscellaneous; Military; Human-Centered design

INTRODUCTION

The concerns surrounding automation complacency (Singh, Molloy, & Parasuraman, 1993), skill degrade (Wiener & Curry, 1980), and 'out of the loop' phenomena (Endsley & Kaber, 1999) increase users' reticence to accept and utilise new systems to their full extent. However, it is clear that the complexity of warfare is continually increasing, resulting in threats that overwhelm even the most experienced operators. Fighting power is built on capability (Ministry of Defence, 2014), with a potential answer to remaining operationally superior lying in the appropriate use of automated systems and tools.

Naturalistic decision making (NDM) research has sought to explore cognitive work performed within complex sociotechnical environments and as a research domain has grown from work conducted with the US Military and Navy (Schraagen, Klein, & Hoffman, 2008). This body of literature has contributed greatly to furthering our understanding into, but not limited to, how complex decisions are made effectively within military and emergency services contexts. NDM research, at its heart, is concerned with unpicking our ability to perform under less than optimal conditions. As a result, NDM produces research that provides practitioners with useful recommendations as to how to further improve and support decision making in the field.

This thesis, utilising both student and practitioner samples, aims to explore the potential influence of several cognitive traits on automation usage decisions. The thesis provides findings that will extend the theoretical understanding of which cognitive factors can either facilitate or act as a barrier to appropriate automation usage decisions.

METHODOLOGICAL APPROACH

This research will take a mixed methods approach by utilising quantitative a priori hypotheses alongside qualitative post-hoc analysis of data collected using a micro-world simulation experiment and questionnaire battery. This mixed methods approach will produce empirically valid conclusions that stem from an ecologically valid context.

RESEARCH QUESTION 1: *WHAT* ARE THE DECISION MAKING STAGES WITHIN THE AIR DEFENCE TASK CONDUCTED BY ROYAL NAVY PERSONNEL?

Procedure

It is common practice at the start of any thesis to conduct a comprehensive review of the current literature on the topic in question. Therefore, to answer this first research question, an exhaustive review of the literature on maritime decision making was undertaken. This illuminated several seminal research projects and papers that had previously used NDM methods to explore military naval decision making. Notable projects identified via this review included the Tactical Decision Making Under Stress (TADMUS) programme conducted in the US (see Cannon-Bowers & Salas 1998) and Holt's work in 1988 looking at assessing the requirement for decision support tools in the UK. Both explored and produced decision making models of the air defence task conducted by naval personnel.

This current project began in 2015 and has a primary focus on the air defence task conducted by Royal Navy personnel. Therefore, to ensure that no significant changes to the decision making process had occurred since 1988, a short questionnaire was sent to Subject Matter Experts (SMEs). This vignette questionnaire aimed to explore the stages of the decision making process that current naval personnel are familiar with today.

Participants

There is limited availability of SMEs, not only because of generic issues with access to Naval personnel, but also because the number of individuals within the Royal Navy that could be considered experts of the air defence task is small. Seven SMEs responded to a questionnaire that was emailed out via a gatekeeper. The results were then thematically analysed. This qualitative method has previously been argued to allow the formation of patterns when exploring complex realities (Braun & Clarke, 2013).

Findings

Although limited by the small sample size number, the findings support previous models of the air defence process developed in both the UK and USA.

Implications:

By comprehensively exploring past literature into this area of decision making and confirming, through a small qualitative study, the stages of the air defence task, it has been possible to tailor the remainder of this thesis to explore gaps within the literature on this domain area.

RESEARCH QUESTION 2: *WHEN* WOULD AN AUTOMATED DECISION SUPPORT TOOL BE USED TO AID THE DECISION MAKING PROCESS DURING THE AIR DEFENCE TASK?

RESEARCH QUESTION 3: WHERE DO THE INDIVIDUAL DIFFERENCES LIE WITH THE DECISION TO USE AN AUTOMATED DECISION SUPPORT TOOL OR NOT? DOES A RELATIONSHIP EXIST BETWEEN CERTAIN COGNITIVE TRAITS (COGNITIVE FLEXIBILITY, NEED FOR CLOSURE, ASSESSMENT & LOCOMOTION AND IMPULSIVITY) AND AUTOMATION USAGE DECISIONS?

Two studies are proposed to answer these research questions. Study 1 will be a mixed design microworld experiment, using a student sample population. Study 2 will be an online questionnaire, sent to Subject matter experts (SME) and Royal Navy (RN) personnel via gatekeepers. Both studies will now be described in full below.

Study 1

A vast amount of research has explored the utility of specific decision support tools during complex decision making environments. For example, within medical (Goud et al., 2009) and military (Biros, Daly, & Gunsch, 2004; St John, Smallman, Manes, Feher, & Morrison, 2005; Cummings & Guerlain, 2007) environments. However, few have looked at the influence individual differences may have upon automation usage decisions. That is, the decision of the individual to choose to utilise such an automated tool. Therefore, this thesis aims to explore a range of individual differences and their potential impact upon the decision to use a generic decision support tool during the complex naval task of classifying radar tracks. To achieve this, a framework for micro-world experiments (Automatic Radar Classification Simulation, ARCS) has been developed to explore when individuals would select to use a generic automated tool to aid their classification decisions.

Participants:

A sample size of 69 participants will be recruited from the University of Liverpool to take part in the microworld experiment. This sample size was computed using GPower 3.1 (Faul, Erdfelder, Lang & Buchner, 2007) with parameters of a medium effect size (0.25), p=0.05 and alpha level 0.8. This sample size will ensure appropriate experimental power (to avoid making a Type II error).

Design:

A mixed-design experiment is proposed (see Table 1)- 3 (Automation background, between-groups condition: control group i.e. no automation, low automation group i.e. automated tool is compared to human operator and high automation group i.e. automated tool is compared to an automaton) x 2 (Accountability, within-groups condition: low vs. high).

Table 1. Experimental Design

		Automation Background (between-participants)		
		Control (A)	Low (B)	High (C)
Accountability (within-participants)	Low (L)	A_L (N=23)	B_L (N=23)	C_L (N=23)
	High (H)	A_H (N=23)	B_H (N=23)	C_H (N=23)

Procedure:

It is proposed to run participants in groups (5 or more, depending on practicality and availability of participants) to replicate working with other personnel around. Participants will be asked to take part in the study twice, with at least 2 weeks break in-between experiments. Participants will be allocated into one of three groups (A, B or C), they will complete the experiment <u>within the same</u> *Automation Background condition*, and <u>will experience both</u> the high and low *Accountability conditions*. Student participants will be taken through a 30-minute training procedure, consisting of a 15-minute presentation to immerse the participants in their task and explain the system and symbology, followed by 15-minutes of practice with the system itself. Following this, participants will complete a 30-minute radar classification task (see Figure 27) during which they will have the option to utilise an automatic classification tool up to 3 times (if in groups B or C).

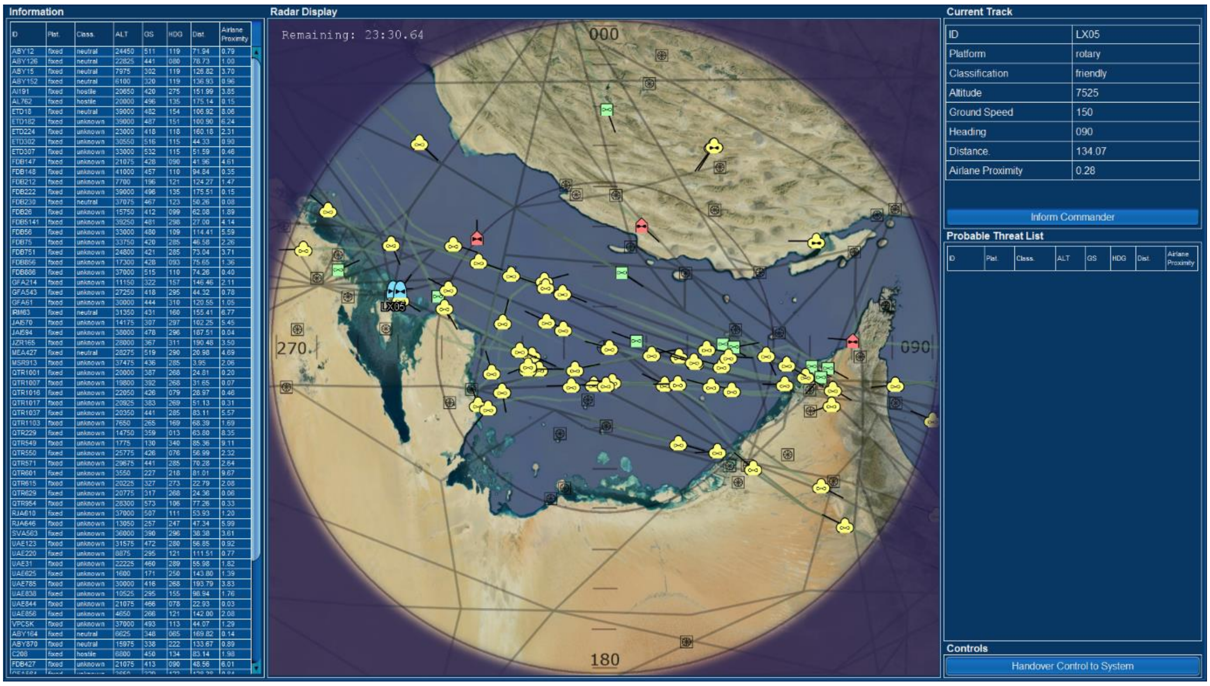


Figure 27 Screenshot of the main task within the Automatic Radar Classification Simulation (ARCS)

Measures:

- All mouse clicks made by participants will be recorded- i.e. all classification decisions and if and when the automated tool was selected. This log of data will allow future analysis to explore all decisions made by participants and, to some extent what information they were looking at to make those decisions.
- At 10 minute intervals, participants will be asked to complete quick NASA-Task Load Index ratings. This is intended to provide a continual gauge of participant workload throughout the task across 6 dimensions (mental demand, temporal demand, physical demand, frustration, performance and effort).
- Individual differences will be measured post- scenario via a questionnaire battery:
    - Cognitive Flexibility Inventory (CFI, Dennis & Vander Wal, 2010)
    - Need for Closure Scale (NFC, Kruglanski et al., 1997)
    - Locomotion & Assessment Scales (Kruglanski et al., 2000)
    - Barratt Impulsivity Scale-11 (BIS-11, Patton, Stanford, & Barrett, 1995)
    - Conscientiousness (John & Srivastava, 1999)
    - Propensity to Trust Scale (Merritt, Heimbaugh, LaChapell, & Lee, 2013)

- The questionnaire battery will also include demographic and qualitative feedback questions to collect rich data relating to the self-perceived view from participants on their performance and the rationale behind decisions related to using and not using the automated decision support tool.

Hypotheses:

Table 2 displays all hypotheses included in this study. Due to the number of hypotheses within this research they have been subdivided into levels (Primary, Secondary or Exploratory). Also presented are the measures that will be used to address each research question.

Table 2. Research hypotheses and proposed analysis for Study 1

Hypotheses Level	Research Question	Measures
Primary	Individuals with access to the DSS during the task (groups B or C) will perform better (i.e. correctly classify more tracks) compared to the control group (A).	- Number & Accuracy of classification decisions made
Primary	Individuals will use Decision Support System (DSS) when task demands exceed cognitive capacity (i.e. when mental/temporal workload is high)	- Workload - Time at which DSS selected.
Secondary	Individuals in low accountability condition will select DSS (more times) compared to individuals in high accountability condition who will only select DSS once if at all.	- Condition Low ($_L$) or High ($_H$) [within subjects] - Number of times DSS is selected
Secondary	Individuals in low automation prime condition (B) will select DSS more than individuals in the high automation prime condition (C).	- Condition (B) or (C). [between subjects] - Number of times DSS is selected
Primary/ exploratory	Explore the relationship between scores on each cognitive trait and automation use	- DSS selection - Scores on each cognitive trait
Primary/ exploratory	Explore the relationship between scores on each cognitive trait and performance on classification task	- Performance (number & accuracy of classification decisions made) - Scores on each cognitive trait - Controlling for DSS use
Secondary	Individuals who score high on Cognitive Flexibility Inventory will use DSS during phases 3-6 (15-30minutes into the scenario) compared to those who score low on the Cognitive Flexibility Inventory.	- Score on Cognitive Flexibility Inventory - Time at which DSS selected
Secondary	Individuals who score high on Need for Closure will use DSS during phases 1-3 (0-15 minutes into the scenario) compared to those who score low on the Need for Closure.	- Score on Need for Closure Scale - Time at which DSS selected
	Individuals who score high on Locomotion will use DSS during phases 1-3 of the task	- Score on Locomotion scale - Time at which DSS selected
	Individuals who score high on Assessment will use DSS during phases 3-6 of the task	- Score on Assessment scale - Time at which DSS selected
Secondary	Individuals who score high on BIS-11 will use DSS during phases 1-3 of the task.	- Score on BIS-11 - Time at which DSS selected
Secondary	Individuals who score high on Conscientiousness will not use DSS as many times as those who score low on conscientiousness	- Score on conscientiousness aspect of Big-5 inventory - Number of times DSS selected
Secondary	Individuals who score high on Propensity to trust scale will use DSS earlier (during phase 1) and more times compared to individuals who score low	- Score on Propensity to trust scale

		- Time at which DSS selected - Number of times DSS selected
Exploratory	Are demographics balanced across each test group (A, B & C)?	- Performance score (number and accuracy of classifications) - Age - Gender - Past experience (i.e. previous study) - Current study/employment
Exploratory	Do the cognitive measures assess state or trait? - Propensity to trust (Merritt et al, 2013)[a] - Cognitive Flexibility Inventory (Dennis & Vander Wal, 2010)[b]	- Scores on cognitive measures at time point 1 - Scores on cognitive measures at time point 2
Exploratory	What factors underpin automation use?	- Selection of DSS - Workload - Demographics (gender, age, past experience etc) - Cognitive measures (CFI, NFC etc.) - Accountability condition - Automation background condition

[a]The Propensity to trust scale was developed recently, therefore it would be worthwhile to explore if this measure assesses the state of an individual or trait of an individual.

[b]The cognitive flexibility inventory was also quite recently developed, therefore, it would be worthwhile to explore if this measures assesses the state of an individual or trait of an individual and if the two features within the scale (Control or Alternative) are reliable within the sample collected.

Implications:

It is hoped that an understanding of the potential relationships between these cognitive traits and automation use can be established from this study. Due to time constraints and access issues, the sample will be students- potentially limiting the applicability of any findings to RN personnel. However, it is hoped that this study can be extended to naval personnel as participants. Increasing the complexity of the microworld and scenarios presented during the 30-minute test phase.

Study 2.

How far can the findings from the microworld experiment conducted with a student sample extend to SMEs? To address this research question several hypotheses are proposed, see Table 3.

Table 3. Research Hypotheses and Proposed analysis for Study 2

Hypothesis level	Research question	Measures
Primary	Does a relationship between self-reported automation use and scores on propensity to trust scale exist?	-Qualitative self-report of automation use during deployment -Scores on propensity to trust scale
Exploratory	What are the opinions of RN personnel towards automation use?	-Qualitative questions
Exploratory	Does a relationship exist between scores on cognitive measures and score on propensity to trust scale	-Scores on all cognitive measures
Exploratory	Are attitudes towards automation influenced by demographics (i.e. gender, job role/experience etc)?	-Demographics

	-Qualitative questions
	-Propensity to trust scores

NB: It is acknowledged that cause and effect cannot be established by the proposed research questions and analysis. However, as the aim of this research is to explore a novel way to approach and understand the human-automation relationship in critical environments, this research is intended as a starting point from which future research can draw from.

Procedure:

An online questionnaire will be distributed to appropriate SMEs and RN personnel via a gatekeeper. The questionnaire will consist of the same individual differences measures as those included in the student questionnaire battery, and a series of qualitative questions to explore self-reported use of automated systems. Additionally, the questionnaire will include demographic questions relating specifically to individuals with experience of conducting and/or being trained for above water warfare roles (i.e. gender, roles, time spent at sea, time spent since being at sea etc.).

Results:

Findings from study 2 will be used to explore any potential differences that may be present between practitioners and students. Where there are similarities, it may be that the findings from study 1 can be extended to highlight recommendations for how new tools are deployed and presented to serving RN personnel. Where there may be differences, it is hoped that through the extension of the micro-world experiment to the SME sample, these differences will be understood in more detail.

IMPLICATIONS

This research will contribute to the theoretical understanding of automation usage decisions in reality. Combining the findings of Study 1 and 2, it is hoped that meaningful conclusions can be drawn from the data collected. The microworld simulation has been designed as such that, although for this work a generic classification tool was programmed, the development of more specific and specialised automated tools can be bolted onto the main software. As a whole it is hoped that this microworld will provide the opportunity for future research within academia to explore the benefits and realistic use of a new tool (utilising student samples) prior to such a tool being rolled out to naval personnel, or indeed any military or emergency services personnel. It is hoped that this will help to address the current issue of new automated decision support tools being ‘brittle’ and not used as intended when moved out of the developers’ lab environment.

REFERENCES

Biros, D., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation, 13*(2), 173-189.

Braun, V., & Clarke, V. (2013). *Successful Qualitative Research: A Practical Guide for Beginners.* London:SAGE Publications Ltd.

Cannon-Bowers, J.A., & Salas, E. (1998). *Making decisions under stress: implications for individual and team training*. Washington, DC: APA Press.

Cummings, M.L., & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors, 49*(1), 1-15.

Denis, J.P., & Vander Wal, J.S. (2010). The cognitive flexibility inventory: instrument development and estimates of reliability and validity. *Cognitive therapy research, 34*(3), 241-253.

Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural and biomedical sciences. *Behaviour Research Methods, 39*(2), 175-191.

Goud, R., de Keizer, N.F., ter Riet, D., Wyatt, J.s., Hasman, A., Hellemans, I.M., & Peek, N. (2009). Effect of guideline based computerised decision support on decision making of multidisciplinary teams: cluster randomised trial in cardiac rehabilitation. *BMJ*, 338:b1440 doi: 10.1136/bmj.b1440.

Holt, J. (1988). Assessing the need for decision support systems. *European Journal of Operational Research, 37*(1), 73-82.

John, O.P. & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement and theoretical perspectives. In L.A. Pervin & O.P. John (Eds.). *Handbook of Personality: Theory and research* (Vol.2, pp.102-138). New York: Guilford Press.

Kruglanski, A. W., Pierro, A., Higgins, E. T., & Capozza, D. (2007). "On the move" or "Staying put": Locomotion, need for closure and reactions to organizational change. *Journal of Applied Social Psychology, 37*(6), 1305-1340.

Kruglanski, A. W., Thompson, E. P., Higgins, E. T., Atash, M. N., Pierro, A., & Shah, J. Y. (2000). To "do the right thing" or "Just do it": Locomotion and Assessment as distinct self-regulatory imperatives. *Journal of Personality and Social Psychology, 79*(5), 793-815.

Merritt, S., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. *Human Factors, 55*(3), 520-534.

Ministry of Defence. *Joint Doctrine Publication 0-10*: UK Defence Doctrine. (2014, 5th Edition).

Patton, J.H., Stanford, M.S., & Barratt, E.S. (1995). Factor Structure of the Barratt impulsiveness scale. *Journal of clinical psychology, 51*(6), 768-774.

Schraagen, J.M., Klein, G., & Hoffman, R.R. (2008). The macrocognition framework of naturalistic decision making. In J.M. Schraagen, L.G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp.3-25). Aldershot, UK: Ashgate.

Singh, R., Molloy, I.L., & Parasuraman, R. (1993). Automation induced "Complacency": Development of the Complacency-potential rating scale. *The International Journal of Aviation Psychology, 3*(2), 111-122.

St. John, M., Smallman, H.S., Manes, D.I., Feher, B.A., & Morrison, J.G. (2005). Heuristic automation for decluttering tactical displays. *Human Factors, 47*(3), 509-525.

Wiener, E.L., & Curry, R. E., (1980). Flight-deck automation: Promises and Problems. *Ergonomics, 25*, 995-1011.

When good operators perform poorly

Sue BROUWERS[a] and Mark WIGGINS[b]

[a]Macquarie University, North Ryde, NSW Australia sue.brouwers@mq.edu.au
[b]Macquarie University, North Ryde, NSW Australia, mark.wiggins@mq.edu.au

ABSTRACT

A key challenge for researchers in applied industrial contexts is explaining situations where high performing operators respond incorrectly or delay a response to a change in the system state. One explanation is that a proclivity to acquire patterns during highly routinized tasks, can *miscue* performance when these patterns change. In the present study, 50 university students undertook an assessment of cue utilization and also engaged in a 16-minute rail control simulation. The rail control task involved routine monitoring with periodic interventions to re-route trains. However, the movement of trains followed a consistent but implicit pattern. Following 11 minutes of exposure, this pattern of train movements was altered. Participants who indicated that they had detected the pattern of train movements and who recorded greater levels of cue utilisation also recorded a relatively greater increase in response latency following the change in the pattern of train movements, consistent with a deterioration in performance. The results suggest that operators who engage in highly routinized activities and who record greater levels of cue utilisation, are vulnerable to miscued performance following a change to the pattern of task-related stimuli. The implications are discussed for training and system design.

KEYWORDS

Learning and training; Cue utilization; Pattern recognition; Miscues; Novel task performance

INTRODUCTION

The acquisition of skilled performance has been characterized by the capacity to rapidly and accurately extract and utilize meaningful information from features in the environment (Abernethy, 1987; 1990; Bellenkes, Wickens, & Kramer, 1997; North et al., 2009; Reischman & Yarandi, 2002), thereby enabling the discrimination of relevant from less relevant cues (Weiss & Shanteau, 2003). For example, in medical diagnostic tasks, learners gain expertise by acquiring knowledge about the specific features that are best able to differentiate diseases (Norman, Rosenthal, Brooks, Allen, & Muzzin, 1989; Norman, Brooks, Allen, Rosenthal, 1990). The importance of cue utilization in a high-stakes setting is illustrated by the 1989 Hillsborough stadium disaster in England (Nicholson & Roebuck, 1995; BBC, 2016). Ninety-six Liverpool football spectators were killed and hundreds injured when thousands of fans were allowed into two football standing 'pens', sealed off by crowd barriers and fences at the front of the stands. The event was live-telecast (BBC, 2016) and despite visible evidence of overcrowding amongst the Liverpool section of the stadium (Hillsborough Independent Panel, 2012), the danger went unnoticed until the crush caused barriers to break (Nicholson & Roebuck, 1995; Taylor, 1990) and injured spectators spilled onto the pitch. Crowd control officials and police later explained that they attributed the crowding, screams as well as other behaviours (such as attempts by some individuals to climb barriers), to fan revelry and rowdiness (Scraton, 1999; Taylor, 1990). By the time the first ambulance arrived, most of the fatalities had already occurred (Gibson, 1990; Scraton, 1999).

Crowd crush conditions have been known to precede stampedes, choke points, stacked bodies and death via asphyxiation (Ahmed, Arabi, & Memish, 2006; Byard, Wick, Simpson, & Gilbert, 2006; Hsu & Burkle, 2012; Lee & Hughes, 2005; Zhen, Mao, & Yuan, 2008). Harding, Amos and Gwynne (2008) have identified several critical risk features for crowd crush conditions that can be used to predict and prevent crowd crushing. Crowds where more than four persons occupy one square metre of space, is one such visual cue (Harding, et al., 2008). Footage of the Hillsborough incident revealed an estimated 8.4 persons occupied each square metre within pen 3, where most of the deaths occurred (Nicholson & Roebuck, 1995). Put simply, there was a failure to recognise a major incident until it was too late (Fruin, 2002; Scraton, 1999), and while the critical features were available, officials either did not detect the features or were miscued, assuming these patterns were indicative of revelry or hooliganism rather than crowd-crushing (Hopkins & Treadwell, 2014). Miscueing refers to the activation of an inappropriate association in memory by a salient feature, thereby delaying or preventing the accurate recognition of an object or event (Rowe, Horswill, Kronvall-Parkinson, Poulter, & Mckenna, 2009; Wiggins & Loveday, 2015). Under certain circumstances, miscueing can result in an expert practitioner's speed and diagnostic accuracy decreasing to that approaching a novice (Shanteau, 1992).

The human propensity for error in reasoning and judgement has been well documented and catalogued (Croskerry, 2003; Dekker, 2014; Hammond, 1996; Leape, 1994; Rasmussen, 1982; Reason, 1990; 2000; Shappell & Wiegmann, 2012; Shorrock, & Kirwan, 2002), and while serious incidents such as industrial, aviation and nuclear power accidents may be rare, operators throughout the world every day, function in settings marked by vaguely defined goals, uncertainty, time pressure and a range of organizational and team-based constraints (Orasanu & Connolly, 1993). From a Naturalistic Decision Making (NDM) perspective, highly skilled operators make 'good decisions' under these conditions, because of their reliance on cues and patterns, which allow them to, often tacitly, recognise previous solutions (Klein, 1998; Klein, Calderwood & Clinton-Cirocco, 1986; Orasanu & Connolly, 1993). A sensitivity to cues or patterns, however, may come with an increased likelihood of being miscued. This is especially likely to occur if the learning experiences of the operator are largely rote or highly automated and repetitive. In today's increasing trend toward automation, there is a danger that operators in these environments are becoming increasing susceptible to performance decrement or misdiagnoses in the face unexpected occurrences.

The aim of the present study was to examine whether the propensity for pattern recognition is associated with a propensity for miscueing. Participants were motor vehicle drivers with no previous experience in rail control, who first undertook a battery of situational judgement tasks (EXPERTise driving battery) which provided a composite assessment of cue-utilization in the context of motor vehicle hazard detection and way-finding. They subsequently engaged in a 16-minute rail control simulation that required them to monitor and periodically re-route trains according to a train-number and track-label matching rule. Participants were not informed however, that trains on only two of the four tracks required re-routing. This pattern represented an opportunity to reduce the workload considerably, as it enabled participants to divert any train arriving on two particular tracks and disregard all other task features (e.g., the number-label matching rule and the trains on the two other train tracks).

To provide participants with the opportunity to acquire the pattern, the first six minutes of the rail task involved participants diverting trains in the absence of any other tasks. Throughout the remaining 10 minutes of the rail task, participants completed a secondary task, which was designed to increase the cognitive demands and incentivise the management of these demands. An abrupt change to the train diversion pattern occurred 12 minutes into the task. More specifically, the pattern of diversions to which participants had been exposed reversed, so that the tracks which could previously be ignored, now required an intervention. Consistent with the experimental design utilized by Brouwers, Wiggins and Griffin (in press), participants were asked to respond as rapidly and as accurately as possible where a train required a diversion and they were not provided with any instructional information about the pattern. Having completed the rail task, participants were asked if they noticed a pattern. Accurate responses were recorded as 'Yes' (for a correctly identified pattern) or 'No' (did not identify a pattern).

Based on the assumption that cue utilization is associated with pattern recognition, it was anticipated that, after the extended period of exposure to the initial pattern of train diversions, participants with greater levels of cue utilization would identify the pattern and, as a consequence, be miscued by the change in the pattern. This would be evident in a statistically significant increase in response latencies (for accurate responses to misrouted trains) immediately following exposure to the change in pattern. More specifically, the expectations of the present study were as follows:

- H1. Participants with higher levels of cue utilization, as measured by EXPERTise v. 2.0, will be more likely to identify the train pattern, than participants with lower levels of cue utilization.
- H2. There will be a significant 3-way interaction effect between cue utilization level, pattern identification and block trials, in which participants who identify the rail pattern and have higher levels of cue utilization, will record a greater increase in mean response latency (following exposure to the miscue), in comparison to participants who fail to identify the pattern and who have lower levels of cue utilization.
- H3. There will be a significant 2-way interaction effect between levels of cue utilization and block trials, in which the mean response latency for participants with higher levels of cue utilization will increase, while the response latencies of participants with lower levels of cue utilization will remain stable.
- H4. There will be a significant 2-way interaction between pattern identification and block trials, whereby the mean response latency for participants who identify the pattern will increase, while the mean response latency for participants who fail to report the pattern, will remain stable.

METHOD

Participants

A total of 50 first and second year university students (28 females and 22 males) were recruited for the study, each of whom received course credit in return for their participation. The participants ranged in age from 18 to 22 years (M = 19.34, SD = 1.21). The inclusion criteria comprised existing motor vehicle drivers who had not been exposed to rail control operations, who had corrected to normal vision, and who were aged between 18 and 22 years. Utilizing a cohort of 18 to 22-year-old drivers enabled comparative assessments of cue utilization, controlling to a limited extent, exposure to driving. Due to the train task visual stimuli, any persons who were red-green colour blind were also excluded from participation.

Instruments

EXPERTise 2.0: EXPERTise 2.0 consists of experimental tasks that have been individually and collectively associated with differences in performance at an operational level (Loveday, Wiggins, Searle, Festa, & Schell, 2013; Loveday, Wiggins, Harris, O'Hare, & Smith, 2013; Loveday, Wiggins, Festa, Schell, & Twigg, 2013).The driving version of EXPERTise (Brouwers, Wiggins, Helton, O'Hare, & Griffin, 2016; Brouwers et al., in press; Wiggins, Brouwers, Davies, & Loveday, 2014) was utilized, as it assesses the acquisition of cues in a specific cohort and at a specific point in time, and it is a context with which participants would be familiar (Wiggins et al., 2014). The five tasks in the EXPERTise driving battery comprise a Feature Association Task, two Feature Identification Tasks, a Feature Discrimination Task and an Feature Prioritisation Task. These tasks are described in detail elsewhere (Brouwers et al., in press; Loveday, Wiggins, Harris, O'Hare, & Smith, 2013; Loveday, Wiggins, Festa, Schell, & Twigg, 2013; Wiggins et al., 2014).

The criterion validity of EXPERTise has been established in a number of different domains in which typologies formed on the basis of EXPERTise performance differentiated workplace-related performance (Loveday, Wiggins, Searle, Festa, & Schell, 2013; Loveday, Wiggins, Harris, O'Hare, & Smith, 2013; Loveday, Wiggins, Festa, Schell, & Twigg, 2013). The test-retest reliability (Kappa = .59, $p < .05$) has been demonstrated with power control operators at six-monthly intervals (Loveday, Wiggins, Festa, Schell, & Twigg, 2013). In the present study, restricting the age of participants (18-22 years) controlled for exposure to driving experience. This ensured that any differences in cue utilization would be unlikely to result from differences in driving experience. Overall, participants had accumulated a mean of 37.56 months (or about 3 years) of driving experience (*SD* = 12.70 months).

Rail control task: A simulated rail control task (used previously by Brouwers et al., 2016 and Brouwers et al., in press) was used as a novel, low-workload context for the present study. In this task, a computer screen depicts a simulated, simplified rail control display (see Figure 1). The four long, horizontal green lines shown in the display represent railway tracks. Each track incorporates an intersection (depicted by white portions on the track), which is controlled by an interlocking switch labelled, 'Change'. This switch is depicted by a small circle icon, located above each track. Upon selecting the "Change" icon, (with a computer mouse), any train travelling on the connected track will be diverted onto the intersecting line. A train is depicted by a short, red horizontal bar that appears at one end of a train line, and travels across the display.

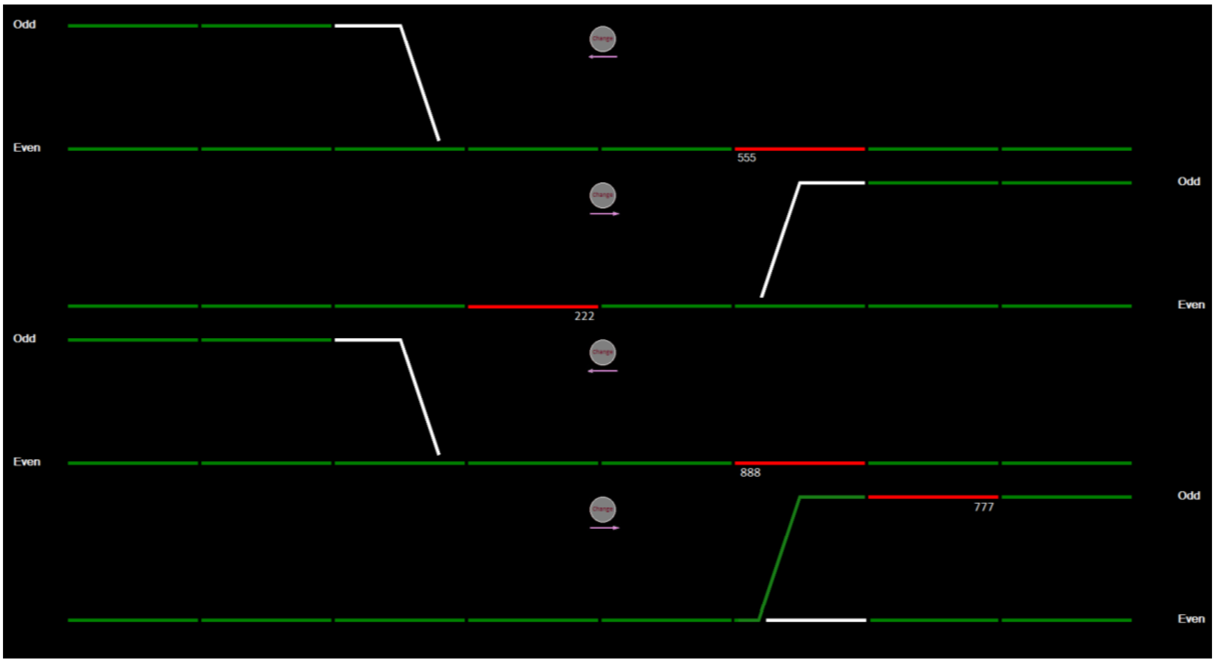


Figure 1. The simulated rail-control display as viewed by participants. Participants were given the following rule: "Even-numbered trains must run along even-labelled lines and odd-numbered trains must run along odd-labelled lines". The red lines represent trains and the four long, horizontal green lines represent railway tracks. If a train is inconsistent with its track, the participant must select the "Change" icon above the train, which will divert the train. The train on the lowermost track in this figure is numbered "777". This train is travelling from left to right, moving across the screen, and has been diverted (correctly) onto an odd labelled track.

Each train has a three-digit number assigned as either odd or even (e.g., 333, 888). Each train line and its associated branch line also have an assigned label: Odd or Even. As the train appears onto the screen, a green line depicts the programmed route of the train. Participants were given the following rule: Even-numbered trains must run along even-labelled lines and odd-numbered trains must run along odd-labelled lines. To ensure this rule was followed, participants were required to periodically select "Change" to re-route a trains whose number did not match its route. All of the trains in the task progressed at the same speed and once a train emerged onto the computer screen, participants had up to seven seconds in which to decide whether or not to reroute the train. Different trains however, appeared between 6-8 seconds of each other. Therefore, at any given time, the display depicted all four train lines occupied with a train. A total of 206 trains appeared on the four rail lines over the course of sixteen minutes, half of which (103 trains) were not required to be re-routed. Designed as a low workload, semi-automated task, the primary goal of the participant was to attend to the display, identify a match between each train and its planned route, and select the 'change' icon as appropriate so that a train's route corresponds with its number. Data recorded from this task included response latency (in milliseconds, from the initial appearance of a train, to the selection of the "Change" icon), and the accuracy of responses (whether trains are diverted when required).

Procedure

After providing demographic information via an on-line questionnaire, participants then either commenced the rail task, which took sixteen minutes to complete, or EXPERTise (on-line tasks), which took twenty minutes to complete. During the completion of EXPERTise, participants were prompted by on-line instructions. The rail task procedure consisted of an initial brief practice trial to orient participants to the rail task. Instructions were also provided to participants in relation to the secondary task and they were familiarized with the paper-based, secondary-task sheet. Once participants indicated that the instructions were understood, the six-minute, single-condition, simulated rail control task commenced. Following this, for the remaining 10 minutes, participants completed the secondary-task sheet concurrently (logging the arrival of trains). The clock used by participants in logging the time of each trains' arrival, was a 22 x 7 cm-sized digital clock, positioned to the right of the computer monitor, so that no portion of the rail display was obscured. At the conclusion of the rail task, participants were asked, "Did you notice a pattern in the rail task?" If they indicated that they had observed a pattern, they were then asked to describe the pattern and their responses were recorded verbatim.

RESULTS

Preliminary analysis

Rail task performance scores: Response latencies were calculated from the initial appearance of a train to the correct selection of the 'Change' icon. The rail task was 16 minutes in duration and the mean response latencies for correct responses were calculated across four time intervals, which included three blocks and one miscue interval. A total of 138 trains appeared on the four rail lines over the course of 16 minutes, half of which (69 trains) were not required to be re-routed (see Table 1). The mean response latencies (in milliseconds) captured over the 3 blocks and the miscue time block, comprised the dependent variables in subsequent analyses.

Table 1

The number of trains that appeared and required re-routing in the rail task

Description of time blocks	Appeared	Required re-routing
Block 1 (6-min)	52	26
Block 2 (5-min) + Secondary task	43	21
Miscue (1-min) + Secondary task	9	5
Block 3 (4-min) + Secondary task	34	17
Totals 16-min	138	69 (50%)

Cue utilization typologies: Prior to further analysis, it was necessary to identify the cue utilization typologies that corresponded to relatively higher or lower levels of cue utilization (Loveday, Wiggins, Harris, O'Hare, & Smith, 2013; Loveday, Wiggins, Festa, Schell, & Twigg, 2013; Wiggins et al., 2014). Consistent with the standard approach to EXPERTise data, *z* scores were calculated for each task, with those corresponding to the Information

Acquisition and Feature Identification tasks reversed so that for all five tasks, higher *z* scores represented greater levels of cue utilization. A cluster analysis identified two groups with centroids, classifying 25 participants in the lower cue utilization typology and 25 participants in the higher cue utilization typology (**Table 2**).

Table 2 Cluster Centroids for the EXPERTise Task Scores

	Cluster 1 (*n*=25) Low cue typology	Cluster 2 (*n*=25) High cue typology
Feature Identification I	-.18	.06
Feature Identification II	-.25	.31
Background Knowledge I	-.64	.81
Background Knowledge II	-.52	.71
Problem Diagnosis	-.67	.68
Information Acquisition	-.42	.36

Pattern identification: The successful identification of the rail task pattern (yes, no) comprised an independent, grouping variable in subsequent analyses. Based on their verbal responses, a total of 42 percent of participants (n = 21) successfully identified the pattern, while 58 failed to identify the pattern (n = 29).

Pattern identification and cue utilization: To investigate whether a greater proportion of participants who identified the pattern, also recorded greater levels of cue utilization, the relationship between cue utilization and pattern identification was examined using a 2 x 2 chi-square test of independence. This analysis incorporated two levels of cue utilization (higher and lower) and two levels of pattern identification (yes, no). The results revealed a statistically significant relationship between pattern identification and the cue utilization, $\chi^2(1, 50) = 8.49$, $p = .004$, phi = .412. While 61.5 percent of participants in higher cue utilization group (n=16) verbally identified the pattern, 20.8 percent of the low cue utilization group (n = 5) verbally identified the pattern (see Figure 2). Therefore, participants with higher levels of cue utilization were 6.1 times more likely (odds ratio = 6.08, 95% CI = 1.72, 21.50, $p = .005$) to report the pattern compared to those participants with lower levels of cue utilization.

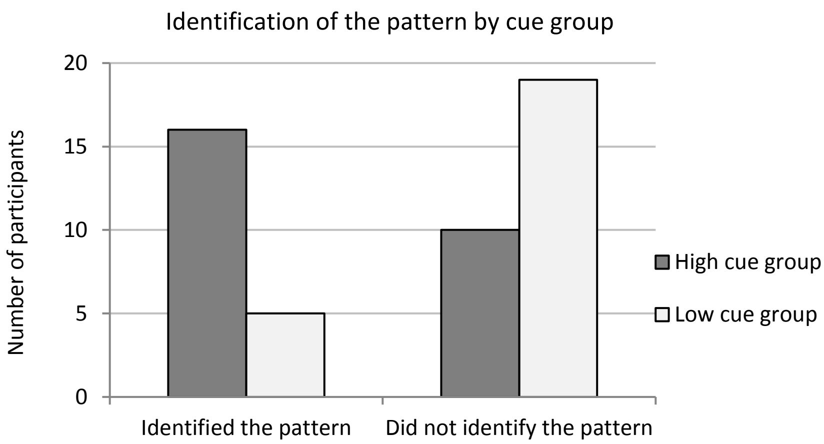


Figure 2. The number of participants who verbally identified the rail task pattern, plotted by cue utilization group.

Impact of the miscue

The mean response latencies of participants, across the four rail task trials, grouped by pattern identification together with cue level, are shown in Figure 3. To investigate whether participants with greater levels of cue utilization and who identified the rail pattern, were miscued by the change in the pattern of trains, a 2 x 2 x 2 mixed repeated ANOVA was undertaken, incorporating two levels of cue utilization (greater, lesser) as a between-groups variable, pattern identification (yes, no) as a between-groups variable, Block 2 (the 4-minute period preceding the miscue) and the Miscue interval as within-group variables, and response latency as the dependent variable.

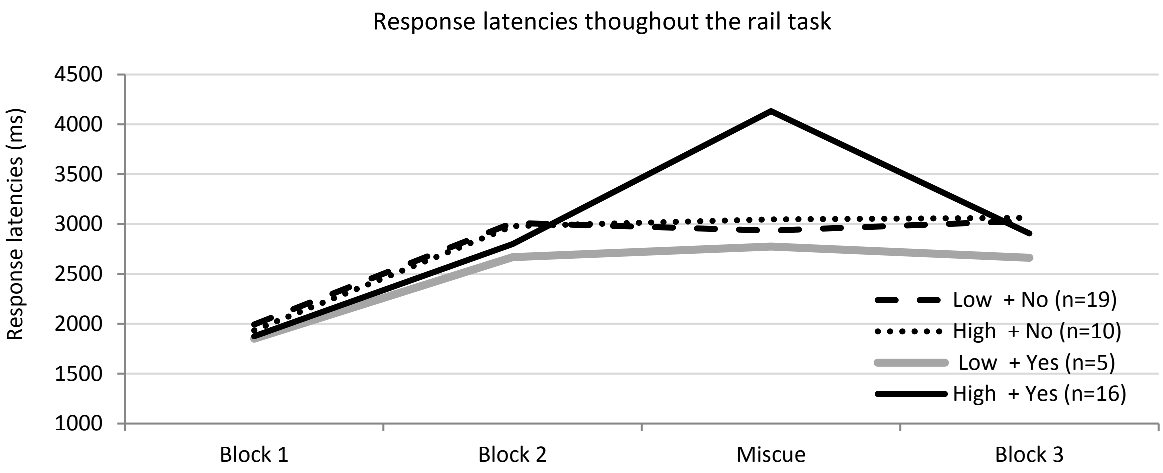


Figure 3. Mean response latencies over time, in the rail task, grouped by cue level and pattern identification (Yes = identified the pattern, No = failed to identify the pattern). The Miscue trial is a 1-min interval that captured response latencies (in milliseconds) to an abrupt pattern change in programmed train diversions.

The result of the 3-way interaction was not statistically significant, $F(1,46) = 2.73$, $p = .105$, indicating that participants who identified the rail pattern and who had greater levels of cue utilization, did not record a greater increase in mean response latency between Block 2 and the Miscue, in comparison to participations who failed to identify the rail pattern and/or who had lesser levels of cue utilization. A statistically significant interaction was evident between levels of cue utilization and block trials, $F(1,46) = 4.39$, $p = .042$, partial $\eta^2 = .087$, in which the mean response latency for participants with higher levels of cue utilization increased from $M = 2869.27$ msecs ($SD = 868.24$) to $M = 3715.27$ msecs ($SD = 1282.84$) while the response latencies of participants with lesser levels of cue utilization remained relatively stable, changing from $M = 2941.96$ ($SD = 689.65$) to $M = 2901.28$ ($SD = 959.62$). This interaction is shown in Figure 4

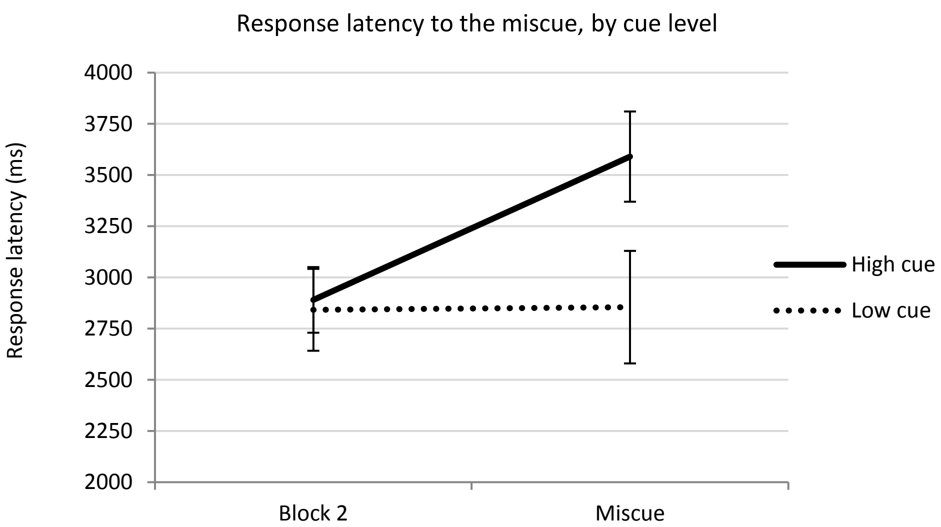


Figure 4. The mean response latencies of participants in the high and low cue group, before and after the pattern change. Error bars represent ±1 SE.

A statistically significant interaction was also evident between pattern identification and block trials, $F(1, 46) = 4.90$, $p = .032$, partial $\eta^2 = .032$, whereby the mean response latency for participants who identified the pattern, increased from a mean of 2768.79 ms (*SD* = 761.23) to 3809.61 ms (*SD* = 1261.32), while the mean response latency for participants failed to report the pattern, remained relatively stable, changing from 3002.19 (*SD* = 792.23) to 2973.31 (*SD* = 1039.46). Figure 5 highlights this interaction.

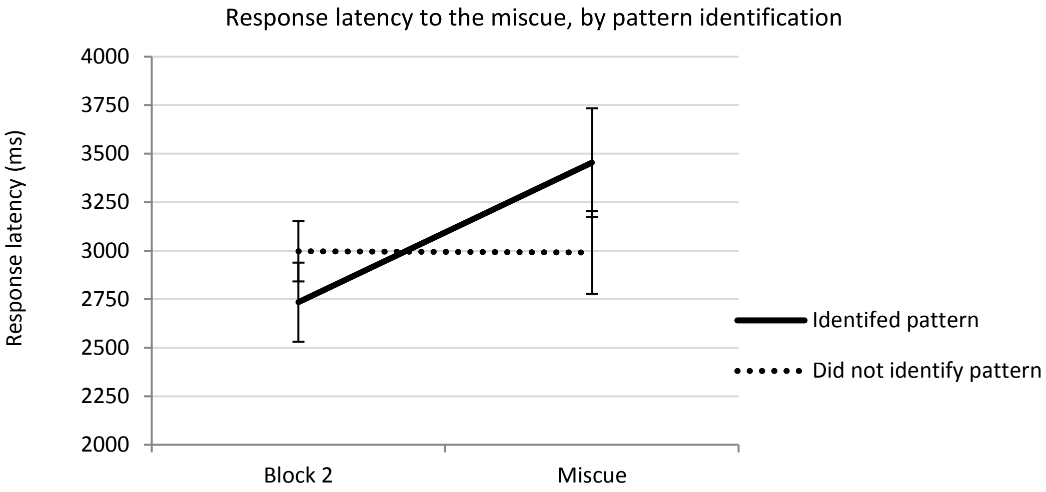


Figure 5. The mean response latencies of participants, grouped by pattern identification (those who did or did not verbally identify the train sequencing pattern), in response to the pattern change in the rail task. Error bars represent ±1 SE.

DISCUSSION

The superior performance of skilled operators across a range of occupational domains, has been attributed to a reliance on cues and patterns, which allow operators to, often tacitly, recognise previous solutions (Klein, 1998; Klein, Calderwood & Clinton-Cirocco, 1986; Orasanu & Connolly, 1993). Due to the dynamic nature of real-world settings, which are often marked by uncertainty and changing patterns, a sensitivity to cues or patterns, may have the disadvantage of increasing the likelihood of being miscued during certain situations. The aim of the present study was to examine whether the propensity for pattern recognition is associated with a propensity for miscueing during a carefully designed rail control task. The participants completed EXPERTise v. 2.0 which is designed to assess cross-task levels of cue utilization (Wiggins et al., 2014). Subsequently, participants engaged in a 16-minute rail control simulation that involved monitoring and periodically re-routing trains according to a train-number and track-label matching rule. Participants were not informed that trains on only two of the four tracks required re-routing. This pattern represented an opportunity to reduce the workload considerably. After 11 minutes of exposure, this programmed pattern was abruptly changed so that trains that previously did not require re-routing, now did and vice versa. This change in pattern was designed to elicit a miscue effect and it was reasoned that participants with higher greater levels of cue utilization (as measured by EXPERTise 2.0), and who successfully identified the rail task pattern (verbally reported the pattern), would be more vulnerable to the change in pattern, and would demonstrate increased response latencies in the rail task, immediately following the pattern change.

Consistent with H1, participants with higher levels of cue utilization, were 6 times more likely to identify the train pattern, compared to participants with lower levels of cue utilization, suggesting that a sensitivity to cues is related to a sensitivity to pattern identification. Consistent with H3, a statistically significant 2-way interaction was also evident, in which the mean response latency for participants with higher levels of cue utilization increased in response to the change in pattern, while the mean response latency of participants with lower levels of cue utilization remained relatively stable. This suggests that a sensitivity to cues may be related to a proclivity for being miscued under particular circumstances. Consistent with H4, a statistically significant, 2-way interaction between pattern identification and block trials was evident in which the mean response latency for participants who identified the pattern increased following the change in the pattern of trains, while the mean response latency for participants who failed to report the pattern, remained relatively stable. This interaction suggested that a sensitivity to detecting patterns may be related to a proclivity for being miscued.

It was also hypothesised that a 3-way interaction would be evident (H2), wherein the participants who identified the rail pattern and who had higher levels of cue utilization, would record a greater increase in the mean response latency (following exposure to the miscue), in comparison to participants who failed to identify the pattern and who recorded lower levels of cue utilization. Since this hypothesis was not supported, the results suggest that, while a sensitivity to cues and patterns are related to a susceptibility of being miscued, these factors in combination, did not account for a significant portion of the variance in train task performance. It may be the case that other factors such as specific visuo-spatial abilities may also explain performance during the task. Future studies that examine cue utilization and pattern recognition in other contexts (aside from rail control), may assist in delineating whether: (a) there are task-specific effects that are not evident in performance across other tasks, and/or (b), whether other abilities/skills, aside from cue utilization, may explain performance decrement in the face of pattern changes.

From a theoretical perspective, the possibility that a capacity to rapidly detect patterns of dynamic stimuli may come with a tendency to be miscued, is consistent with the view that an over-reliance on environmental regularities can compromise the performance of even highly skilled operators (Kahneman & Klein, 2009). There are a number of potential training-related implications that arise from the findings of the present study. For example, in high-stakes settings, it may be important to expose learner-operators to a wide range of real-world or simulation settings that vary in complexity, time-pressure and cue availability, enabling them to acquire cues yet also detect and adapt to dynamic changes. There are also tangible selection-related opportunities with regard to the use of EXPERTise v. 2.0. The capacity to identify levels of cue utilization may aid in differentiating job applicants who are more or less likely to advance and acquire skills and who may benefit most from real world or dynamic simulation settings.

CONCLUSION

In summary, the findings of the present study suggest that individuals who readily acquire patterns and cues may also be those operators who are more at risk of misapplying rules or misdiagnosing situations, particularly in highly routinized environments. A proclilvity for miscueing, may go some way to explaining why good operators can perform poorly. The use of tools such as EXPERTise 2.0, which can be used to identify the cue utilization skill of operators, may be useful in identifying high performers or 'tomorrows' experts', who may benefit most from exposure to a wide range of real-world or simulation settings that vary in complexity, time-pressure and cue availability, allowing them to acquire critical cues yet also remaining adaptable to dynamic changes.

REFERENCES

Abernethy, B. (1987). Anticipation in sport: A review. Physical Education Review, 10, 5-16.

Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. Journal of Sports Sciences, 8, 17–34.

Ahmed, Q. A., Arabi, Y. M., & Memish, Z. A. (2006). Health risks at the Hajj. The Lancet, 367, 1008-1015

BBC (2016). Hillsborough disaster. BBC News UK. Retrieved 6 May 2016 from http://www.bbc.com/news/live/uk-england-merseyside-36102998

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. Aviation, Space, and Environmental Medicine, 68, 569-579.

Brouwers, S., Wiggins, M. W., & Griffin, B. (in press). Reducing the impact of cognitive load on task performance: An outcome of cue utilisation.

Brouwers, S., Wiggins, M. W., Helton, W., O'Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. Frontiers in Psychology, 7, 1-12.

Byard, R. W., Wick, R., Simpson, E., & Gilbert, J. D. (2006). The pathological features and circumstances of death of lethal crush/traumatic asphyxia in adults - a 25-year study. Forensic Science International, 159, 200-205.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. Academic medicine, 78, 775-780.

Dekker, S. (2014). The field guide to understanding 'human Error'. Burlington, USA: Ashgate Publishing:

Furin, J. (2002). The causes and prevention of crowd disasters. First International Conference on Engineering for Crowd Safety, London, England, March 1993. Revised for www.crowdsafe.com, January 2002. Elsevier Science Publishers. Retrieved 10 May 2016 from http://www.crowdsafe.com/fruincauses.pdf.

Gibson, R. M. (1990). Report of the Medical Working Party. London, UK: HMSO.

Hammond, K. R. (1996). Human judgment and social policy: Incredible uncertainty, inevitable error, unavoidable justice. New York: Oxford University Press.

Harding, P. J., Amos, M., & Gwynne, S. (2008). Prediction and mitigation of crush conditions in emergency evacuations. In Pedestrian and Evacuation Dynamics 2008 (pp. 233-246). Berlin: Springer.

Hillsborough Independent Panel (2012). Hillsborough Disclosure Website: http://hillsborough.independent.gov.uk/. Crown Copyright 2012: United Kingdom Government.

Hopkins, M., & Treadwell, J. (2014). Football Hooliganism, Fan Behaviour and Crime: Contemporary Issues. London: Palgrave Macmillan.

Hsu, E. B., & Burkle, F. M. (2012). Cambodian Bon Om Touk stampede highlights preventable tragedy. Prehospital and Disaster Medicine, 27, 481-482.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. American Psychologist, 64(6), 515-526.

Klein, G. (1998). Sources of Power: How people make decisions. Cambridge, MA: MIT Press.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In Proceedings of the Human Factors Society 30th Annual Meeting (pp. 576-580). Santa Monica, CA: Human Factors Society.

Leape, L. L. (1994). Error in medicine. Journal of the American Medical Association, 272, 1851-1857.

Lee, R. S., & Hughes, R. L. (2005). Exploring trampling and crushing in a crowd. Journal of Transportation Engineering, 131, 575-582.

Loveday, T., Wiggins, M. W., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In P. Latorre Carmona, J. S. Sanchez & A. L. N. Fred (Eds.), Pattern recognition - applications and methods (Vol. 204, pp. 1-11). New York: Springer-Verlag Berlin Heidelberg.

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2013). An objective approach to identifying diagnostic expertise among power system controllers. Human Factors: The Journal of the Human Factors and Ergonomics Society, 55, 90-107.

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. Human Factors, 55, 125-137.

Nicholson, C. E., & Roebuck, B. (1995). The investigation of the Hillsborough disaster by the Health and Safety Executive. Safety Science, 18, 249-259.

Norman, G. R., Brooks, L. R., Allen, S. W., & Rosenthal, D. (1990). Sources of observer variation in dermatologic diagnosis. Academic Medicine, 65(9), S19-20.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. Archives of Dermatology, 125, 1063-1068

North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. Applied Cognitive Psychology, 23, 878-894.

Orasanu, J., & Connolly, T. (1993).The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.). Decision Making in Action: Models and Methods (pp. 3-20). Norwood, CT: Ablex.

Rasmussen, J. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. Journal of Occupational Accidents, 4, 311-333.

Reason, J. (1990). Human error. New York: Cambridge university press.

Reason, J. (2000). Human error: models and management. British Medical Journal, 320, 768-770.

Reischman, R. R., & Yarandi, H. N. (2002). Critical care cardiovascular nurse expert and novice diagnostic cue utilization. Journal of Advanced Nursing, 39, 24-34.

Rowe, R., Horswill, M. S., Kronvall-Parkinson, M., Poulter, D. R., & McKenna, F. P. (2009). The effect of disguise on novice and expert tennis players' anticipation ability. Journal of Applied Sport Psychology, 21, 178-185.

Scraton, P. (1999). Policing with contempt: The degrading of truth and denial of justice in the aftermath of the Hillsborough disaster. Journal of Law and Society, 26, 273-297.

Shanteau, J. (1992). The psychology of experts an alternative view. In G. Wright & F. Bolger (Eds.). Expertise and decision support (pp. 11-23). New York: Plenum Press.

Shappell, S. A., & Wiegmann, D. A. (2012). A human error approach to aviation accident analysis: The human factors analysis and classification system. Aldershot, England: Ashgate Publishing.

Shorrock, S. T., & Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control. Applied ergonomics, 33, 319-336.

Taylor, L. J. (1990). Final Report into the Hillsborough Stadium Disaster, CM 962. Lord Justice Taylor, London, HMSO. Accessed May 2016 from: http://www.southyorks.police.uk/sites/default/files/hillsborough%20stadium%20disaster%20final%20report.pdf

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. Human Factors: The Journal of the Human Factors and Ergonomics Society, 45, 104-116.

Wiggins, W., & Loveday, T. (2015). Diagnostic expertise in organizational environments. Surrey, England: Ashgate Publishing.

Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014). Trait-based cue utilization and initial skill acquisition: Implications for models of the progression to expertise. Frontiers in Psychology, 5, 1-8.

Zhen, W., Mao, L., & Yuan, Z. (2008). Analysis of trample disaster and a case study-Mihong bridge fatality in China in 2004. Safety Science, 46, 1255-1270.

Decision-making in Extreme Environments: What contributes to effective teamwork?

Olivia BROWN[a], Nicola POWER[a], Emma BARRETT[a]
[a]*Department of Psychology, Lancaster University, UK*

ABSTRACT
Extreme environments create unique interpersonal, physical and emotional challenges. This PhD research will seek to further understand what allows teams to make effective decisions when working in extreme environments. Specifically it will explore two types of extreme teams: (i) emergency responses teams; and (ii) expedition teams. The focus will be on team cohesion and other aspects of teamwork that have been shown to relate to cohesion. Firstly, a literature review will be performed to identify gaps in understanding and clarify terminology. Secondly the research will focus on exploring what factors facilitate team decision-making during major incidents. This will involve gathering data from a live exercise with Emergency Services teams. Thirdly, the research will aim to track how cohesion emerges in expedition teams. Daily diaries will be used to assess if daily events and interpersonal relations can predict fluctuations in team cohesion. Team behaviours related to team perfomance (coordination, cooperation and communication) will be used to assess team outcomes, as well as Subject Matter Expert evaluations. As this research will involve gathering data in real-world contexts, it will use a Naturalistic Decision-Making framework. Fundamentally, this research will contribute to our understanding and offer recommendations on how to improve teamwork when operating in stressful environments.

KEYWORDS
Research; Decision-Making; Team-efficacy; Extreme Environments; Cohesion.

INTRODUCTION

Naturalistic Decision-Making (NMD) research explores how experts use knowledge to make decisions in complex, real-world settings (Zsambok & Klein, 1997). The PhD research described in this paper will adopt an NDM framework to explore the decision-making of teams operating in extreme environments; complex, dynamic, high-stakes settings wherein success is contingent upon effective teamwork (Militello, Sushereba, Branlat & Finmore, 2015). Specifically, it will involve research with two types of extreme teams: (i) the emergency services responding to major incidents; and (ii) expedition teams who are trekking across harsh landscapes (e.g. arctic, desert). Decision-making during major incidents has been described as inherently uncertain (Alison & Crego, 2008), due to events unfolding rapidly and unexpectedly. Expedition settings are also characterised by threats of danger, difficult living conditions and interpersonal challenges (Palinkas & Seudfeld, 2008). Though there is a large body of research that has proposed solutions to effective team decision-making in business teams, this PhD research will argue that these findings cannot be applied to extreme settings. For example, Cheng, Wang and Zhang (2011) reported that when relationship conflict was reduced, task conflict lead to better decision-making. The study involved teams of business students who were asked develop a plan to rescue a failing clothing manufacturing company. The authors suggest that the presence of a leader during team discussion was able to moderate relationship conflicts and make best use of functional task conflict. Task conflict may increase decision quality by increasing decision options and stimulating creativity (De Dreu, 2006). However, in the context of extreme environments, task conflict might negatively impact upon performance, as it can delay decision-making by increasing deliberation about decision options. To overcome these inconsistencies between business contexts and the context of extreme environments, this PhD will conduct research ‘in the wild’, using NDM methods (McAndrew & Gore, 2013).

Cohesion is a commonly researched concept in the team literature that has been positively linked to team performance outcomes. Cohesion is defined as the shared attraction within a group, reflecting the tendency for the group to remain united in reaching a common goal (Casey-Campbell & Martens, 2009). A recent meta-analysis, including 70 studies of cohesion, emphasised that it was an essential component to effective team performance (Salas, Grossman, Hughes & Coultas, 2015). The authors sought to clarify the dimensionality and measurement of cohesion, emphasising that research should prioritise task and social cohesion. Mikalachki (1969) define task cohesion as the shared commitment to the task and social cohesion as the interpersonal bonds existing between members. For example, when task cohesion is high, a team will be committed to working together interdependently to achieve a common goal. When social cohesion is high there will be a high level of shared liking within the group and so team members will be motivated to work together to achieve shared goals. Given the nature of extreme environments, cohesion will likely be an important for effective team decision making, in facilitating interpersonal relationships and encouraging a joint task focus. Indeed, research has indicated that task cohesion can protect against the deleterious effects of time urgency (Zaccaro, Gualitieri & Minionis, 1995). Zaccaro et al., (1995) asked three person teams to identify the best sites for oil drilling based on information given during the experiment. The study had four conditions, manipulating level of task cohesion and time given to complete the experiment. Analyses indicated that teams in the high task-cohesion, time urgent group performed as well those in the high task-cohesion no time urgency condition and significantly better than those in the low task-cohesion conditions. The authors concluded that the results were indicative of higher levels of coordination and communication in the task-cohesive team, which protected against the effects of time urgency and allowed them to perform as well as the teams not under time pressure.

Though cohesion will be an important part of this research, it was decided that it should not be the sole focus. Stage one of this PhD (the literature review) identified how the literature on cohesion is fraught with inconsistencies over definitions and measurements, with widespread disagreement over how best to operationalise it (Friedkin, 2004). Salas et al., (2015) refer to it as an ‘umbrella’ term as it overlaps and relates to many other aspects of team behaviour, which means that despite its common identification as a key antecedent to effective teamwork, there is little consistency about what it actually means in practice. For this reason, this PhD research will not only explore cohesion in teams, but further investigate other facets of teams that appear to interact with cohesion and behaviour. For example, leadership, group composition, trust, goal formation and team climate. Team processing, as an outcome will be measured around three constructs: communication, cooperation and coordination. Subject matter expert evaluations will supplement this. The literature has consistently supported the role of coordination, cooperation and communication in producing effective team behaviour (Mathieu, Heffner, Goodwin, Salas & Cannon-Bowers, 2000; Kozlowski & Bell, 2003).

A unique feature to this research is its access to two related, but distinct, extreme environments: emergency response contexts and expedition contexts. Much of the literature on extreme environments has tended to treat them as being holistically distinct from non-extreme environments, but it rarely distinguishes between the different types of extreme environments. For example, comparing extreme environments that include rapidly forming teams (e.g., emergencies) versus extreme environments that include teams who work together in that environment for a period of weeks or months (e.g., expeditions). It is anticipated that certain facets of teams will be more important in different contexts. Social cohesion born out of a sense of liking and mutual attraction will likely be of more importance in long-term stable teams, who will be spending a great deal of time together, than teams who will be rapidly formed and then disbanded. This PhD research will thus make a novel contribution by allowing for comparisons across these two different extreme environments.

RESEARCH QUESTION 1: WHAT FACILIATES EFFECTIVE TEAM DECISION-MAKING IN EXTREME ENVIRONMENTS?

This PhD research will begin with an extensive review of the literature on team decision-making. This will include a review of effective team behaviours, with a further examination of other facts of teams that enable these behaviours (e.g., cohesion, team climate, leadership).

Data

Data will be gathered from electronic search engines (e.g., PsycInfo). Papers will be included if they are; written in English; published in peer-reviewed articles and include an element of empirical testing. The search terms used to identify papers will include: ‘decision-making’, ‘effective team behaviour’, ‘coordination’, ‘cooperation’, ‘communication’, ‘cohesion’, ‘team climate’, ‘team trust’, ‘leadership’, ‘extreme environments’. Papers will be grouped according to context, with priority given to non-organisational ones.

Procedure

Guidelines of how to conduct a literature will be followed (Mohr, Liberati, Tetzlaff & Altman, 2009). Papers will be sorted through by assessing the validity of methodology and the relevance of content. Following this a mind map will be completed to allow for visualisation of different aspects of effective team behaviours. It is anticipated that this will allow links to be made across different variables in particular contexts. For example, research has indicated that leadership is vital in maintaining cohesion and building a more uniformed team climate during arctic expeditions (Schmidt, Wood & Lugg, 2004).

RESEARCH QUESTION TWO: WHAT FACTORS DURING MAJOR INCIDENT REPSONSE CAN FACILITATE TEAM DECISION MAKING?

The second study will explore how multi-agency emergency response teams make decisions during major incidents. Data will be collected from a live multi-agency training exercise, taking place in March 2017, over the course of 12 hours based around a counter-terrorism attack. It is being organised by a Police Service in the North of England. Members from each of the blue light services will be present, responding to the live event as if it were a genuine major incident. Following the event, participants will be interviewed and given questionnaires. Specifically, the research will explore the role of goal consistency, leadership, cohesion, trust and communication.

Participants

Participants will be from the UK Emergency Services, including the Police, Fire and Rescue, and Ambulance service. It is estimated that 100 participants will take part, all of who will be experienced in Major Incident response.

Measures:

Semi-Structured Interview

The researchers will conduct semi-structured interviews with emergency responders as soon as possible after the exercise (either on the day, or in the week following). No more than two interviewers will conduct the interviews. The interview will be designed to explore the beliefs, perceptions, preferences and experiences of the participants. Interviews will be transcribed and thematically analysed (Braun & Clarke, 2006) to explore the cognitive processes associated with decision-making during the exercise. Specifically, we are interested in the team variables identified in the literature review, and so questions will seek to unpack themes such as trust, leadership, goal consistency, communication and cohesion. Table 1 provides some example questions.

Table 1. An example of interview questions

Question	Purpose
Did you find that you experienced team-based uncertainties? (e.g., communication, role confusion)	This will allow the researchers to establish what level of support the team provided to each member of the emergency services.
To what extent do you feel that a shared liking within the team enabled an effective working unit? Does it matter? Is it more about getting the job done?	It is hoped that this question will allow the researchers to explore the importance of social cohesion in emergency response teams. With teams formed quickly, it might not matter that a sense of liking is established. It may be more important that team members are confident in one another to complete the task at hand.
Did you feel confident in your teams' ability to respond to the incident? Did you feel that each member of the team was committed to the same team goals?	It is hoped that this question will allow the researchers to establish levels of task cohesion. It might be that this has a more important role in emergency response, with shared goal focus viewed as an important part of effective teamwork.

Questionnaires

Questionnaires will be administered either electronically using Qualtrics or in pen and paper format after the exercise. The majority of questions will be presented on a Likert scale, asking participants to rate statements from one to seven, with one being completely disagree and seven completely agree. The questionnaire will include a measure of perceived team performance, designed by the researchers to assess to what extent responders employed communication, cooperation and coordination behaviours. It is hoped that this questionnaire will be developed throughout the course of the PhD to provide a new measurement of team performance for teams operating in difficult conditions. The questionnaire will also include the Group Environment Questionnaire (GEQ; Carron, Widmeyer & Brawley, 1985). This measure of cohesion was originally designed for sports teams but has since been adapted and utilised in several contexts (Chang & Bordia, 2001; Carless & De Paola, 2000). In addition, it will include the McAlister (1995) measure of team trust. This measure of trust differentiates between cognitive and affective trust. Cognitive trust relates to trust surrounding the knowledge and cognitive capabilities of others. Affective trust relates more closely to emotional bonds between partners, where the social ties between individuals provide the basis of the trust. The scale allows each to be measured separately. In the case of a major incident, where teams are brought together rapidly, it is hypothesised that cognitive trust will be more prevalent in this context. Participants will also be asked to rank six statements in order, to indicate which goal was of primary importance to them during the incident. This will allow goal consistency.

Subject Matter Experts

Subject Matter experts will be consulted to independently assess the performance of the responders. This will allow the researchers to compare their reports with the scores on the perceived team performance scale, described in more detail above.

Future Implications

This research will explore, analyse and increase understanding of teamwork in emergency response contexts. By focusing on effective team behaviours, findings will aim to improve the training of the emergency services, and thus increase the ability to respond to major incidents.

RESEARCH QUESTION THREE: HOW DOES TEAM COHESION EMERGE OVER TIME AND IN RESPONSE TO INTEPERSONAL AND SITUATIONAL CHALLENGES IN EXPEDITION TEAMS?

This research will explore how cohesion develops over time in expedition teams. The research will involve six expedition teams, visiting three different countries (Greenland, Kyrgyzstan and Mongolia) with each expedition being three weeks in length. The research will focus on tracking how cohesion emerges in the teams over time, particularly in response to situational and interpersonal challenges that will be identified via the use of daily diaries during their expedition. Smith and Barrett (in press) conducted similar research, tracking a team of four taking part in a desert expedition. The results indicated that daily events and the use of coping strategies significantly predicted positive and negative affect in the team members. This research will adopt a similar methodology, but with a focus on cohesion as oppose to coping strategies.

Participants

Participants will be recruited from a pre-existing relationship with an agency that organises expeditions for groups of students, all aged between 16 and 18, accompanied by adults. The groups will consist of 10 students and 2 accompanying adults. The expedition will last for three weeks.

Measures

Personality

The few studies that have compared team cohesion and personality have utilised the Big Five Factor Model (Costa & McCrae, 1992). Participants will therefore be given the revised Neuroticism-Extraversion-Openness Personality inventory (NEO PI- R; Costa and McCrae, 1992) to assess their personality. By adopting the same framework used in previous studies, team personality composition (mean variance etc.) will be compared against task and social cohesion (Van Vianen & De Dreu, 2001). From previous findings, it is hypothesised that teams scoring highly in agreeableness will report high task and social cohesiveness. It is also hypothesised that teams scoring highly on neuroticism will be less cohesive. It might be that cohesion emerges at different rates according to patterns in team composition.

Daily Diary

The daily diaries will be adapted from Atlis, Leon, Sandal and Infante (2004), used in prior expedition projects. The diaries will be used to track changes in group dynamics (e.g., trust, interpersonal conflict) and external events over the course of the expedition. The daily diaries will also include items to measure group cohesion and trust. The items will be adapted from the Group Environment Questionnaire (Carron, Widmeyer & Brawley, 1985) and from McAllister's (1995) measure of cognitive and affective trust. The diary will also ask team members about their perceived team climate, and their perception of team goals. It is hypothesised that cohesion will vary according to external and internal daily events. Table 2 provides example questions about daily events. For example, it might be that on a day where team trust is high and the team are experiencing good weather conditions, team cohesion is also high.

Table 2. An example of a daily event diary used by Smith and Barrett(in press)

Please tick next to each event you encountered today

Daily event		Daily event	
Problems with gear and equipment e.g., clothing, tools, communication, navigation equipment etc.		Feeling of camaraderie/closeness with team mates	
Fear of being injured		Worried about family/friends	
Concern about the well-being of my team mates		Concern about how effectively my team mates and I are working together	
Enjoyment of the environment		Personal hygiene (wanting to be cleaner)	
Loneliness, homesickness		Lack of privacy, personal time	
Satisfaction in making good progress today		Tension or argument with my team mate(s)	
Concerns about the effectiveness of safety of the decisions I made today		Satisfaction that equipment is working properly	
Feeling down/low stressed out because my team mates is/are feeling that way		Satisfaction that I am able to cope with challenges	
Diarrhoea		Muscle, joint or injury	
Constipation		Headache	

Worried about encountering bad weather		Lack of sleep	
Seasickness		Other (please briefly explain)	

Team performance

Following completion of the expedition, team leaders will be interviewed to assess performance. Questions will focus on the extent to which communication, coordination and cooperation was employed by team members. For example, providing back-up behaviour to a team member who was showing signs of difficulty would be indicative of cooperative team behaviour.

Future implications

The findings from this study will identify how team cohesion emerges and fluctuates in teams under stress. This will aid in preparing groups for expeditions and extend out into other groups (i.e., Special forces) working for prolonged periods of time in similar harsh environments. Smith and Barrett (in press) emphasised how furthering understanding in expedition teams is an attractive analogue in providing a greater understanding of the experiences of security personnel.

CONCLUSION

This research will explore facets of effective teamwork within teams operating in extreme environments. It will focus on cohesion and other facets of teams that have been shown to relate to cohesion and effective team performance. It is expected that findings will provide useful information to improve the training and preparation of both emergency services and expedition groups, to encourage more effective team functioning. In addition, by researching two distinct contexts, it is hoped that the findings will allow for comparisons across different types of extreme environment. This will form the final and perhaps most important element to this research, by emphasising the need for differentiating between extreme contexts and developing methodologies to compare data across contexts. For example, it is anticipated that social cohesion will be a more integral component of effective teamwork in expedition teams than in major incident response teams. It is expected that as expedition teams are required to spend a considerably longer time together, it will be more important to establish a degree of shared liking within the group. By evidencing the differences in effective team behaviour in different contexts, it is hoped that this PhD can generate contextualised and grounded applied recommendations about the types of team-based factors that are important for different types of teams. It is also hoped that this will spur further research in this area by exploring teamwork through a contextualised NDM lens.

REFERENCES

Alison, L.J., & Crego, J. (2008). Policing Critical Incidents: Leadership and Critical Incident Management. Devon, UK: Willan Publishing.

Atlis, M. M., Leon, G. R., Sandal, G. M., & Infante, M. G. (2004). Decision processes and interactions during a two-woman traverse of Antarctica. *Environment and Behavior*, *36*(3), 402-423.

Barrett, E., & Martin, P. (2014). *Extreme: Why some people thrive at the limits*. Oxford University Press. U.K.

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: a meta-analytic clarification of construct relations. *Journal of Applied Psychology, 88*(6), 989-1004.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. *Small group research*, *31*(1), 71-88.

Carron, A. V., Widmeyer, W. N., & Brawley, L. R. (1985). The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of sport psychology*, *7*(3), 244-266.

Casey-Campbell, M., & Martens, M. L. (2009). Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, *11*(2), 223-246.

Chang, A., & Bordia, P. (2001). A multidimensional approach to the group cohesion-group performance relationship. *Small Group Research*, *32*(4), 379-405.

Cheng, Long, Wang, Zhong-Ming, & Zhang, Wei. (2011). The effects of conflict on team decision making.(Report). *Social Behavior and Personality: An International Journal, 39*(2), 189

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5.

De Dreu, C. K. W. (2006). When too little or too much hurts: Evidence for a curvilinear relationship between task conflict and innovation in teams. *Journal of Management*, *32*, 83–107.

Friedkin, N. E. (2004). Social cohesion. *Annual Review of Sociology*, *30*, 409-425.

Kozlowski, S.W.J., & Bell, B.S. (2003). Work groups and teams in organizations. In W.C. Borman, D.R. Ilgen, & R.J. Klimoski (Eds.), Handbook of psychology: Vol. 12. Industrial and organizational psychology. London: Wiley.

McAndrew, C., & Gore, J. (2013). Understanding preferences in experience-based choice: A study of cognition in the "wild". *Journal of Cognitive Engineering and Decision Making*, *7*(2), 179-197.

McAllister, D. J. (1995). Affect and Cognition based trust as foundations for interpersonal cooperation in organizations. *Academy Of Management Journal*, *38*(1), 24-59

Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., & Cannon- Bowers, J.A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273–283.

Mikalachki, A. (1969). *Group cohesion reconsidered: A study of blue collar work groups*. School of Business Administration, University of Western Ontario

Militello, L. G., Sushereba, C. E., Branlat, M., Bean, R., & Finomore, V. (2015). Designing for military pararescue: Naturalistic decision-making perspective, methods, and frameworks. *Journal of Occupational and Organizational Psychology*, *88*(2), 251-272

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA StatementThe PRISMA Statement. Annals of Internal Medicine, 151(4), 264-269

Neely, J. G., Magit, A. E., Rich, J. T., Voelker, C. C., Wang, E. W., Paniello, R. C., ... & Bradley, J. P. (2010). A practical guide to understanding systematic reviews and meta-analyses. *Otolaryngology--Head and Neck Surgery*, *142*(1), 6-14.

Power, N., & Alison, L. (2016). Offence or defence? Approach and avoid goals in the multi-agency emergency response to a simulated terrorism attack. *Journal of Occupational and Organizational Psychology*. doi:10.1111/joop.12159.

Salas, E., Grossman, R., Hughes, A. M., & Coultas, C. W. (2015). Measuring team cohesion: Observations from the science. *Human factors*, *57*(3), 365-374.

Van Vianen, A. E., & De Dreu, C. K. (2001). Personality in teams: Its relationship to social cohesion, task cohesion, and team performance. *European Journal of Work and Organizational Psychology*, *10*(2), 97-120

Zaccaro, S. J., Gualtieri, J., & Minionis, D. (1995). Task cohesion as a facilitator of team decision making under temporal urgency. *Military Psychology*, *7*(2), 77.

Zsambok, C.E., & Klein, G.A. Eds. (1997). *Naturalistic decision making*. Mahwah, NJ: Erlbaum.

Weak Signals in Healthcare: A Case study on Community-Based Patient Discharge

Eva-Maria CARMAN[a], Mike FRAY[a], and Patrick WATERSON[a]
[a]*Human Factors in Complex Systems Research Group, Loughborough Design School, Loughborough University, Loughborough, UK*

ABSTRACT

To adjust performance to ensure the success of a task and prevent error, it is necessary to anticipate, identify and respond to variations in the work system. The objectives of this study were to develop a framework for the analysis of signals, which provide an indication of variations in the system, in the healthcare environment and qualitatively investigate signals in the context of community-based patient discharge. In addition to the signals, both traditional (Safety-I) and proactive safety (Safety-II) elements were investigated with six expert groups, from the field of community-based patient discharge. The signals identified and the safety elements were analysed using the SEIPS 2.0 model. The sources of the signals were identified as originating from work system elements. The proposed framework and method provide a preliminary basis for the investigation of signals and assists in highlighting the role that these can play in safety behaviour.

KEYWORDS

Sensemaking; Health; Weak Signals; SEIPS; Safety-II.

INTRODUCTION

Healthcare can be described as a complex socio-technical system (Braithwaite, Clay-Williams, Nugus, & Plumb, 2013; Buckle, Clarkson, Coleman, Ward, & Anderson, 2006; Carayon & Friesdorf, 2006; Janowitz, Gillen, Ryan, Rempel, Trupin, Swig, Mullen, Rugulies & Blanc, 2006), which due to the rapidly changing and dynamic environment, exhibits properties of emergence, adaptation and self-organization (Braithwaite et al., 2013). Emergent behaviours result as numerous tasks require a more improvised response due to the unpredictable nature and the unanticipated events that comprise the daily routine in this field (Braithwaite et al., 2013; Kerfoot, 2004). As a result of the work and system nature of the healthcare environment, safety in this context not only refers to the safety of the workers but more predominantly to the safety of the patients. Patient safety can be effected by errors (Ahmed, Adam, & Al-Moniem, 2011), which can then potentially lead to adverse incidents or events (Brown, Hofer, Johal, Thomson, Nicholl, Franklin & Lilford, 2008). Additionally numerous factors, including organisational factors (Johnson, 2004) affect patient safety and therefore it is key to look at patient safety from a system's perspective.

Research in healthcare will benefit from adopting not only the traditional definition of safety but the more recently developed definition of Safety-II. The concept of safety has recently been expanded to include both a definition of an absence of harm, whereby the number of adverse events is as low as acceptably possible labelled as Safety-I (Hollnagel, 2014), as well as a definition that focuses on the ability to succeed under varying conditions, labelled Safety-II (Hollnagel, 2014). One element of Safety-II is the ability to adjust performance to ensure success of the task and this requires anticipating, identifying and responding to signals indicating changes in the system (Hollnagel, 2014). Signals can be defined as sensed information regarding emerging events (Ansoff & Mcdonnell, 1990), and include indicators or cues from the environment (Rasmussen, 1983) which require interpretation and sense-making (Weick, 1995). The strength of these signals can vary resulting in different requirements regarding interpretation and abilities of sense-making. Strong signals provide a specific indication and are more readily recognized (Guillaume, 2011) whereas weak signals are often vague in nature (Ansoff & Mcdonnell, 1990), and need to be actively sought out and created by processing interrelated existing events, prior knowledge and future expectations in order to understand the information they provide (Macrae, 2014a). Through the early detection of unexpected events, they may be addressed in a more cost-effective and timely manner (Vogus & Sutcliffe, 2007), but failure to notice the warning signs may result in the risks being normalised, and remaining dormant until an adverse event occurs (Macrae, 2014a, 2014b).

Weak signals may provide an opportunity to achieve proactiveness through the required awareness, monitoring and constant vigilance needed for the identification of these signals and the up-to-date information regarding ongoing operations they provide (Vogus & Sutcliffe, 2007). Effective risk management requires continuous identification and addressing the problems that threaten safety (Macrae, 2014b) and identifying weak signals may offer means of reducing risk and responding to hazards earlier. This highlights the role that weak signals can play in safety behaviour. Despite accident reports increasingly stating signals were present prior to an incident that would have altered the course of the event if they had been acted upon, research exploring weak signals and the role they may play in safety, especially in healthcare, is limited.

The aims of this research were two-fold, namely develop a conceptual framework for the investigation of weak signals in the healthcare environment and to explore both Safety-I and Safety-II elements in the environment. The results in this paper include the developing conceptual model as well as the results from the first case study.

METHOD

Framework

To develop an integrated framework, literature was drawn from different fields to develop a conceptual framework for the investigation of weak signals to be assessed in upcoming case studies. The framework was developed for the analysis of weak signals in the context of the work, actions and events in the system in which they occur, specific for the healthcare context. This framework was developed based on literature from strategic management theory (Ansoff & Mcdonnell, 1990), systems ergonomics (Holden et al., 2013; Karsh et al., 2006), naturalistic decision making theory (Zsambok & Klein, 1997), the work on weak signals by Macrae (2014a), as well as the work on error by Rasmussen (1983) and Reason (1991). It is aimed through the use of different case studies conducted within the healthcare context, that the model will be verified and expanded on. The aim of the framework is to enhance the understanding of signals, specifically where they may originate and how they may be manifested specifically in healthcare, which will aid the development of training and tools utilising weak signals.

Research Design

An explorative qualitative method was adopted to investigate aspects that lead to performance failure and success as well as weak signals within the community-based patient discharge field using a focus group methodology. This qualitative approach was adopted due to the fuzzy nature of weak signals and as the field of Safety-II is still emerging. The focus groups drew upon the experience of the staff involved in the discharge process for patient discharge into the community to investigate why discharges fail, could the failure be prevented and the characteristics that ensure that the discharge is a success. A total of 6 focus groups across three directorates of the Nottinghamshire Healthcare NHS Foundation Trust were conducted.

Participant Characteristics

A total of 39 participants to part across the six focus groups. For five of the six focus groups, seven participants partook in the focus groups with the sixth focus group having a total of four participants. The mean total number of years involved in patient care across all six focus groups was 16.6 years (±10.6) and the mean number of years in the current position was 3.6 years (±3.6). The current positions held by the participants in the focus groups included community and district nurses, locality managers, community physiotherapists and occupational therapists, assistant practitioners, and a team leader of a care home team.

Protocol

Each focus group comprised of two consecutive components, where the focus of each component was on aspects and system structures that promote success (in line with the Safety-II definition) and aspects that could go wrong, influencing factors and possible weak signals present (in line with the Safety-I definition). Each component was approximately 45 minutes in duration with a 20-minute break between the two sessions. Prior to the start of the focus group, the project was described to the participants and the project information sheet, the informed consent sheet, as well as demographic information sheet was distributed among the participants, and returned before the start of the first component in the focus group.

The emphasis of the first component in the focus group was on the elements of the discharge process that work well and improve patient safety. These questions were developed using the SEIPS 2.0 model by incorporating the work system elements (Holden et al., 2013) and based on literature on Safety-II. During the main discussion, the group was encouraged to develop a definition of a good discharge from the perspective of the staff as well as that of the patient. Participants were encouraged to discuss how stable their work conditions are, if their work requires a high degree of improvisation and how predictable the work situation is. Following the development of the definition, the discussion was guided by one of the researchers through the following series of questions:

- What is the best way or optimal way to perform your work? What personal elements ensure a good discharge? (Person-related) What needs to be in place (requirements)?
- What can happen unexpectedly during the task and how do you prepare for it? (Task-related)
- Are tools in place that assist with this? (Tool-related)
- What do you require from your team/unit for the discharge to be a success? (Team/group/unit/department)
- What organizational elements assist in ensuring the discharge is a success? (Organizational factors)

The emphasis of the second component in the focus group was on the potential elements for error recovery and identification of weak signals. The main discussion was guided through the following series of questions by one of the researchers:

- What could go wrong with this task? (Error)
- What external factors would influence this task? (External Factors)
- How do you know the task is going wrong? (Signals)
- When you know it is going wrong, how do you correct yourself? Can you pre-empt the task? (Reaction/Monitoring)
- Do you use this knowledge next time you do this task? (Learning)

All focus groups were conducted using the question structure described above. The questions for the two components were used to loosely guide the discussion, but the participants were encouraged to freely discuss any topics that arose as a result of these questions. The discussions for both components of the focus groups were recorded using two audio recorders and one researcher recording field notes. During both discussions, one of the researchers compiled a summary of the key points raised by the group in the discussion on a white board or flip chart. The results were qualitatively analysed using a themantic analysis approach (Braun and Clarke, 2006) by identifying common themes across the groups and the signals identified were categorised using the SEIPS 2.0 model. Ethical approval for this project was granted by the Loughborough University Ethics Approval (Human Participants) Sub-Committee and the Nottinghamshire Healthcare NHS Foundation Trust.

RESULTS

Framework

This work draws on research from numerous fields, including human factors, strategic management theory, natural decision making theory as well as the concepts of safety-I and safety-II to expand the knowledge and understanding of weak signals. The framework was developed for the analysis of weak signals in the context of the work, actions and events in the system in which they occur specific for the healthcare context and the preliminary framework is depicted in Figure 1.

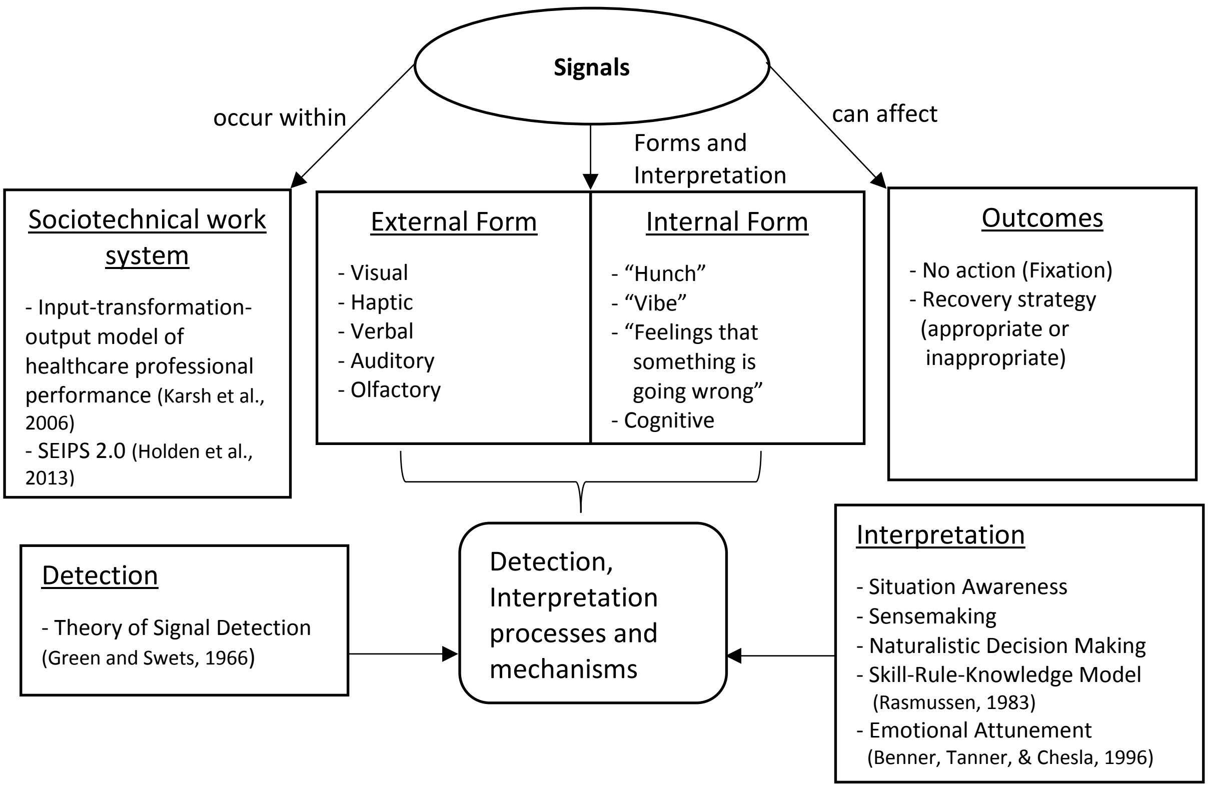


Figure 1: A conceptual framework for the investigation of weak signals within the healthcare context.

The work by Ansoff (1975) on weak signals in strategic management theory and the work by Macrae (2014a) on weak signals in aviation were used to provide the basic definition and premise for the conceptual framework developed for the investigation of weak signals. The left aspect of the framework included the Input-transformation-output model of healthcare professional performance (Karsh et al., 2006) in order to determine the aspects of the system from which the signals could originate. This model was selected as it provides a general multi-level model of a work system, as well as having considered open systems theory. The framework was then further expanded to incorporate more specific elements from the second version of the Systems Engineering Initiative for Patient Safety (SEIPS) model (Holden et al., 2013). The SEIPS 2.0 model was selected as it provides a framework for the analysis of processes and the relationship of various elements that occur in healthcare specifically (Carayon et al., 2006).

The forms of the signals have been described in the framework as either being internal or external. An external signal may also generate an internal signal, but the external source or signal that causes the experience of an internal signal may not always be present or known. The external signals include visual, haptic, verbal, auditory or olfactory cues. The internal signals include the experience of a "hunch", "vibe" or a general sense of "something going wrong". Signal detection theory (Green and Swets, 1966) was included in the framework as it provides possible factors that may influence how a signal is perceived. These include the strength of the signal and the individual's bias in perceiving the signal (McLeod, 2015). Theoretical concepts and models included to explain possible interpretation processes and mechanisms in the framework consist of the concepts of situation awareness (Endsley, 1995), sensemaking (Weick, 1995), naturalistic decision making (Zsambok & Klein, 1997), emotional attunement (Benner, Tanner, & Chesla, 1996) and the skill-rule-knowledge model of behaviour (Rasmussen, 1983). The skill-rule-knowledge model of behaviour (Rasmussen, 1983) was included as it may explain the processing and influence signals may have on performance. Signals can affect outcomes in that as a result of fixation (Reason, 1991) no action may be taken or alternatively a recovery strategy is implemented which may either result in an appropriate or inappropriate outcome. By considering the source and type of information these signals provide, insight regarding the status of the system and areas of risk may be revealed (Macrae, 2014a).

Weak Signals

In the second component of the focus group, participants were asked to discuss how they knew the discharge may not be going as expected and the signals that indicated this. Additonally it was discussed how they would respond to this and make an adjustment or adaptation they thought was necessary. These signals, if they are acted upon, have the potential to change the progression of the task and may assist in ensuring a successful outcome.
The identified signals were analysed using the SEIPS 2.0 model by identifying from which element in the sociotechnical work system they originated. The sources of the signals could be categorized as the following elements from the work system: "person", "tasks", "tools" and "internal environment". Examples identified in the focus groups and the categorization of these examples is presented in Table 1.

Table 1. Examples of the signals identified in the focus groups and their classification with regards to the sociotechnical work system and from

Example given	Source (Sociotechnical Work System)	Form
Patient's physical state *For example*: Patient does not look well Patient is not at the anticipated level Patient's behaviour Cleanliness of Patient	Work system: Persons	Examples consist of external forms of signals including visual cues and olfactory cues, as well as internal forms such as cognitive cues.
State of patient's home *For example:* Untouched medication Cluttered environment Cleanliness of home environment	Work system: Internal environment	Examples consist of external forms of signals including visual cues and olfactory cues
Family's response *For example:* Numerous phone calls Family is intense or disengaged Family/Carers look as if they are not coping Family's expectations	Work system: Persons	Examples consist of external forms of signals including visual cues and auditory cues.
Patient documentation *For example:* Prior phonecall Referral	Work system: Tools	Examples consist of external forms of signals including auditory cues, as well as internal forms such as cognitive cues.
Patient history *For example:* Previous experience with the patient Known psychological disorder	Work system: Tasks	Examples consist of internal forms of signals including cognitive cues.

Examples of signals originating from "person(s)" in the system included the patient's physical state and feedback from the patient and their family. This is highlighted by the following quote from one participant: *"they're [the patient] telling you they are not coping"*. Additionally the indicators from feedback from the family are highlighted by the following quote: "*Just to add to that, the patient might tell you their fine, but then you might get a family member go… into the kitchen… and 'can I have a word'. They will tell you they are fine but they are not"*. Signals originating from the "internal environment" included the state of the patient's home. Signals originating from the "tools" included the signals generated as a result of patient documentation such as the referrral form. The referral form may contain information that could possibly act as a warning signal to community staff prior to them visiting the patient, for example if the patient is a "high volume service user" with a history of failed discharges.

Following this, the examples were classified into the different forms of signals according to the framework depicted in Figure 1. The examples related to the patient's physical state that could be categorised as external included visual cues, for example seeing the patient does not look well, olfactory cues which would provide an indication of if the patient is not coping with tasks of daily living, such as personal hygiene. Internal signals identified included cognitive cues, for example the patient not being at the anticipated level of recovery. This type of signal was classified as internal, as it would have originated based on information processing from other external cues such as the documentation of the current state of the patient and the external visual cue that would then highlight the indiscrepancies between the two sources of information.

The participants mentioned that they felt that the identification of these signals is a necessary component of their current work as their work requires them to adapt the patient's treatment plan accordingly so that a readmission would not occur. In response to these signals, several participants mentioned that they would restructure their time, delegate work to other staff members and link in with other services to address the emerging issues. Additionally, they mentioned updating documentation, reporting back to the care teams, and setting up follow-ups as possible response to the signals. In the groups, it was also mentioned that it was not always possible to respond to the signals identified. The final aspect discussed during the focus groups considered the potential learning opportunities and how one would pass the knowledge regarding signal identification on to more junior members. These included the need for reflection with different staff members, regular meetings to allow for feedback, team handovers to share best and worst practice, clinical supervision, and sharing experiences and information with different occupational groups.

Safety I and Safety II Aspects

The safety aspects discussed in the focus groups included aspects which could go wrong including errors, influencing factors and the various system elements that aid in task success. Common errors identified included errors relating to inappropriate or missing equipment, missing medication and inadequate packages of care. Error producing conditions identified by the groups consisted of a lack of communication between the different services involved in the process, and missing or incomplete information or documentation. An example of the effect and resulting problems a lack of cmmunication can cause is highlighted in the quote: "patients gone home to a different place, can't find your patient but you know they are out".

Potential factors that would influence the task and task-related behaviour identified by the groups included patient-related factors, time-related factors, and organizational factors. The organizational and managerial factors identified may not only influence the worker and task but also may affect an individual's ability to adapt and adjust their performance. The elements that assist in promoting a good discharge were categorized according to the SEIPS 2.0 model, and the results for the different elements are depicted in Figure 2.

Examples of work system elements that promote a good discharge categorized as person-related, included experience, open communication, ability to improvise, and having the confidence in asking questions. An example of open communication and the importance of it is highlighted by the following quote by one of the participants: *"where there's been some level of communication between where they're coming from to where they're going to ensures that there's a smooth transition of care"*. Additional examples categorized as person-related included utilizing all available resources, understanding the job-roles of the individuals involved in the process, good team work and being proactive, for example by "chasing" discharges for that day in the morning. This is highlighted by the one participant: *"I think for me its chasing up that discharge early in the morning, making sure its planned for that day. So now when we have future discharges I will ring in the morning to make sure all the plans are done and that they've not moved ward."*. Examples of task-related elements included the information provided during the task being up-to-date and accurate, effective cooperation and coordination between the services involved, good timing of tasks, for example other services are timeously informed about the discharge and ensuring cut-off times are considered and maintained to ensure patient safety, as well as the necessary and appropriate equipment being in place prior to the patient being discharged. Examples of tool-related elements included well completed documentation forms including referrals, therapy forms and discharge letters, having access to computer records to access the latest information on the patient, and standard operating procedures. The standard operating procedures were described as an aid in specific cases in that it eased the discharge process across localities, which is explain in the following quote: *"but sometimes you need to fall back on standard operating procedures so that you can kind of, everyone is going from the same sheet across the localities"*. Examples of organization-related elements included organizational structures such as integrated teams, good intra-organizational communication and designated staff members, such as having a key contact person within the acute hospital. An example of how integrated teams promote a good discharge is highlighted in the following quote: *"I think we all support each other and be in integrated teams and working with the physios and OTs. You have got access to people, sending them [the patient] back into hospital is always going to be the last resort, it would be a visit of what can we get into place, what equipment can we get, who can we get involved to try keep them at home"*.

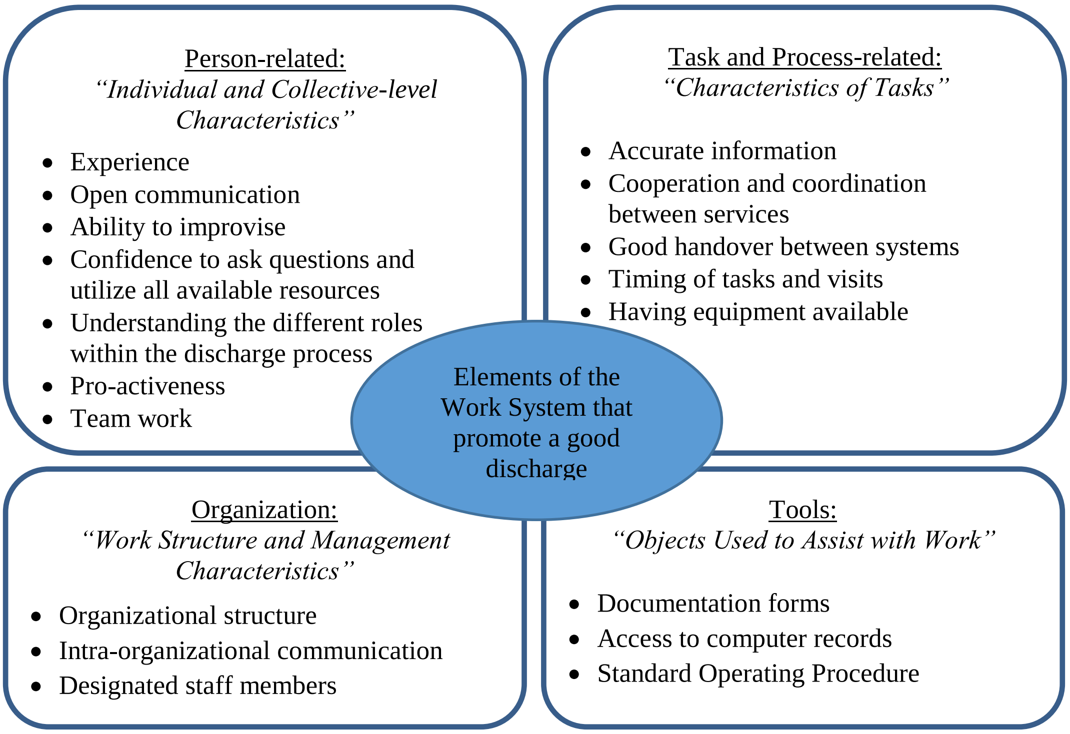


Figure 2: The categories of the results found for the different elements of the work system from the SEIPS 2.0 model that promote a good discharge.

CONCLUSION

The method above qualitatively investigated Safety-I elements, signals as well as other Safety-II elements. The Safety-I elements addressed in this study included specific potential errors and error producing conditions related to the discharge process that may result in adverse events whereas the Safety-II elements investigated included signals, learning opportunities and work elements that may assist in task success. These Safety-II elements may aid in rendering a system more resilient by improving the ability to succeed under varying conditions (Hollnagel, 2014). Additionally, weak signals may also provide a means for effective risk management in that they provide an opportunity to be proactive and identify aspects that threaten safety (Macrae, 2014a), and consequently one can respond to hazards earlier.

The proposed method provides a preliminary basis for the investigation of signals and work elements that may aid in task success. The ramification and practical implication of this research is that it provides a basis for developing work processes so that current aspects that staff feel work well are incorporated into new procedures. Additionally incorporating Safety-II aspects when designing an intervention, whether it is a tool, training or redesigning a work process, by incoporating aspects staff feel work well, it may promote greater ownership of the intervention. The model on weak signals may assist in the development of a possible tool to triage and maybe highlight discharges tht may be particularly difficult sooner rather than later.

The participants provided positive feedback regarding the focus groups as a source of sharing ideas and as a platform to discuss task aspects that work well. Further investigations are required to identify additional elements that aid in task success as well as the factors that promote or inhibit signal identification.

ACKNOWLEDGMENTS

The project was commissioned as a result of a successful joint bid for funding by Health Partnerships, a Division within Nottinghamshire Health Care NHS Foundation Trust and Loughborough University Design School. This project was conducted in partnership with Health Partnerships, Loughborough University Design School and Nottingham University Hospitals (NUH). The authors are grateful to all the colleagues at the Health Partnerships, Loughborough University Design School and NUH for their help and support with data collection and analysis.

REFERENCES

Ahmed, N. G., Adam, S. M., & Al-Moniem, I. I. A. (2011). Patient safety: Assessing nurses' perception and developing an improvement plan. *Life Science Journal*, 8(2), 53–64.

Ansoff, I. (1975). Managing Strategic Surprise by Response to Weak Signals. *California Management Review*.

Ansoff, I., & Mcdonnell, E. (1990). *Implanting strategic management* (Second Edi). United Kingdom: Prentice Hall International.

Benner, P., Tanner, C. A., & Chesla, C. A. (1996). *Expertise in nursing practice: caring, clinical judgement and ethics*. Springer Publishing Company.

Braithwaite, J., Clay-Williams, R., Nugus, P., & Plumb, J. (2013). Health Care as a Complex Adaptive System. In E. Hollnagel, J. Braithwaite, & R. L. Wears (Eds.), *Resilient Health Care* (pp. 57–73). Ashgate Publishing Ltd.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.

Brown, C., Hofer, T., Johal, a, Thomson, R., Nicholl, J., Franklin, B. D., & Lilford, R. J. (2008). An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Quality & Safety in Health Care*, 17(3), 170–7. http://doi.org/10.1136/qshc.2007.023655

Buckle, P., Clarkson, P. J., Coleman, R., Ward, J., & Anderson, J. (2006). Patient safety, systems design and ergonomics. *Applied Ergonomics*, 37(4), 491–500. http://doi.org/10.1016/j.apergo.2006.04.016

Carayon, P., & Friesdorf, W. (2006). Human Factors and Ergonomics in Medicine. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (Third, pp. 1517–1537). Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/0470048204.ch58

Carayon, P., Schoofs Hundt, A., Karsh, B.-T., Gurses, A. P., Alvarado, C. J., Smith, M., & Flatley Brennan, P. (2006). Work system design for patient safety: the SEIPS model. *Quality & Safety in Health Care*, 15 (Suppl 1), i50-8. http://doi.org/10.1136/qshc.2005.015842

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32–64.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Guillaume, E. G. (2011). *Identifying and responding to weak signals to improve learning from experiences in high-risk industry.* TU Delft, Delft University of Technology.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, a A., & Rivera-Rodriguez, a J. (2013). SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, 56(11), 1669–86. http://doi.org/10.1080/00140139.2013.838643

Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. Farnham: Ashgate Publishing Ltd.

Janowitz, I. L., Gillen, M., Ryan, G., Rempel, D., Trupin, L., Swig, L., Mullen, K., Rugulies, R. and Blanc, P. D. (2006). Measuring the physical demands of work in hospital settings: design and implementation of an ergonomics assessment. *Applied Ergonomics*, 37(5), 641–58. http://doi.org/10.1016/j.apergo.2005.08.004

Johnson, K. (2004). Keeping Patients Safe: An Analysis of Organizational Culture and Caregiver Training. *Journal of Healthcare Management / American College of Healthcare Executives*, 49(3), 171–179.

Karsh, B.-T., Holden, R. J., Alper, S. J., & Or, C. K. L. (2006). A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. *Quality & Safety in Health Care*, 15 (Suppl 1), i59-65. http://doi.org/10.1136/qshc.2005.015974

Kerfoot, K. (2004). Attending to weak signals: the leader's challenge. *Pediatric Nursing*, 30(1), 79–80.

Macrae, C. (2014a). *Close Calls*. Palgrave Macmillan.

Macrae, C. (2014b). Early warnings, weak signals and learning from healthcare disasters. *BMJ Quality & Safety*, 23(6), 440–5. http://doi.org/10.1136/bmjqs-2013-002685

McLeod, R. W. (2015). *Designing for Human Reliability: Human Factors Engineering in the Oil, Gas, and Process Industries.* Elsevier Science.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3), 257–266. http://doi.org/10.1109/TSMC.1983.6313160

Reason, J. (1991). Too little and too late: a commentary on accident and incident reporting systems. *In Near Miss Reporting As a Safety Tool* (pp. 9–26). Elsevier. http://doi.org/10.1016/B978-0-7506-1178-7.50006-X

Vogus, T. J., & Sutcliffe, K. M. (2007). Organizational resilience: Towards a theory and research agenda. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 3418–3422. http://doi.org/10.1109/ICSMC.2007.4414160

Weick, K. E. (1995). *Sensemaking in Organizations*. London: SAGE Publications.

Zsambok, C. E., & Klein, G. A. (1997). *Naturalistic Decision Making*. L. Erlbaum Associates.

How can we design Tactile Interactive software for Argument Construction in Criminal Intelligence Analysis?

Celeste GROENEWALD, B.L.William WONG, Simon ATTFIELD, Peter PASSMORE and Neesha KODAGODA

Middlesex University, London, UK

ABSTRACT

Argumentation construction refers to the different ways in which people can formulate a well-defined argument that can withstand scrutiny in a court of law. Different domains have enhanced the research on argumentation construction, but each enhancement relates either to 'evidence' or to the 'relevance of evidence', thus making these elements the corner stone elements of argumentation construction. By attempting to understand how Criminal Intelligence Analysts understand and manage these corner stone elements, as well as how these elements differ to the law domain counterparts, we aspire to formulate design guidelines for a software program that is tailored to how Criminal Intelligence Analysts think and argue during sense-making activities. This paper outlines the relevant literature and why the researchers consider 'evidence' and the 'relevance of evidence' as the corner stone elements of argumentation construction. The results section summarises the outcomes of two qualitative studies. The first study aims to understand how Criminal Intelligence Analysts perceive and manage uncertain information and how this eventually leads to the creation of evidence (as exhibits) for a court of law. The second study aims to understand how Criminal Intelligence Analysts know which information is relevant for the task at hand, especially when uncertainty is high. The results of the two studies still need to be combined and put into practise (as design guidelines) to test the effectiveness and validity of the results obtained in the two studies. This is listed as future work in the conclusion section.

KEYWORDS

Uncertainty; Sense making; Argumentation Construction; Security; Government and Law

INTRODUCTION

The aim of this research is to answer the question, "How can we design Tactile Interactive software for Argument Construction in Criminal Intelligence Analysis?" This question originated by working on the Visual Analytics for Sense Making in Criminal Intelligence Analysis (VALCRI) project, in an attempt to understand how Criminal Intelligence Analysts (from now on referred to as Analysts) think, which then could inform designers on how to design software programs to assist Analysts with their daily analytical activities.

The VALCRI project has various requirements, expressed as design guidelines, which influence the direction and nature of research on this project. The user interaction requirement for this project is outlined to support tactile reasoning, which is defined as an "interaction technique that supports sense-making by the direct manipulation of information objects in the user interface" (Takken and Wong, 2013). Takken and Wong's (2013) reasoning for this is that "when one is presented with a set of information that can be freely moved, manipulated, grouped or re-arranged in a visuo-spatially manner, this interaction method can help us discover meanings or relationships". Their experiment found that participants were able to construct twice as many relationships between items in a dataset, when they were allowed to interact directly with the data items using their hands, than the group who were deprived thereof.

Another requirement of the VALCRI project is to support Analysts with argumentation construction. Argumentation construction refers to the different ways in which people can formulate a well-defined argument that can withstand scrutiny in a court of law. The argumentation construction process must therefore form part of the tactile reasoning process, from the onset of sense-making activities and not as an afterthought. This is a challenge, because it implies that Analysts must be afforded the ability to construct arguments from the onset of sense-making activities, even when uncertainty is high, when little or no information is available and the construction of an argument should not hinder the flow of the sense-making process. Analysts concentrate on the exploration of information when uncertainty is high, so that they can gain traction to reach one or more conclusions. The main purpose of argumentation construction, as found in the law domain, is to verify various conclusions. There is therefore a gap in what the VALCRI project aims to achieve and what the available literature offers on argumentation construction.

The research in this paper attempts to understand and address the gap between what the VALCRI project aims to achieve with argumentation construction and what is currently available within the law domain literature. The literature within the law domain was selected as a starting point, is it contains ample research on the topic of argumentation construction. The literature review section outlines 'evidence' and the 'relevance of evidence' as two of the corner stones of argumentation construction. This also forms the basis for the undertaking of the two qualitative studies outlined in this paper. The first study attempts to understand how Analysts perceive information when uncertainty is high and how this leads to the creation of evidence (as exhibits) for a court of law, thus addressing the first corner stone of argumentation construction. The second study aims to understand how Analysts know which information is relevant for the analytical task at hand, especially when uncertainty is high, thus addressing the second corner stone of argumentation construction.

The section below briefly outlines the relevant literature and why the researchers consider 'evidence' and the 'relevance of evidence' as the corner stone elements of argumentation construction.

LITERATURE

Argumentation has been around since Pre-Socratic times and Wigmore contributed to argumentation research in 1931, by creating an argumentation diagram (also known as an argumentation map). This shows the relevant components of an argument as a graphical representation and it was designed to assist the law domain with the construction of robust arguments for trial cases in a court of law. The notation was however difficult to learn and understand until Toulmin in 2003 presented a simplistic model for representing arguments in general. Toulmin's model is now preferred by scholars over Wigmore, but most still reverts to the hierarchical layout that Wigmore introduced to represent the argumentation components and the relationships between them. In the work of Pennington and Hastie (1993) which started as early as the nineteen eighties, introduced the role of narrative as way to convey and explain the argumentation components. This extended the concept of argumentation to have a social component, which assists the logical and rational side. But, it is this overwhelming logical and rational side which prompted the integration of artificial intelligence and law (AI&Law) in the mid nineteen nighties, by looking for ways to assist the formulation of argumentation using quantitative models (Nissan, 2009). Even though different scholars have progressed research on argumentation construction, they all seem to have two elements in common from which these enhancements were made possible. The section below outlines the relevant research and the common argumentation elements of 'evidence' and the 'relevance of evidence'.

Toulmin

When looking at Toulmin's work (2003), he has formulated six components in constructing a general argument namely, *claim*, *datum*, *warrant*, *backing*, *qualifier* and *rebuttal*. In order for a claim (or conclusion) to be accepted as true, Toulmin asks us to formulate our persuasion by means of a datum (or grounds of acceptance or premise). If the validity of the datum is under question, then a warrant is used to link the datum with other ground truths, which shows the relevance of the datum in support of the claim. One can strengthen a warrant with additional backing in the form of supporting information to strengthen the relevance. A qualifier indicates the applicability of the link from the datum to the warrant and could be used as a means to ring-fence the argument to be relevant to a set of conditions. A rebuttal is a counter argument, which contradicts the initial claim. This can be summarised as follow: The datum serves as evidence, for example a birth certificate stating England as the country of birth, can be the grounds of acceptance for the conclusion that the person in question is British. This evidence can be backed up (backing) with law documents outlining that a person with a birth certificate, stating England as the country of birth, is a British citizen. The relevance or the applicability of the evidence can be through the use of a warrant, which links the evidence to additional ground truths such as: All people with a birth certificate which states that they were born in England, are legally regarded as British citizens. The qualifier further strengthens the relevance of the evidence by including that not only is it true for people who have England as a birth country, but this argument is also valid for people of Wales, Scotland and Northern Ireland. If a person would like to disprove the conclusion that the person in question is a British citizen, then it needs to be on the basis that the evidence is invalid (birth certificate is a counterfeit) or that the relevance of the law should be interpreted differently. The corner stones of Toulmin's argumentation structure is therefore 'evidence' and the 'relevance of evidence' and the constructs for manipulation is summarised in Table 1. This means that the datum and the backing can be used to (dis)prove the available evidence. Similarly, the warrant and qualifier can be used to (dis)prove the 'relevance of evidence'.

Table 7 – Summary of Toulmin's argumentation components in relation to 'evidence' and the 'relevance of evidence'

Scholar	Domain	Evidence	Relevance of evidence
Toulmin (2003)	Argumentation	Datum, Backing	Warrant, Qualifier

Wigmore

Wigmore diagrams (1931) are constructed using mainly the components of evidence and forces. Evidence are the statements or assertions made and the forces are the degree to which evidence supports (affirmatory) or opposes (negatory) another piece or pieces of evidence. Both the affirmatory and negatory forces play a role in determining the relevance of evidence as the affirmatory supports the argument and the negatory supports the counter-argument (rebuttal). For this reason, modern day lawyers are still focused on evidence and the relevance of evidence when constructing arguments (Allen et al., 2015). This can be summarised as follow: The person in question is British (as a statement). This statement is supported by the affirmatory force (AF-1), in the form of a birth certificate stating that the person in question was born in England. AF-1 is supported by the affirmatory force 2 (AF-2) in the form of the law, stating that people with birth certificates, stating that their country of birth is England, is recognised by law as British citizens. A negatory force (NF-1) against AF-2 could be in the form of the law that states that AF-2 is not applicable when the birth certificate is a counterfeit. The second negatory force (NF-2) against AF-1 could be in the form of a witness admitting that s/he counterfeited the birth certificate for the person in question. The corner stones of Wigmore diagrams are therefore 'evidence' and the 'relevance of evidence' and the constructs for the manipulation is summarised in Table 2. This means that the various statements can be regarded as evidence and the relevance of those statements are supported by forces, which either supports or disproves the evidence.

Table 8 - Summary of the components in Wigmore diagramd in relation to 'evidence' and the 'relevance of evidence'

Scholars	Domain	Evidence	Relevance
Wigmore (1931)	Argumentation & Law	Statements as evidence	Forces
Allan (2015)	Law	Legal Evidence	Relevance to legal consequences

Pennington and Hastie

Pennington and Hastie (1992) concluded that showing mere evidence was not enough in a court of law and that narrative played an important part in conveying the facts to the jury. Their work concentrates on knowing what is required to successfully *explain* or *explain away* either the 'evidence' or 'the relevance of evidence'. This is outlined in their Story Model and their model consists out of the following three components: "(a) evidence evaluation through story construction; (b) representation of the decision alternatives by learning verdict category attributes; and (c) reaching a decision through the classification of the story into the best-fitting verdict category" (Pennington and Hastie, 1992). Point (a) concentrates on 'evidence' and points (b) and (c) concentrate on the 'relevance of evidence' in relation to determining legal consequences. This can be summarised as follow: Although all of the evidence points to a counterfeited birth certificate for the person in question, the lawyer can convey this information to the judge in such a way to depict the person in question as a victim of his/her parent's decisions. In other words, the person in question did not know that the parents counterfeited the document and believed the document to be true and within the law. And since the person in question has been brought up in England since the age of one, the law should be lenient and take into consideration the person's good behaviour until the age of twenty-one and reward the person with British citizenship. The corner stones of Pennington and Hastie's narrative are therefore related to explaining the 'evidence' or the 'relevance of evidence' and the constructs for the manipulation is summarised in Table 3. This means that the jury could be convinced to interpret the 'evidence' or the 'relevance to the evidence' differently, depending on the message that the lawyer wants to convey.

Table 9 - Summary of the components in Pennington and Hastie's narrative in relation to 'evidence' and the 'relevance of evidence'

Scholars	Domain	Evidence	Relevance
Pennington & Hastie (1992)	Narrative & Law	Explain or explain away	Explain or explain away

Bex at al.

Bex et al.'s (2010) hybrid theory concentrates on explaining the events in a judicial case by using narrative to explain the relevance of the causal connections between the events and using the evidence to anchor the facts in general and acceptable common sense rules. This can be summarised as follow: The events that led to the counterfeited birth certificate could be (a) The parents illegally immigrating to England; (b) The parents telling the child that s/he was born in England; (c) The child growing up believing that s/he is British; (d) The child, now a young adult, finding out that the birth certificate is a counterfeit.

Each event can be grounded in evidence or acceptable common-sense rules for example (a) counterfeit passports of the parents; (b) The parents did not want anybody to find out about their illegal status, especially to protect the child from being deported; (c) There was no reason for the child to distrust the information which the parents offered; (d) The young adult should not be held legally responsible for the parents actions and the law should assist him/her rather than deport him/her. The corner stones of Bex et al.'s (2010) hybrid theory is therefore 'evidence' and the 'relevance of evidence' and the constructs for the manipulation is summarised in Table 4. This means that the various causal relationships can be explained by narrative and each causal event can be backed up by evidence, either as physical documents or generally accepted common sense rules.

Table 10 - Summary of the components in Bex et al.'s (2010) hybrid theory in relation to 'evidence' and the 'relevance of evidence'

Scholar	Domain	Evidence	Relevance
Bex et al. (2010)	AI & Law	Evidence as Anchors	Narrative to explain causal relationships

Artificial Intelligence

Artificial Intelligence have examined the possibility of using Bayesian logic (Nissan, 2009) to assign probabilities to different pieces of evidence and can use knowledge maps and ontologies to determine the relevance of evidence to a particular legal consequence. This in turn allows an artificial agent to quickly review the available evidence and then construct arguments and counter-arguments, which should ease the burden on lawyers, judges and juries. This can be summarised as follow: The probability of the young adult's knowledge of the illegal acquisition of the false birth certificate (evidence) could be calculated based on the normal behaviour of child-parent behaviour as defined in the knowledge graph or ontology (relevance of the evidence). If the probability on young adult's knowledge of the illegal behaviour of the parents is less than 51%, then it could be concluded that the young adult is innocent and should be exempted from deportation. The corner stones of AI & Law is therefore 'evidence' and the 'relevance of evidence' and the constructs for the manipulation is summarised in Table 5. This means that the various evidential pieces and its relevance to the current argument, can be explained by the weights of the evidence and the relation to the various relationships, as defined in the ontologies or knowledge graphs.

Table 11 - Summary of the components in AI in relation to 'evidence' and the 'relevance of evidence'

Scholar	Domain	Evidence	Relevance
Nissan (2009)	AI & Law	Assign probabilities	Link to knowledge graphs or ontologies

The section below looks at how the sense-making domain regards evidence and the process of attaining it.

Pirolli and Card's (2005) Notional Model of Analytical Sense making provided researchers with a model, clearly illustrating the complexities of the processes involved to move from raw data to presenting a conclusion to the decision makers. These processes involve sixteen steps governed by three main loops that consist out of foraging, sense making and reality-policy loops. Foraging involves direct manipulation of the data such as searching, filtering, reading, extracting and the organisation of data. The sense making loop involves the building of a case by means of schemas, hypotheses and searching for support to substantiate these theories. The third loop governs the policies and procedures under which the other two loops operate.

Wong (2014) built upon this by defining the concepts of fluidity and rigour. Fluidity refers to how easily the system allows the Analyst to express their various types of thinking processes and rigour refers to the ability to test the validity of outcomes of processes. In our research we have found that fluidity and rigour can also refer to the type of sense making that is required to occur at different stages of the analysis. Fluid sense making refers to the exploration of ideas and the extension of knowledge. Rigour sense making refers to the verification stage of sense making when the Analyst is required to verify the correctness of ideas and knowledge gained. Both of these stages require Analysts to be "rigour" in the form of due-diligence, as it is part of their obligations. Ultimately, the system's design should match fluidity and rigour through interaction and processes, as well as the fluidity and rigour of sense making in the form of exploration and verification.

Klein et al. (2007) introduced the concept of using data elements as anchors to "create understanding and guide subsequent inquiry". Wong and Kodagoda (2015) used this and further investigated how the analytical inferencing process propagates from one conclusion to another and suggested that Analysts use both data and non-data anchors, which affords the process of gaining traction to reach a conclusion. These anchors have a variable level of certainty attached to it, which in turn influence the certainty of the conclusion reached. This falls in line with the guidelines of the Association of Chief Police Officers (ACPO) Core Investigative Doctrine, which states that Analyst's conclusions should adhere to the highest degree of certainty to minimise personal bias and stereotyping (ACPO, 2005).

Evidence in the law domain therefore serves as anchors which validates the certainty of a conclusion, whereas the anchors in the sense making domain servers as exploratory elements to assist with reaching a conclusion of variable certainty. In the law domain, evidence refers to something tangible that could be handed over to a jury for inspection (i.e. a murder weapon), proven facts (i.e. a medical examiner's DNA report), witness statements or a collection of carefully formulated exhibits as support for evidential facts (Allen et al, 2015). In the sense-making domain, the Criminal Intelligence Analysts mainly works with information which contains very few facts from the onset (if any at all) and their job is to determine the facts and create exhibits as evidence for a court of law. It is

this difference between what is considered as evidence, the certainty it encapsulates and the role it plays, what sets the fluidity and rigour stages apart within sense making activities in relation to argumentation construction.
This suggests a rethink about how argumentation schemas can be used in the sense making domain to capture the Analysts' creative and exploratory path in finding evidence, which eventually will support their conclusion with the highest degree of certainty, but which does not hinder them from reaching a conclusion or forcing them to commit to outcomes prematurely.

If evidence and relevance can be regarded as the foundation concepts of argumentation, then understanding how the sense making domain understands these concepts, might shed light on how to construct arguments in this domain. Various other considerations play a role. The VALCRI system is an intentional system, which means that the system's functionality should not only match the procedural processes of the analytical task, but also the thinking processes (as intent) of the Analyst. This means that we as researchers should understand how Analysts think and then design tools to assist those intentions, rather than hindering creativity with inflexible processes that will not allow them to continue to the next step until the current is complete. The underlying rigour should still be present, so that when the Analyst has found enough information s/he can test the validity and make the required adjustments as needed. We therefore strive to develop tools, which would not just follow one processing path, but which allows the Analyst to construct many paths from a point of intent in a visible and tactile manner. The law of requisite variety says that "R's capacity as a regulator cannot exceed its capacity as a channel for variety" (Ashby, 1958). In our case, this means that the design of the system should as a minimum, accommodate the known number of possible intentions from the Analyst. By researching 'evidence' and the 'relevance of evidence' as the corner stone elements of argumentation construction and how Analysts perceive each - should allow us to determine what those intentions are and then design our tools in such a way to accommodate each, thus satisfying the law of requisite variety.

The next section outlines the relevant research questions.

METHODOLOGY

The research questions of interest are:

- RQ1 How do Analyst perceive and manage (un)certainty?
- RQ2 How do Analysts determine and manage Relevance of Evidence and information?
- RQ3 How can the Analysts populate and use a low commitment argumentation schema?

RQ1 and RQ2 are answered through the data analysis process conducted on Cognitive Task Analysis (CTA) interviews from five experienced Operational Criminal Intelligence Analysts in the UK and Belguim. The VALCRI project aims to accommodate the thinking processes of all Analysts, regardless of where they are located, thus it was essential to gather data from both countries. Two senior interviewers, using the Critical Decision Method (Klein et al., 1986), investigated the inference and sense making processes of the Operational Criminal Intelligence Analyst participants from different police forces in the UK and Belgium. The interviewers wanted to understand how each Analyst resolved a particularly memorable case. Participants are more able to recall the details associated with a memorable case and the influencing factors it had, than just an ordinary case. Each interview lasted around sixty to a hundred minutes. A third party anonymised, transcribed and reviewed the transcripts due to the sensitivity of the contents it contained. A third junior researcher performed the data analysis on five of the transcripts from the CTA interviews. The data analysis was performed in line with the phases as outlined per Crandall et al. (2006) and the phases are, preparation, data structuring, discovering meanings and representing findings. The Open Coding technique as part of Grounded Theory (Corbin and Strauss, 1990) was used to code the data during multiple passes through the data. This in turn made it possible to find thematic themes in support of the research questions.

RESULTS AND CONCLUSION

The results of RQ1 are described in detail in the work of Groenewald et al. (2017a). By attempting to understand how Analysts perceive and manage uncertainty, allows for a better understanding on how Analysts reach a state where they can conclude on findings and subsequently construct exhibits as evidence in a court of law. The research results as found in Groenewald et al. (2017a) outline eleven problems that hinder sense making and which collectively influence the level of uncertainty surrounding outcomes of sense making activities. By considering the inverse of each problem, reveals the likely aspirations that Analysts could have had and why a particular strategy was successful in overcoming each problem. The identified aspirations are: certainty, believability, plausibility, simplicity, clarity, connectivity, creativity, finding new possibilities; finding meaning / information; determining correctness (data quality) and increasing understanding. These aspirations influence the collective level of certainty surrounding the outcomes of sense making activities, similarly, to how the problems contribute to the collective level of uncertainty. By understanding how Analysts gradually increase certainty of outcomes, may assist with understanding the requirements of assisting argumentation construction in the sense-making phase that are exploratory, with varying levels of certainty and low commitment from Analyst to use all of the outcomes.

The results of RQ2 are described in detail in the work of Groenewald et al. (2017b). By attempting to understand how Analysts identify significant information, allows for a better understanding on how Analysts perceive, manage and differentiate between information that is relevant and that which is not. The research results as found in Groenewald et al. (2017b) outline the concept of entities, which follows a lifecycle within a sense making environment. It was observed that these entities (which refer to both people and objects), are abstracted from the available information by the Analyst. They are then catalogued and their gradual changes in significance tracked for the duration of the analysis or until the entity loses its value in relation to the Analyst's goals. Significance refers to the implication the entity has on the current thinking regime of the Analyst. If the significance of an entity increases then the Analyst considers it as explaining or revealing more than what is currently known or understood and brings the Analyst closer to the answer. When the significance decreases then the Analyst found contradictory information, hit a dead end or the current explanation seems less likely than initially thought. So both types of significance enhances the Analyst's understanding, but in different ways and yielding different results. This abstraction of information into entities takes place when the most certain information is available or when the Analyst deems something interesting or strange. This ability to differentiate between what is normal and what is considered interesting and strange, may be a product of years of experience that the Analyst cultivated. The entity therefore has a lifecycle starting with creation, followed with growth by means of increased significance, stagnation through decreased significance, and then either death when the entity has no significance left or if there is no more information to continue the cycle. Resurrection is possible if new information becomes available and if that information is significant to the goals of the Analyst.

RQ1 assists with understanding how Analysts gradually increases their understanding surrounding analytical outcomes and how those outcomes evolve into conclusions with the highest possible level of certainty. RQ2 assists with understanding how Analysts orient themselves in the analysis and how they determine which entities have significant impact on the goals of the analysis. This in turn allows the Analyst to determine what is relevant and worth investigating further and what is not.

RQ3 has not been investigated yet, but the aims (as future work) to use the results from RQ1 and RQ2 to determine how to design an argumentation schema that would allow analytical outcomes to evolve over time with different levels of certainty and low commitment to using all of the outcomes. RQ1 relates to the element of *evidence* and RQ2 relates to the element of *relevance* as found in the law domain. If designed correctly, then it may be possible to evolve a low commitment argumentation schema into a rigour argumentation schema. This would allow Analysts to take advantage of the progress in argumentation construction from the law domain, but at the same time allow them to playfully explore the information and enhance their understanding without overcommitting to something that may be too rigid for their sense making needs.

ACKNOWLEDGMENTS

The research results reported here has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) through Project VALCRI, European Commission Grant Agreement N° FP7-IP-608142, awarded to B.L.William Wong, Middlesex University and partners. We are also very grateful for the enthusiasm of the police analysts in sharing with us their experiences that made this report possible. Celeste Groenewald has been awarded a VALCRI PhD studentship to complete her research at Middlesex University, London, UK.

REFERENCES

ACPO (2005). Investigative and evidential evaluation (extract from ACPO (2005) Core Investigative Doctrine). Available at: http://www.app.college.police.uk/app-content/investigations/linked-reference-material/

Allen, C., Taylor, C. and Nairns, J., 2015. Practical guide to evidence. Routledge.

Ashby, W. R. (1958). Requisite variety and its implicatons for the control of complex systems. Cybernetica, 1(2), 83-99.

Bex, F., Koppen, P., Prakken, H., Verheij, B., (2010). A hybrid formal theory of arguments, stories and criminal evidence. Artif Intell Law, doi:10.1007/s10506-010-9092-xToulmin, S. (2003). The Uses of Argument. Cambridge University Press, 2 edition

College of Policing (2009). Practice Advice on The Management of Priority and Volume Crime: The Volume Crime Management Model. Second Edition. NPIA. Available at: http://library.college.police.uk/docs/acpo/VCMM-191109.pdf

Corbin, J.M. and Strauss, A., (1990). Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative sociology, 13(1), pp.3-21.

Crandall, B., Klein, G., Hoffman, R. (2006). Working Minds – A Practitioner's Guide to Cognitive Task Analysis, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England

Groenewald, C., Wong, B.L.W, Attfield, S., Passmore, P., Kodagoda, N. (2017a). How Analysts Think: Navigating Uncertainty – Aspirations, Considerations and Strategies. In the 13th International Conference on Naturalistic Decision Making 2017, Bath, Uk

Groenewald, C., Wong, B.L.W, Attfield, S., Passmore, P., Kodagoda, N. (2017b). How Analysts Think: Navigating Uncertainty - Exploring the 'Lifecycles of Entities' in a sense-making environment. Unpublished HFES 2017 submission.

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting, 1, 576–580. Cited in Klein, G. (2008). “Naturalistic Decision Making”. HUMAN FACTORS, Vol. 50, No. 3, June 2008, pp. 456–460. DOI 10.1518/001872008X288385.

Klein G., Phillips J. K., Rall E. L., Peluso D. A. (2007). A Data-Frame Theory of Sensemaking. In Hoffman R. R. (Ed.), Expertise out of context Proceedings of the Sixth International Conference on NDM (pp. 113–155). Erlbaum.

Legislation.gov.uk (1997). Police Act 1997. Part III Authorisation of Action in Respect to Property. Section 93. Available at: http://www.legislation.gov.uk/ukpga/1997/50/section /93

Nissan, E., (2009). Legal evidence, police intelligence, crime analysis or detection, forensic testing, and argumentation: an overview of computer tools or techniques. International Journal of Law and Information Technology, 17(1), pp.1-82.

Pennington, N. and Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. Journal of Personality and Social Psychology, Volume 62(2), Pages 189-206

Pirolli, P. and Card, K. (2005). The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, Conference: Proceedings of International Conference on Intelligence Analysis, Available at: https://www.researchgate.net/publication /215439203_The_sensemaking_process_and_leverage_points_for_Analyst_technology_as_identified_throug h_cognitive_task_analysis

Takken, S., and Wong, B. L. W. (2013). Tactile Reasoning : Hands-on vs Hands-off – what’s the difference? In 11th international conference on naturalistic decision making NDM.

Toulmin, S. (2003). The Uses of Argument. Cambridge University Press, 2 edition

Wigmore, J. (1931) The Principles of Judicial Proof (2nd Edition) Little, Brown & Co.

Wong, B. L. W. (2014). How Analysts Think (?): Early Observations. In 2014 IEEE Joint Intelligence and Security Informatics Conference (pp. 296–299). IEEE

Training support for Naturalistic Decision Making: Serious gaming for adaptive performance of military personnel

Yelim MUN[a], Anja van der HULST[b], Esther OPRINS[b] and Jan Maarten SCHRAAGEN[a&b]

[a]*University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands*

[b]*The Netherlands Organisation for Applied Scientific Research TNO, PO Box 23, 3769 ZG Soesterberg, The Netherlands*

ABSTRACT

Introduction: For effective decision making in the 21st century where operational environments are complex and uncertain, there is a strong need for training support and its practical application to naturalistic, real-world settings. In this contribution, we focus on training of adaptive performance using NDM approach. In particular, we introduce serious games (SGs) as a potential means to train adaptive performance of military personnel. **Method**: The design of our SG is explained with the aim of enhancing adaptive performance and effective decision making in complex and uncertain environments. The rule change element is introduced and other game design elements are described taking the perspective of Naturalistic Decision Making (NDM) approach. **Results**: The findings from a game design validation session show that participants were able to assess the situational change and adapt their actions during the SG play. Challenges of practical application of the training support into the Dutch Major's school are described. **Discussion**: We discuss future directions of SG improvement within an NDM approach.

KEYWORDS

Decision making; adaptive performance; education and training; serious game; military.

INTRODUCTION

The Naturalistic Decision Making (NDM) approach has been frequently examined and applied in various domains, especially the domain of defense and security (i.e., Klein, Klein, Lande, Borders, & Whitacre, 2015). Effective and efficient decision making of commanding officers not only affects their own work performance but also the lives of soldiers and civilians. Incidents such as the USS Vincennes shooting down a commercial airliner in 1988, the Tarnak farms incident where American F-16 bombed Canadian soldiers in Afghanistan in 2002, and the attack of an Afghan hospital by the US military in 2015 show that decisions made under time pressure and with incomplete and often conflicting information can have adverse consequences. Therefore, many scholars have investigated decision-making in real work environments to help high-stake decision makers in various domains including the military. However, sound decision making in the 21st century involves several challenges because operational environments are becoming more complex, changing, uncertain and unpredictable. To effectively handle such environments, adaptive performance, defined as a capability to change and adapt according to situational demands (Good, 2014), should be enhanced. Despite the need for training professionals in decision-making under uncertain circumstances, there has been a scarcity of research into the development and application of such training support tailored to real working environments. This study aims to explore the game design of training support for ill-structured decision-making in uncertainty to enhance adaptive performance of future commanders from the perspective of an NDM approach.

Coping with uncertainty

According to Lipshitz and Strauss (1997), 'uncertainty' is summarized as *ambiguity*, *doubt*, and *unpredictability* causing difficulty to make sound decisions. The topic of uncertainty management has been studied in various domains such as education, cognitive psychology, and human factors using related constructs. For example, the term 'adaptive performance' refers to adaptability in real-time task performance under uncertain and dynamic circumstances (Good, 2014). In that sense, 'cognitive flexibility' is one of the strong predictors of adaptability, viewed as an adaptive reaction to changing and uncertain situations by rapidly reforming one's knowledge (Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987). Another related construct to uncertainty management, 'resilience', refers to human and system interaction under complexity and uncertainty (Bergström, van Winsen, & Henriqson, 2015). Finally, 'accelerated learning' shares similarity with uncertainty management in that the aim of accelerated learning is to enhance work proficiency dealing with complexities including uncertainty (Hoffman et al., 2014). Scholars have investigated uncertainty and its related constructs to understand and improve human performance despite the challenges of complexities and uncertainties. As earlier research of uncertainty management focused on acquiring more information before the moment of decision making to decrease the level of uncertainty (i.e., Lipshitz & Strauss, 1997), more studies nowadays emphasize the exposure and experience to decision making under such challenging operational environments to better deal with uncertainty (i.e., Carbonell, Stalmeijer, Könings, Segers, & Merriënboer, 2014) rather than ways to avoid or diminish uncertainty. Although many researchers examined how people make decisions under uncertain and complex situations, as Klein (2015) expressed his concerns, more studies should focus on training methods and their applications, which can aid effective decision making under uncertainties and complexities that are realistic and hold direct implications for the workforce. Therefore, we will introduce a design of a training support for adaptive decision making of military personnel.

Training support: Serious gaming

We introduce Serious Games (SGs) as a potential training means to support adaptive decision making of military personnel. SG refers to a class of games used for the purpose of learning and training beyond entertainment (Ritterfeld, Cody, & Vorderer, 2009). With technological advancement, numerous scholars have investigated SGs as an alternative learning tool for education and training. In particular, SGs have been frequently studied for workforce training in domains such as medicine (i.e., Graafland, Schraagen, & Schijven, 2012), business (i.e, Chang, Chen, Yang, & Chao, 2009), security and safety (i.e., Nesbitt et al., 2015). Also, SGs are widely used in dynamic decision making research (i.e., Gonzalez, 2004). For training of adaptive performance within the NDM perspective, SGs may be a valuable tool in that it offers a naturalistic environment where learning is situated in context (Gee, 2005). Considering that scholars have emphasized the instruction method of exposure and experience to operational environments that are challenging due to complexity and uncertainty (Spiro, Coulson, Feitovich, & Anderson, 1988), SGs are a practically suitable means to provide naturalistic virtual environments where dynamic, novel, and complex decision making can be practiced. Furthermore, using SGs as a training support for the military seems valid because of its time and cost effectiveness in comparison to traditional field training (Roman & Brown, 2008). Finally, SG is a familiar training means to military in that SGs are commonly used for military training for over 200 years.

Previous studies focused on training adaptive performance using SGs. Cañas et al. (2005) investigated the efficacy of strategy change during SG training, which encouraged dynamic decision making under changing environments. Glass et al. (2013) examined SGs—embedded within a rapid switching and dynamic virtual environment inducing complex decision making—as an intervention to increase adaptive performance. Gonzalez (2004) examined the relationship between time pressure, cognitive abilities, and dynamic decision making using SGs in uncertain and changing environments. Other studies (i.e., Chen, Thomas, & Wallace, 2005; Good, 2014; Marks, Zaccaro, & Mathieu, 2000; Stokes, Schneider, & Lyons, 2010) investigated SG support for adaptive performance embedding complex decision making tasks under changing and uncertain environments. These studies hold value in that they reported training outcomes to better aid adaptive performance and they used realistic decision-making tasks via SG support. However, direct application of these studies into real work training is somewhat limited in that all of these studies used university students as participants in controlled laboratory settings, hence taking a microcognitive rather than a macrocognitive perspective. Therefore, SG design within an NDM approach may further facilitate the practical applications of SG training support for the adaptive performance of military decision-making.

METHODS: GAME DESIGN

In this section, we will address the theoretical foundation of the SG and how the SG was designed to enhance adaptive performance via rule change element. We will first describe the macrocognition (Klein et al., 2003) and the recognition-primed decision (RPD) model (Klein, 1993). These concepts are theoretically embedded in the game design as their relevance to adaptive performance in uncertainty. Then, the specifics of the game design will be described.

Macrocognitive functions and processes

Macrocognition refers to research on functions of cognition and how these interact within realistic environments (Schraagen, Klein, & Hoffman, 2008), including Naturalistic Decision Making. The rationale of using macrocognitive functions to explain the game design is that the definition of macrocognition is closely related to adaptive performance in complex, naturalistic environments. Therefore, we will describe our game design as aligning with the description of macrocognition by Klein et al. (2003).

- Naturalistic decision making; complex decision; decision making under time pressure, high stakes, high risks: The game is designed for naturalistic decision making in that players (role of company commanders) have to make decisions in a military relevant context such as sending units to gather intelligence, to find civilians, order to capture enemy intel, etc. Players make complex decisions due to an abundance of information, lack of information, changing situations and unknown situations. Players are limited to take only two actions per case and the narrative requires players to take rapid actions within a particular time limit. Risks are high because units can be heavily damaged and cause a death toll depending on the decisions made.

- Planning-> sensemaking and situation assessment-> problem detection-> adaptation and re-planning:
  These are the attentional processes that we expect players to experience during the game play based on their prior working experience. After the briefing, each player plans how to tackle the problem. As players make a series of decisions via cases, sensemaking and situation assessment occur. For example, players gather situational cues such as the interaction of decoys with mounted enemy weapon. Then, players assess the situation based on the decoy that when approaching the pathway, the mounted weapons near the pathway detect and destroy all air and ground vehicles immediately. Sound decision based on the situational assessment will be sending the units on foot where the radar of mounted enemy weapon cannot reach. We embedded situational cues throughout the narrative in a way that the number of cues increases gradually for effective training, avoiding data overload. During the rule change phase, players detect problems in that how they previously made decisions no longer applies to the changed situation. In this phase, the feedback they receive contains results of actions following the new situations. For example, the mounted enemy weapon no longer detects and destroys the air vehicle due to the malfunction of its radar after the solar eruption. For course, the causation between the solar eruption and malfunction of enemy detection radar is not explicit. It is the role of players to assess the change of situations and conduct appropriate sensemaking. Therefore, in order to complete the mission successfully, players must adapt and re-plan according to the new situation.
- Vague key variables, goals and their interactions; uncertainty management:
  The narrative contains various unknowns such as no available intelligence on behaviour of enemy robots, missing information on the map, and the rule change. No players are told that rules change in the narrative. Players have to detect the changes based on given situational cues and feedback and adapt accordingly. Feedback is given only as a course of chosen actions instead of showing which options were the best. Players have no information about the long-term effects of decision making in each case. Therefore, the game is purposely designed such that players are exposed to complex decision making under uncertainty.

RPD model

In this section, we will examine the game design following the aspects of the RPD model, which is an aspect of NDM. The rationale of using RPD model is due to its strong relevance to the goal of game design and its training application for naturalistic decision making support.

- Four aspects (plausible goals, relevant cues, expectancies, actions) of situation assessment:
  As we mentioned in the general description of the game design, the selection of two decisions out of four actions is not designed to judge the options and select the best one. Instead, the focus is on decision-making based on situation assessment. First, players are informed about the goals of the mission (i.e., rescuing civilians, gathering intelligence on new robot army) and they are plausible in that players have resources such as units and advanced technology to achieve the goals. Second, relevant cues are embedded in case descriptions and feedback. Third, there are expectancies that players (trained military personnel) can complete the mission by successfully rescuing the civilians and defend the area against the enemy. Fourth, players take courses of action (i.e., sending the combat unit to a housing area to search for remaining civilians) based on their situational assessment.
- Three levels of decision making:
  During the rule learning and consolidation phase, the level of decision-making is designed to be simple and more complex as players proceed. The level is controlled by the amount of information processing for effective decision-making, amount of relevant cues and available information. The control of complexity and gradual presentation is critical for effective training, avoiding overload of complexity (Field, Rankin, Pal, Eriksson, & Wong, 2011). When players enter the rule change phase, the level of decision-making is at the most complex due to the unexpected and uncertain changes of the situation.
- Conditions of using RPD model:
  This game is designed to support decision making of military officers. We expect that the ideal users for this game have sufficient working experience in military service. The available data in the game is rather perceptual and intuitive in terms of relevant cues and feedback in that no statistical or numerical data is provided for affecting decision making. We emphasize the adaptive performance during the rule change phase but not on the justification of the decision making in general. In fact, the game does not provide feedback whether one decision is better than another. Also, the decision tasks are ill-defined and high level without structured, psychomotor action. Ambiguitiy is prevalent throughout the narratives and players have to deal with uncertain, unpredictable and complex situations. Due to the rule change, the situation is unstable.

We designed and developed a PC-based, complex decision-making SG, using the theoretical perspective of NDM approach as explained above, to increase the implication of the training tool for military training of adaptive performance. The game is scenario-based, designed for individual players (military personnel) with a rich narrative. Two scenarios are created following the game design framework (see Figure 1). For each scenario, players (decision makers) begin the game with the briefing, which contains the goal of the game, background information, and the current situation. Players complete the game by making 21 decisions (cases) following the descriptions of cases in each scenario. A case refers to an assignment for players. Maps were added in the game to help players visualize the location. Decision making in the first nine cases is the rule-learning phase, where players learn about the situation and three main rules about the virtual environment (i.e., behaviors of the robot enemy). Following each case description, players receive four different actions. Once players choose two actions, feedback is given automatically about the results of chosen actions (see Figure 2). For instance, a player (a company commander in scenario 1) can select two actions among a) Send a combat platoon unit to a nearby 5-story hospital building to search for civilians, b) send out armed scouts to a nearby water treatment facility to gather intelligence, c) send out armed scouts to a nearby warehouse to gather intelligence, and d) Send a combat platoon unit to a nearby bus station to search for civilians. Such selection is geared to encourage learning by doing and feeling of control as a player rather than weighing the best option. After the rule-learning phase, a second briefing is given with more information. Afterwards, players confirm the rules during the consolidation phase. This phase is created with the purpose of reassuring learning in case players did not understand the rules initially. Finally, an event occurs that changes the initial rules during the rule-change phase. For example, a solar eruption occurs changing the functions and behaviors of enemy robots in scenario 1. The players are not told that the rules are changed. They have to figure it out by themselves based on the decisions they make and their feedback.

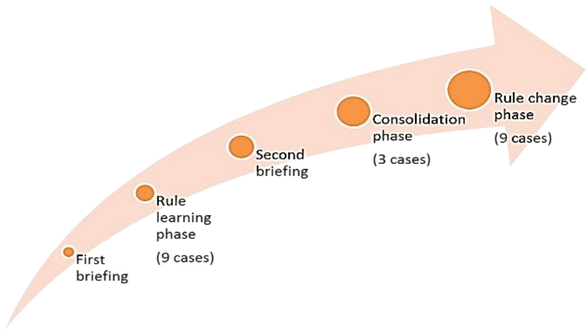


Figure 28. Game design framework

The key game design for supporting adaptive performance is the rule change phase. Rule change is our main game element that is designed to expose players to an adaptive performance-inducing environment. The rule change element is derived from tests, which measure cognitive flexibility (i.e., Wisconsin Card Sorting Test). Naturalistic and situated context is added to the rule change element, which is tailored for military decision-making. Our assumption is that by providing a naturalistic experience of rule change via a strong narrative, military officers can enhance adaptive performance leading to effective decision-making under complex and uncertain situations. The further details of game mechanics are available in the Mun et al. (2016) study.

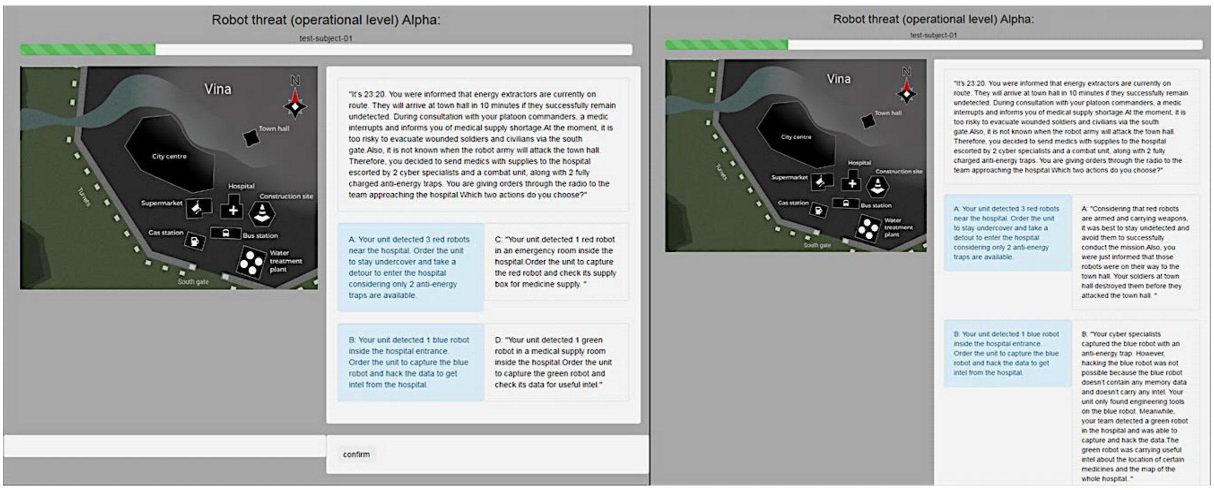


Figure 29. Action- and feedback-based game play

RESULTS

As our study is in work-in-progress status, we will share the brief findings regarding the practical application of the training support for enhancing adaptive performance. First, we tested the SG whether the discussed design of the SG induces expected behaviors of players such as detecting the rule change and adapt their planning for effective decision making. After the initial testing, we improved the game and used the game in the military training as a pilot study.

Validation of SG design

Before the validation, we improved the game based on the feedback from playtesting done by SG experts at TNO. Afterwards, we validated the game design during an introductory workshop for Game Master's students at the University of Amsterdam. The rationale of using student participants although taking an NDM approach is that these students have extensive experience and knowledge in games. Therefore, we assumed that these participants can provide us more insights on improving the game prior to its application in military training. Twelve Game Master's students (1 female, 11 male) participated in game testing. We introduced the purpose of the game testing and general instruction on how to play the game. At the end of the game play, students filled in a survey and informal group reflection was given. The total testing took about 90 minutes.

From the testing, we found that individual differences of players are strong in terms of time of game completion, and decision making patterns. Out of 12 participants, 11 participants reported that they were able to detect the rule change and acted accordingly. The moment of detection varied per player. The survey investigated how the players experienced the game. It consists of students' ratings on game difficulty, level of engagement, motivation, and concentration as well as open questions for suggestions. Figure 5 shows how the participants experienced the game. Both engagment and motivation was positively reported. Concentration and difficulty varied among the students. Some participants expressed during the group reflection that making decisions in military context was difficult due to their lack of prior knowledge. For concentration, some participants mentioned that they were distracted due to fatigue. In fact, this session was given at the end of the introductory workshop, which lasted 9 hours. Based on the suggestions and feedback from participants, we modified the game such as adding more visual aids and modifiying parts of narrative.

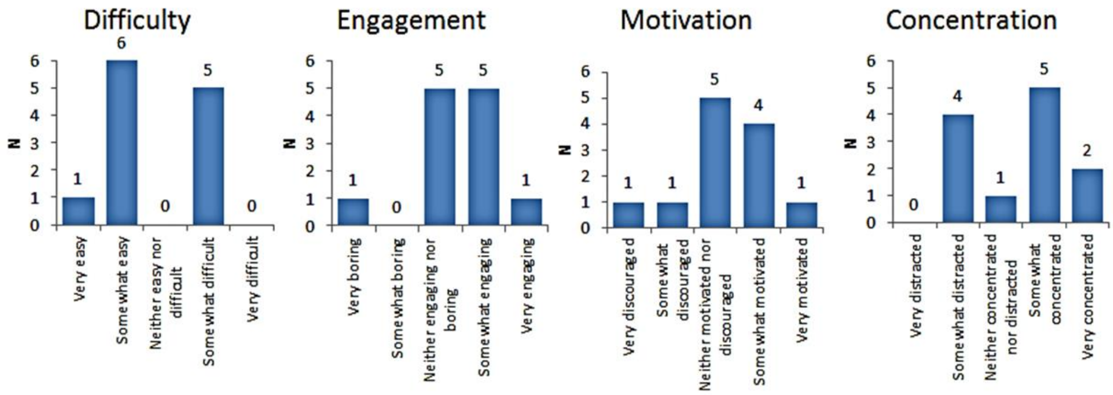


Figure 30. Students' assessment on the game testing

Application to military training

After the revision of the game, the game was used in the military training during the Dutch Major's course. Seventeen Major's course students (military officers) participated. The session began with the purpose of the training followed by the introduction to the game. The officers played the first scenario in the morning session and the second scenario in the afternoon session. Each game session contained individual and group reflection upon completion. In between the game play, participants were asked to fill in questionnaires and participate in non-game adaptability related tasks. From the notes taken during the reflection session, we conclude that the application of the game for adaptive performance was positively received by the participants. However, challenges of applications into a real military training setting occurred. Besides the technical challenges, some officers felt discomfort with the changes in narrative. Although many officers found the SG as an interesting training tool for adaptive performance, others had difficulties accepting the scenarios as the narrative with the rule change conflicts with their prior knowledge and experience. The data analysis is currently in progress.

DISCUSSION

This study explored the game design to support the training of adaptive performance in a naturalistic decision-making context for military personnel. We introduced rule change as a design element to induce adaptive performance. Then, we explained the game design using the macrocognitive functions and processes described by Klein et al. (2003). Moreover, we described the SG design in terms of the aspects of situation assessment as well as levels of decision making and conditions of the RPD model (Klein, 1993). Following the SG design, we reported findings from a game testing session and application of the SG in the Dutch Major's school training.

Although the game design contains various elements encouraging the abovementioned training aspects, it is important to actually check that players train adaptive performance via those game elements. For example, some players from the game workshop responded to the questionnaire in that they did not feel the time pressure although it was embedded heavily in the narrative. Considering that one of the NDM conditions is high time pressure, the addition of another game element for time pressure might be necessary. Another issue is the conflict between players' prior knowledge and the training context for adaptive performance. From the military training, we observed that some officers felt confusion and discomfort due to the rule change. This concept of training is new for military personnel and some confusion is inevitable as the officers have strong prior knowledge on the military context. Therefore, we should further investigate on the state of players' knowledge structure during the SG play and the effects of strong or weak prior knowledge on the military context. In terms of immersion, we received overall positive feedback from both military officers and game students that our SG training tool is engaging. However, it is difficult to check whether players use balanced intuition and analysis in their decision making as Klein (2008) emphasized. As the two pilot sessions were focused on the validation of the training concept and its application, rather than embedding RPD model to the testing, future research should pay specific attention to measuring macrocognitive functions to assess the effects of the SG using NDM approach. Moreover, how the RPD model cognitively takes place within the decision making during SG play such as mental simulation should be examined in detail. Also, more attention should be paid to other design elements that can train adaptive performance. Lastly, a more empirical study should be conducted to examine whether the developed SG can effectively train and improve adaptive performance of decision makers.

ACKNOWLEDGMENTS

We thank Ronald Jongen for developing the game into a PC version. Moreover, we thank Marjoleine 't Hart for her military expertise on improving the game narratives. We also thank Karel van den Bosch and Andrea Jetten for their support on the game design and development. We acknowledge the support of SG experts in TNO for their valuable feedback after play testing the game. We appreciate the participation of Game Master students at the University of Amsterdam and the participation of the Dutch Majors School. This research was partially funded by the Human and Organizational Adaptability program (V1520) at TNO.

REFERENCES

Bergström, J., van Winsen, R., & Henriqson, E. (2015). On the rationale of resilience in the domain of safety: A literature review. *Reliability Engineering & System Safety*, *141*, 131–141. http://doi.org/10.1016/j.ress.2015.03.008

Cañas, J., AntolÍ, A., Fajardo, I., & Salmerón, L. (2005). Cognitive inflexibility and the development and use of strategies for solving complex dynamic problems: Effects of different types of training. *Theoretical Issues in Ergonomics Science*, *6*(1), 95–108. http://doi.org/10.1080/14639220512331311599

Carbonell, K. B., Stalmeijer, R., Könings, K., Segers, M., & Merriënboer, J. van. (2014). How experts deal with novel situations: A review of adaptive expertise. *Educational Research Review*, *12*, 14–29. http://doi.org/10.1016/j.edurev.2014.03.001

Chang, Y. C., Chen, W. C., Yang, Y. N., & Chao, H. C. (2009). A flexible web-based simulation game for production and logistics management courses. *Simulation Modelling Practice and Theory*, *17*(7), 1241–1253. http://doi.org/10.1016/j.simpat.2009.04.009

Chen, G., Thomas, B., & Wallace, J. C. (2005). A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *The Journal of Applied Psychology*, *90*(5), 827–41. http://doi.org/10.1037/0021-9010.90.5.827

Field, J., Rankin, A., Pal, J. van der, Eriksson, H., & Wong, W. (2011). Variable uncertainty: Scenario design for training adaptive and flexible skills. In *Proceedings of the 29th Annual Conference of the European Association of Cognitive Ergonomics* (pp. 27–33). Rockstock, Germany: ACM. http://doi.org/10.1145/2074712.2074719

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning*, *2*(1), 5–16. http://doi.org/10.2304/elea.2005.2.1.5

Glass, B. D., Maddox, W. T., & Love, B. C. (2013). Real-time strategy game training: Emergence of a cognitive flexibility trait. *PLoS One*, *8*(8), 1–7. http://doi.org/10.1371/journal.pone.0070350

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, *46*(3), 449–460. http://doi.org/10.1518/hfes.46.3.449.50395

Good, D. (2014). Predicting real-time adaptive performance in a dynamic decision-making context. *Journal of Management & Organization*, *20*(6), 715–732. http://doi.org/10.1017/jmo.2014.54

Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *British Journal of Surgery*, *99*(10), 1322–1330. http://doi.org/10.1002/bjs.8819

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (Eds.). (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York, NY: Psychology Press.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456–460. http://doi.org/10.1518/001872008X288385

Klein, G. (2015). Reflections on applications of naturalistic decision making. *Journal of Occupational and Organizational Psychology*, *88*(2), 382–386. http://doi.org/10.1111/joop.12121

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A.Klein, J.Orasanu, R.Calderwood, & C.E.Zsambok (Eds.), *Decision making in action: models and methods* (pp. 138–147). Norwood, CT: Ablex Publishing Corporation.

Klein, G., Klein, H. A., Lande, B., Borders, J., & Whitacre, J. C. (2015). Police and military as good strangers. *Journal of Occupational and Organizational Psychology*, *88*(2), 231–250. http://doi.org/10.1111/joop.12110

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems*, *18*(3), 81–85. http://doi.org/10.1109/MIS.2003.1200735

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalitic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. http://doi.org/10.1006/obhd.1997.2679

Marks, M. a, Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *The Journal of Applied Psychology*, *85*(6), 971–986. http://doi.org/10.1037/0021-9010.85.6.971

Mun, Y., van der Hulst, A., Oprins, E., Jetten, A., van den Bosch, K., & Schraagen, J. M. (2016). Serious gaming design for adaptability training for military personnel. In *NATO Modelling and Simulation Group Symposium*. Bucharest, ROU.

Nesbitt, P., Kennedy, Q., Alt, J., Fricker, R., Whitaker, L., Yang, J. H., … Patton, S. (2015). *Understanding optimal decision-making in wargaming*. Monterey, CA. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA587848

Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.). (2009). *Serious games: Mechanisms and effects*. New York, NY: Routledge. http://doi.org/10.1007/s13398-014-0173-7.2

Roman, P. A., & Brown, D. (2008). Games–Just how serious are they. In *Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (pp. 1–11). Orlando, FL. Retrieved from http://bisimulations.com/sites/default/files/file%7B_%7Duploads/Games%7B_%7DHow%7B_%7DSerious%7B_%7DPaper.pdf

Schraagen, J. M., Klein, G., & Hoffman, R. R. (2008). The macrocognition framework of naturalistic decision making. In J. M. Schraagen, L. G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp. 3–25). Hampshire: Ashgate Publishing Limited.

Spiro, R. J., Coulson, R. L., Feitovich, P. J., & Anderson, D. K. (1988). *Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains*. Champaign, IL.

Spiro, R. J., Vispoel, W. P., Schmitz, J. G., Samarapungavan, A., & Boerger, A. E. (1987). Cognitive flexibility and transfer in complex content domains. *Executive Control Processes in Reading*, 177–199.

Stokes, C. K., Schneider, T. R., & Lyons, J. B. (2010). Adaptive performance: A criterion problem. *Team Performance Management: An International Journal*, *16*(3/4), 212–230.

An Exploration of the Action Decision Making of Coaches Working in a Football Association Girls' Centre of Excellence.

[a]David PEARS, [a]Toni O'DONOVAN, [a]Joanne HILL, [a]Andrew MITCHELL, [a]Jeffrey ALDOUS and [b]Stephen HARVEY

[a]*University of Bedfordshire*
[b]*West Virginia University*

ABSTRACT

The action decisions made by coaches working in a Football Association Girls' Centre of Excellence were examined from the perspective of The Naturalistic Decision Making Framework (NDM). Eight qualified and experienced coaches were observed in a naturalistic setting (training and matches) across a full football season and interviewed to explore decisions made around critical incidents. Initial findings suggest that NDM may provide an appropriate lens through which to view the process of action decisions made by coaches in this context. A greater understanding of coaches' decision making using the NDM paradigm can inform and influence future coach development programmes.

KEYWORDS

Decision Making; Education and Training; Sports Coaches; Soccer.

INTRODUCTION

Continued investigation into not only 'what' coaches do, but understanding the underlying knowledge structures and cognitive processes that underpin 'why' and 'how' they do it, could provide useful information for the development of coach education programs, and generate theory that is true to the complex realities of coaching (Smith and Cushion, 2006). Lyle (2010) suggests that decision making in particular is an integral part of the everyday practice of coaches and that it is an important element of coach expertise (see also Abraham et al, (2006); Nash and Collins (2006)). Apart from the most basic elements of coach behaviour that take place instinctively or in a reactive manner, such as saying hello when meeting a performer, most other functions can be viewed as 'action decisions' (Lyle, 2010, p.27).

The nature of action decision making can be differentiated in terms of time pressure, with three decision types offered in the literature: a) deliberative, b) non-deliberative and c) semi-deliberative. The term deliberative is used for decisions that were made by the coach where time allowed for deliberation. Non-deliberative decisions, on the other hand, are made with very little time available for deliberation. However, many coach decisions are made with some deliberation, but where there is a fairly immediate time pressure and might almost be seen as intuitive. Indeed, Launder (1993, p.2) described the coach as the 'master of the instantaneous response.' This type of decision might be described as semi-deliberative.

Klein and Weick (2000) categorised decisions using the alternative categories, rational, intuitive and experiential. Rational decisions are made by quantifying a range of options based on deliberative analysis. Intuitive decisions appear to be made speedily and without deliberation. Lastly, experiential decisions are made by recognising the issue at hand and using memory and experience to decide on a course of action. The decisions made by coaches on a minute by minute basis could be described as semi-deliberative or experiential. This type of decision is the subject of a branch of cognitive psychology called Naturalistic Decision Making (NDM).

NDM describes "how people actually make decisions in real world settings" (Klein (2008, p.456). It is based on research on practice where decisions are complex, time constrained, where the stakes may be high, where the goals are vague, where the landscape is constantly changing and not all of the information needed is available including medical practitioners and uniformed services (Klein, 2008). Putting to one side the high stakes element that was derived from studies of the military and fire fighters, coaching appears to match well with the criteria for NDM contexts.

Lyle (2010) suggests that effective decision making is best viewed through the NDM paradigm as it describes and explains how effective and accomplished decision makers make and take decisions. NDM is not a theory itself, but a framework for 'understanding how people make decisions in real-world contexts that are meaningful and familiar to them' (Lipshitz et al, 2001, p.332). Indeed, Lipshitz (1993) stated that nine different NDM models had been developed in parallel.

Bowes and Jones (2006, p.235), when describing coaching, speak of "the fluid nature of the activity, comprising endless dilemmas and decision making, requiring constant planning, observation, evaluation, and reaction…" and coaching has been described as 'messy' by Jones, Harris and Miles (2009, p.377). Although Lyle (2007) would agree that coaching may be complex, he would suggest that the messiness described by Bowes and Jones (2006) and Jones (2009) may be overstated and, in turn, understates the measure of control that expert coaches are able to exert using a range of tools that he lists as 'coping strategies, routines, allowing degrees of freedom in expectations and targets, by the use of thresholds and key markers to decide what to attend to, and by using a set of performance plans to maintain a focus on the main goals' (p.124).

Rather than had been previously thought, these decisions are not made after careful deliberation of many options that are compared and the best option sought. Rather people appear to select a single action and consider if it will work. This seems to be determined by reliance on a kind of synthesis of their knowledge. It might be described as 'satisficing' (Simon, 1957). That is, using the first workable option rather than seeking the optimum solution. However, these solutions might not be apparent to novice or even intermediate coaches. Therefore, experience must be gained and internalised.

Research into the use of NDM in sports coaching is limited; however, Lyle (1999) describes an unpublished study of coach decision making that he conducted in the field of competitive volleyball. This study suggested that there might be a predominant cognitive mechanism involved in coach decision making. This mechanism was named the 'Slow Interactive Script Model'. He makes the suggestion that coaches continually refine action against 'scripts' or an imagined model of the situation at hand. They consider the situation, its determinants and likely future outcomes based on their prior experiences that appear to have formed a mental model against which they judge situations and determine action. These pre-formed mental models help coaches to recognise situations and choose appropriate actions. It is important to state though, that these scripts are mediated by the current context or circumstances too. More recently, Lyle, Harvey and Muir (2015) examined NDM as a framework for examining sports coaches' decision making. Again, the Slow Interactive Script Model was observed along with the notion that coaches use 'key attractors' to provide focus for their attention and that subsequent breaches of thresholds relating to these key attractors provide a trigger for the management of decision making.

The purpose of this study is to contribute to an emerging body of evidence on sports coaches' behaviour and practice; in particular, to complement descriptive evidence of the cognitive complexity of soccer coaches' behaviour with insights into the cognitive organisation and strategies accompanying the coaches' action decisions. More specifically, the project investigated the in-practice and in-game decision making of coaches working in a football centre of excellence for girls. A mixed-method approach comprising systematic observation, coach interviews and critical incident analysis was used to provide a more holistic examination of the pedagogical and cognitive strategies used by these coaching practitioners (Potrac, Jones and Armour, 2002) – essentially, considering 'why' and 'how' coaches make decisions during the 'act of coaching intervention'. The research provides an opportunity to bridge systematic observation and interpretive accounts of practice to inform the development of intervention strategies and an exploratory model of coaching pedagogy. Essentially, the purpose is to examine the extent to which the NDM framework can 'explain' coach action decision making.

METHODOLOGY

The study utilised an in-depth case study within the interpretive paradigm (Bloom et al, 1999) using a variety of methods to describe and analyse coach decision making in the particular context of a Football Association (FA) Girls' Centre of Excellence. It examined the in situ or covert elements of decision making (rules of thumb, emergent stories or scripts, thresholds etc.). In so doing, it attempted to provide an insight into the practice in this centre and potentially provide a framework for further study in this and other coaching contexts.

Participants and Context

The participants were coaches (n=8) working in a FA Girls' Centre of Excellence (GCoE). The coaches mean age was 26.50 years (SD=3.02) and each had 10.75 Years of experience on average (SD=3.96). One coach held UEFA A licence, three held UEFA B licence and four held the UEFA C Licence. Each age group (under 11, under 13, under 15 and under 17) is coached by two coaches, one a head coach, one an assistant head coach.

In total 36 training sessions and 12 matches were observed. Each participant coach was observed in training on either 4 or 5 occasions (M=4.5, SD=0.54). In matches, only the head coaches (n=4) were observed and for three matches each.

The FA's website states that the **mission statement** for girls' talent development is "To allow players opportunities to access appropriate levels of coaching and support throughout the talent pathway, enabling the programme to reach our ultimate goal in producing elite English female players to compete on the world stage."

This population is of particular interest as no research concerning GCoE currently exists. At the time the field research was conducted there were 31 GCoE currently operating under licence from the FA across England and they are the similar to the academies/centres of excellence that operate for male players. Most GCoE are part of clubs that also participate in the men's professional game, for example, Arsenal, Manchester United and Norwich City, but other types of organisations such as Devon Football Association and Teesside University run a small number of centres. This differs from the centres for male players that are exclusively operated as part of professional clubs such as those mentioned above.

While research into academies/centres of excellence in the male game is growing exponentially, (Brown and Potrac, 2009; Ford et al., 2010; Ford & Williams, 2011; Jones, Armour and Potrac, 2003; Partington, Cushion & Harvey, 2014; Potrac, Jones and Armour, 2002; Smith and Cushion, 2006), there appear to be no studies that have taken place in a GCoE and therefore none concerning the coaches that operate there.

Procedure

Prior consent was sought from the participants and a full explanation given as to the purpose of the interviews and the study in general. Assent was also gained from the players and their parents. Although the players were not participants they were the subject of the participants' decisions and each match and training session was also video recorded while the coaches wore a radio microphone to enable the researcher to listen to the coach interact during the coaching episodes while being able to do so from a distance.

Pre and Post coaching episode interview

Immediately before and after each coaching episode (training sessions and matches) the coaches were guided through an interview. A semi-structured interview format was used as it "*allows the interviewer to ask specific questions while maintaining flexibility with regard to how the interview develops as a result of exchanges with the respondent.*" (Breakwell in Breakwell, Hammond & Fife-Shaw (Eds.) (1995).).

The pre coaching episode interview allowed the researcher to have access to the deliberative thinking that precedes the coaching episode (match/training). This will provide context for the post coaching episode interview. For example, were the session aims met? This can only be judged against the stated aims. The post coaching episode interview will be used to highlight critical incidents from the episode that might be examined through further questioning. The interviews were transcribed verbatim by a professional transcriber and checked by the lead researcher for accuracy.

Interview Coding and Analysis

The interviews were coded using NVivo version 11. Three primary areas of interest were identified from the literature; 1) Situational Assessment (Scanning and Interpretation), 2) Triggering and 3) Management of the Decision Process. Further secondary and tertiary level areas were identified under these main headings. These will be discussed in greater detail in the discussion section.

A general inductive approach as outlined by Thomas (2003) was used. Thomas (2003, p.238) suggests that, "*many researchers and evaluators, who are unfamiliar with any of the traditional approaches to qualitative analysis, wish to have a straightforward set of procedures to follow without having to learn the underlying philosophy and technical language associated with many qualitative analysis approaches."* Inductive analysis seeks to make sense of the situation being studied through observation and subsequent analysis of emergent themes. "*Inductive analysis begins with specific observations and builds toward general patterns. Categories or dimensions of analysis emerge from open-ended observations as the evaluator comes to understand program patterns that exist in the empirical world under study."* Patton (1990, p44).

For the purpose of the content analysis, the basic unit of analysis was operationally defined as any comment made regarding the participant's action decision making. Each response was then examined and statement types within the response categorised by the lead author in relation to features of NDM (i.e. Situation Assessment (primary) – Key attractors (secondary) – Coach Identfies Key Attractors (tertiary)).

RESULTS AND DISCUSSION

Initial analysis highlights the breadth of factors that impact on coaches' decision making in game and training contexts. It was particularly apparent that medium-long term factors influenced their decisions in many situations alongside the immediate match play decisions that were necessary to take. In this section I present two examples from the initial data analysis that exemplify the process of decision making for the coaches in this study

Example One (Match)

The incident below outlines where a decision was made during a match where the team were not playing well against strong opposition and were losing.

> "…but then because they weren't getting enough of the ball, they couldn't do their in-possession targets as a unit and then when they weren't getting the out-of-possession, when they were just losing the ball in transition, they weren't quick enough to do it out of possession…"
>
> Coach E (Head Coach, Under 15)

This example provides evidence that Coach E conducted a situational assessment and that a threat was interpreted. This was based on his expert opinion of what he saw happening on the pitch. It is informed by his expectations of what performance should look like.

However, the threat was not the score or losing the match as might be expected in elite sport but that the players were not able to meet their training targets and that the outcome of the match might affect their self-confidence. The threat to the long term aims of the centre acted as a trigger to make a decision. The coach decided to create new targets that were achievable under the circumstances. This would protect the long term aims.

> "…so it's almost a case of taking it out of the way, let's put that to one side and let's give them targets where it's collectively they can get through the game, as less psychologically scarred as possible."
>
> Coach E (Head Coach, Under 15).

Example 2 (Training)

Coach A (Head Coach, Under 11) pulls Molly (pseudonym) a 10 year old player to one side after Molly does not receive the ball after calling for it from Carrie (pseudonym) an 11 year old player. The coach's responses highlighted that the following cognitions contributed to her action decision;

There were a number of triggers that prompted the decision. Molly was not doing something that was part of a specific action plan for her.

> "…what we're trying to work on at the minute with Molly is her thinking which the other player is at that level, so what I was trying to get across to her was, erm, why is, why is Carrie making that decision and you're not."

The coach could see Molly was frustrated. Therefore, the need to act and the timing of the act were considered.

> "…because I could see her getting wound up that she wasn't getting the ball…"

The player's frustration acted as the trigger. Subsequently, a rule of thumb was employed. Coach A explained that she would not have intervened in the same way if the player was 9 years old as she would not have expected the player to be as capable. It is more likely that she would have praised the effort.

In addition to the prompts or triggers the nature of the decision to speak to Molly away from the group was thought through. It was informed by her knowledge of Molly and her lack of ability to deal with criticism in front of others due to self-confidence issues.

> "The reason why I didn't stop the group and speak to them all is because Molly is struggling right now with confidence, erm, for some reason she doesn't respond very well if we stop the full group and make a point about her"

CONCLUSION

The initial findings suggest that the assertion made by Lyle (2010) and Harvey, Lyle and Muir (2015) that NDM may provide an appropriate lens through which to view the process of action decisions made by coaches may be correct. The coaches appear to use their experience and understanding of what is expected of players at this level to assess their performance. They are able to recognise and interpret these situations. There is strong evidence that they utilise trigger/thresholds that lead to action. These actions are often managed by the use of heuristics. Scanning, interpretation, triggering and decision management appear to be nested within or mediated by the long term aims of the centre. These aims are primarily concerned with the long term development of the player and not with results in matches.

This clearly has implications for coach development. For example, Readinger et al. (2005) suggested that Decision Skills Training might be used to develop the decision making skills of coaches. Lyle (2010) compiled a comprehensive list of suggestions as to how the NDM paradigm might be used to view coaches' action decision making and inform coach development. They are too numerous to list here, but include, adapting the language of NDM to coaching, situated learning, problem based learning, directed and purposeful reflection, using case studies and scenarios and the development of 'rules of practice'(p. 38). Therefore, further purposeful study of coach decision making from NDM perspective offers great potential for coaches and academics alike to better understand coach decision making and to assist in the development of expert coaches.

REFERENCES

Abraham, A., Collins, D., & Martindale, R. (2006). The Coaching Schematic: Validation through expert coach consensus’. *Journal of Sport Sciences, 24* (6), 49-564.

Bowes, I., & Jones, R.L. (2006). Working at the Edge of Chaos: Understanding coaching as a complex interpersonal system. *The Sport Psychologist, 20*(2), 235-245.

Bloom, G.A., Crompton, R., & Anderson, J.E. (1999). A systematic observation study of the teaching behaviours of an expert basketball coach. *The Sport Psychologist, 13,*157- 170.

Breakwell, G.M. (1995). Interviewing. In G.M. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology*. London: Sage

Brown, G. & Potrac, P. (2009). 'You've not made the grade, son': de-selection and identity disruption in elite level youth football. *Soccer and Society, 10*(2), 143-159.

Ford, P.R., Yates, I. & Williams, A.M. (2010). An analysis of practice activities and instructional behaviours used by youth soccer coaches during practice: Exploring the link between science and application. *Journal of Sports Sciences, 28*(5), 483-495.

Ford, PR. & Williams, A.M. (2011). The developmental activities engaged in by elite youth soccer players who progressed to professional status compared to those who did not. Psychology of Sport and Exercise, 13, 349–352.

Harvey, S., Lyle, J.W.B., and Muir, B. (2015). Naturalistic Decision Making in High Performance Team Sport Coaching. *International Sport Coaching Journal, 2,* 152-168.

Jones, R.L., Armour, K.M. & Potrac, P. (2003). Constructing expert knowledge: A case study of a top level professional soccer coach. *Sport, Education and Society, 8*(2), 213-229.

Jones, R.L., Harris, R., & Miles, A. (2009). Mentoring in sports coaching: A review of literature, *Physical Education and Sport Pedagogy, 14*(3), 267-284.

Klein, G.A. (2008). Naturalistic Decision Making. *Human Factors 50*(3), 456–460

Klein, G.A., & Weick, K.E. (2000). Making better decisions. *Across the Board, 37*(6), 16-22.

Launder, A. (1993). Coach education for the twenty first century, *Sports Coach, 16*(1), 2.

Lipshitz, R.(1993). Converging themes in the study of decision making in realistic settings. In G.A. Klein,

J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision Making in Action: Models and*

Methods (103-137). Norwood, CT: Ablex.

Lipshitz, R., Klein, G.A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioural Decision Making, 14*(5), 331-352.

Lyle, J. (1999). Coaches’ decision making. In N. Cross & J. Lyle (Eds.). *The Coaching Process: Principles and Practice for Sport* (210-232). Oxford: Butterworth Heinemann.

Lyle, J. (2007). Modelling the complexity of the coaching process: a commentary. *International Journal of Sport Science and Coaching, 2*(4), 407-409.

Lyle, J. (2010). Coaches’ decision making: a Naturalistic Decision Making analysis. In J. Lyle, J. C.J. Cushion (Eds.). *Sports Coaching: Professionalisation and Practice* (27-42). London: Elsevier.

Nash, C. & Collins, D. (2006). Tacit Knowledge in expert Coaching: Science or Art? *Quest, 58*(4), 465- 477.

Partington, M., Cushion, C.J., & Harvey, S. (2014). An investigation of the effect of athletes’ age on the coaching behaviours of professional top-level youth soccer coaches. *Journal of Sport Sciences 32*(5),403-414

Potrac, P., Jones, R.L., & Armour, K.M. (2002). ‘It’s all about getting respect’: The coaching behaviours of an expert English soccer coach. *Sport Education and Society, 7*(2), 183 – 202.

Readinger, W.O., Hutton, R.J.B., Klinger, D.W. (2005). Recognition-primed decision making in sport: Improving individual, team and coaching decision making. In P. Morris, P. Terry, & S. Gordon (Eds.). Proceedings of the 11th World Congress of Sport Psychology. International Society of Sport Psychology, Sydney, Australia.

Simon, H.A. (1957). *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* New York: John Wiley and Sons

Smith, M., & Cushion, C.J. (2006). An investigation of the in-game behaviours of professional, top-level youth soccer coaches. *Journal of Sports Sciences, 24*(4), 355–366.

Thomas, R. (2003) A general inductive approach for qualitative data analysis. *American Journal of Evaluation 27*(2), 237-246.

Surgeons' Non-Technical Skills Rapidly Acquired by Trainees–Can Human Factors and Virtual Reality Create Synergistic and Predictive Effects?

Matthew PEARS[a*], David PEEBLES[a], Paul WARD[a], Yeshwanth PULIJALA[a], Ashraf AYOUB[b], Gurjog BAGRI[a], Minhua MA[a]

[a]*University of Huddersfield, UK*
[b]*University of Glasgow, UK*

ABSTRACT

There are difficulties in acquiring surgical skills through implicit and indirect teaching methods. These skills include decision-making, situational awareness, teamwork and communication, and leadership. Industries exploiting Human Factors methodologies are imperative to developing training methods capable of improving skill acquisition. This project aimed to create a virtual reality Non-Technical Skills (NTS) trainer for a complex Oral and Maxillofacial surgical procedure. The aim was to investigate how and why experts implement advanced cognitions and behaviours, how cognitions differ from trainees, and how differences can be taught explicitly. After task analysis of the procedure, cognitive interviews with ten surgeons gained approximately 300 minutes of various knowledge categories surrounding NTS expertise. With PhD collaboration, a training tool currently in development was theorised to significantly increase NTS and technical skills (i.e. tools/procedure knowledge). Evidence for the necessity of impactful NTS curricula training, using modern technology and Human Factors utilisation was the desired outcome.

KEYWORDS

Situational Awareness; Education and Training; Cognitive Task Analysis; Non-Technical Skills; Surgery

INTRODUCTION

Difficulties in Achieving Expertise

In the United Kingdom (UK), admissions for surgical procedures have been increasing at a substantial annual rate. The Royal College of Surgeons (rcseng.ac.uk, 2016) reported a 27% increase in the number of admissions for surgical procedures over 10 years from 2003. There were 4.7 million admissions in 2013/2014 with almost one third of these for surgical procedures. However, there has not been an increase in surgical trainee numbers. If only a small reduction in trainee numbers, over time this may amount to a significant negative impact if assumed an approximately linear decline. For example, in Japan Mizuno et al (2014) reported a steady decrease of surgeons between the years of 1996-2006 accumulating to a substantial 14.8% loss. This has suggested that in its current state a growing surgeon-to-procedure ratio may form, if procedures increase disproportionately to trainee numbers. Decreased OR training time constraints are also placed upon trainees. This has prevented opportunities for skill development (Alaker, Wynn, & Arulampalam, 2016; Blackmur, Clement, Brady, & Oliver, 2012). Consequentially, during the limited hours of training with conventional methods, in-depth practice with varied cases and unexpected complications have likely not occurred. This has meant that surgical trainees may not have sufficient practice to achieve the subsequent generally cited career analysis benchmarks of approximately 10,000 deliberate training hours to achieve expertise (Schaverien, 2009; Chrispen & Hoffman, 2016).

Difficulties in Communicating Expertise

To compound this issue, when a situation requires a problem to be addressed or an action to be chosen, an expert's attention, information gathering, and skills combine information significantly different to trainees (Vannaprathip et al., 2016). Consequently, this can create 'automated cognition' (Sullivan, Yates, Inaba, Lam, & Clark, 2014) and cause related behavioural outputs. As expert surgeons rely on increased automatic decision–making models when performing surgery, they may not be able to appropriately deconstruct their cognitive processes into steps (Moulton, Regehr, Lingars, Merritt, & MacRae, 2010). Thus, when trainees receive OR training, their instructors may find it difficult to provide reason behind approximately 70% of their decisions during critical procedures (Pugh, Santacaterina, Da Rosa, & Clark, 2010).

Surgical Training via Simulation

Upon identifying the discussed limitations in current surgical courses, compensatory training outside of the OR has developed to become a fundamental prerequisite within most curricula. As Aim, Lonjon, Hannouche, and Nizard (2016) stated many global surgical societies, such as the Accreditation Council for Graduate Medical Education, American College of Surgeons, and French National Authority for Health, have highly recommended training outside of the OR. Indeed, one approved surgical training method is simulation. Although antithetical in terms of both ecological and construct validity, simulation beneficially provides the ability to break down a system into its sub-components and train/assess individuals on each component. Importantly, as trainees can experience greater learning curves due to the complexities in modern OR environments (Banerjee & Gavade, 2015), simulations have been suggested to improve students' initial learning curve management relative to surgical skill development, with transferability to the OR (Aydin, Ahmed, Khan, & Dasgupta, 2015).

Limitations of Current Simulation

However, current simulations being comprised mainly of high-fidelity systems, have many limitations such as difficulty of implementation, high initial cost, and are usually highly demanding (Stephanidis et al., 2015). A further limitation has more recently being highlighted from a Human Factors (HF; Salvendy, 2012) perspective. This limitation surrounded the general assumption within surgical simulation literature that teaching and assessing only technical skills can predict most abilities of the individual (Graafland, Schraagen, Boermeester, Bemelman, & Schijen, 2015). Therefore, simulated training in many existing medical and surgical educational syllabi and curricula have oriented around practice of technical skill. Yet, technical skills themselves are not sufficient to preserve patient safety, singularly prevent, or reduce errors (Flin et al, 2007). There are other skills required from the surgeon in addition to technical skills- appropriately named non-technical skills (NTS; Yule, Flin, Paterson-Brown, & Maran, 2006).

Importance of Non-Technical Skills Training

These NTS are derived from understanding the successful implementation of HF in other domains. Such domains include Civil aviation (Flin et al, 2003), oil exploration (Flin, O'Connor, & Mearns, 2002; Robb & Miller, 2012), anaesthesia (Flether, McGeorge, Flin, Glavin, Maran, 2002), and nuclear power, and have prompted the production of NTS training to be implemented to the same strength in surgery (Agha & Fowler, 2015). Based upon the classification system by Flin, Youngson, and Yule (2015) four major categories have been revealed and consist of Decision-Making, Situational Awareness, Communication and Teamwork, and Leadership. NTS involve cognitive and social abilities indirectly related to an individual's clinical knowledge, dexterity, and use of equipment, yet are arguably as necessary for the high demand of the surgical workplace. Identified as requirements to achieve surgical excellence and competence (Yule, Flin, Paterson- Brown, & Maran, 2006) NTS skills accompany appropriate levels of surgical knowledge and technical skill. The use of simulation to teach NTS has been challenging due to the nature of non-technical constructs, yet has been gradually increasing. For example, although under-researched there has been a rise in virtual simulation to create surgical decision-making training tools (Flowers & Aggarwal, 2014; Lin, Park, Liebert, & Lau, 2015; Sarker, Rehman, Ladwa, Chang, & Vincent, 2009; Servais et al., 2006; Tran et al., 2013; Lin, Park, Liebert, & Lau, 2015).
Current Non-Technical and Human Factors Training

Despite this increase, the combined limitations of both NTS and simulations withhold the rate of implementation into surgical and medical education. Yet, when removing simulation from the equation, a plethora of HF and NTS training has been incorporated into various courses worldwide. The World Health Organisation's (WHO) patient safety course has a module specifically detailing HF for students to understand their role in patient safety (who.com, 2015). There has also been an increase in the development and rollout of NTS skills training, and primarily orients around using the NTS for Surgeons framework (NOTSS; see Flin, Youngson, & Yule, 2015). For example, the Royal Australasian College of Surgeons (RACS) has provided monthly NOTSS courses (Surgeons.org, 2016). In addition, the Royal College of Surgeons of Edinburgh (RCSE) Patient Safety Board has taught NOTSS classes directed towards senior trainees and consultant surgeons, with junior surgical trainees similarly trained via e-learning NOTSS modules (Flin, Youngson, & Yule, 2015; see rcseng.ac.uk, 2016).

Future Compulsory Non-Technical Skills Training

Providing an example of the range of NTS implementation, the specific surgical speciality of Oral and Maxillofacial Surgery's (OMFS) Intercollegiate Surgical Curriculum (2016) programme syllabus also initialised the inclusion of NTS training. The programme consists of an overview into the common components of the core surgical training programme. These include early year training in OMFS, along with intermediate, and final stage entry requirements. The selection criteria necessary to be accepted for intermediate and final stage training modules, as well as specialised areas, have included criteria of 'Personal Skills'. These 'Personal Skills' are broken down into the similar components of Judgement under Pressure, Communication Skills, Problem Solving, Situational Awareness, Decision-Making, Leadership and Teamwork, and Organisation and Planning. As of 2015, these non-technical clinical skills are also becoming compulsory training requirements in similar areas. For example, the NHS Health Education program (hee.nhs.uk, 2016) have mandatory HF training for Core Surgical Trainees (westmidlandsdeanery.nhs.uk, 2017).

Delivery of Non-Technical Skills Training

These proceedings highlighted the importance of NTS and the roll-out of HF training worldwide, and made it clear that a framework for NTS teaching was necessary. Yet, a formal framework on how to teach NTS has yet to be implemented (Bourrier, Jambon, Garbay, & Luengo, 2016). Furthermore, although the issues found within high-fidelity simulations may be reduced or solved, the Association for Simulated Practice in Healthcare (ASPiH, 2016) have promoted the implementation of the full spectrum range of simulation usage, from high-fidelity mannequins to low-fidelity VR simulation equipment. This additionally created a lack of clarity for the role of simulation and its various delivery methods and fidelities for NTS acquisition. Fortunately, Yiasemidou, Glassman, Tomlinson, and Gough (2017) helped to clarify this issue as they indicated that 86.67% of surgeons agreed that simulation could be used for NTS acquirement. More importantly they suggested that similar educational effects from contemporary low-fidelity simulators can occur when compared to high-fidelity systems. This suggested NTS and low-fidelity simulation has growing support for a place within future surgical training methods.

Surgical Training via Virtual Reality Simulation

Currently however, research literature has yet to acknowledge the efficacy of a modern, low-fidelity VR Head Mounted Display (HMD) tool when used for surgical skill acquisition. The capabilities of modern HMD's have excelled far greater than previous stages, with results from existing literature showing great promise (Aiken & Berry, 2015; Felnhofer et al, 2015; Gahm, 2015; Koo et al, 2015; Nori et al, 2015; Olmos, 2015; Shaw et al, 2015; Shiban, Reichenberger, Neumann, & Mulberger, 2015). HMD's being used in academic research, such as the Oculus Rift (Oculus.com, 2017), have been shown to induce a significant increase in a user's sense of Presence (Waterworth, Waterworth, Riva, & Mantovani, 2015; for Presence see Cummings & Bailenson, 2015). This was compared to conventional display technology within a range of virtual activities. HMD's additionally have a strong degree of sensory richness and like reality, the user is the source of change.

AIMS OF PROJECT

Based upon the literature reviewed, the OMFS department was chosen to be the focus of training as Coffey-Zern, Calvi, Vorrasi, and Meara (2015) stated the use of simulation has been scant when compared to general surgery and anaesthesia residency education. Additionally, collaboration with a PhD project (see Pulijala et al, 2017) which also utilised VR to train OMFS trainees' declarative and procedural knowledge (i.e. tools/procedure/anatomy), material will combine to create a novel 'all-inclusive' system aimed at teaching both technical and non-technical skills.

Consequently, from a Human Factors perspective the overarching aim of this project was to increase patient safety in the OR. Broken down, this meant the specific aim was to create a NTS training tool which would significantly increase trainees' NTS proficiency with transferability to the OR, as measured by the Situation Awareness Global Assessment Technique (SAGAT; *see* Endsley, 1995). The underlying principals behind this were that users could increase their NTS which would assist them in improving their cognitions and behaviours when performing a complex Le Fort 1 Osteotomy. From an educational and psychological perspective, a specific aim was to uncover the efficacy of NTS skill training within a low-fidelity VR headset. The overarching goal was to highlight the importance of synergy between contemporary technology and Human Factors as a potential primary tool for the deliverance of NTS training.

METHOD

Participants

Participant Criteria

Subject Matter Experts (SME's) were included to help in the knowledge elicitation process. Training methods can be significantly more effective when developed through the accurate identification of SME's cognitive processes (Merrill, 2002; Schaafstal, Schraagen, & van Belro, 2000; Velmahos et al, 2004). SME's may not be able to deconstruct their full cognitive processes into clear steps (Moulton, Regehr, Lingars, Merritt, and MacRae, 2010). Therefore, two or more SME's for Cognitive Task Analysis (see *Method section 2.2.2*) are necessary due to the negative properties of automaticity possessed by experts that can limit the effectiveness of their explanations. By including two or more SMEs, the data gathered from each interview can be combined.

Selected participants

Nine surgeons located across India were contacted and agreed to participate in individual semi-structured interviews. Surgeons were chosen from leading Indian institutions such as Manipal University-Manipal, Nair Hospital Dental College-Mumbai, SDM college of Dental Sciences-Dharwad, Manipal College of Dental Sciences-Mangalore, and KLE Institute of Dental Science-Belgaum. These participants were primarily selected due to the increased availability of consultants and Head of Departments when compared to the U.K. As previously stated, the aims of two separate PhD projects were collaborated to enable four authors of the current project to travel to the mentioned institutions over a 3-week period, with one of the collaborated aims being to interview participants with arranged appointments. As institutional location, both regionally and internationally, may implement differing surgical practice and procedures, an additional participant was from Glasgow Dental Hospital- U.K and aided in the standardization of the final content of a Le Fort 1 Osteotomy Task Analysis (TA; see Method section 2.2.1) and training material to conform with typical UK procedure. Opportunistically, 3 novice OMFS post-graduate trainees within three of the listed India institutes were also cognitively interviewed to highlight expert-trainees NTS skill level differences.

Based upon the three classes of methods for expertise criteria of career analysis, sociometric analysis and performance analysis (Chrispen & Hoffman, 2016), five out of ten participants met the criteria of 10,000 training hours/10+ years, hierarchical social/team positioning, and lower comparative error rates. Five participants did not meet the criteria of one and/or two classes as three participants had experience of five to ten years and were not at the top of the sociometric ranking, and it was not established the success rates of these five participants.
Apparatus and Materials

Task Analysis of Surgical Procedure

A TA (Sheperd & Stammers, 2005) of the Le Fort 1 osteotomy was performed before interviewing participants to comprehensively, systematically, and analytically document how the task should be accomplished. To the researcher's knowledge a TA for a Le Fort 1 Osteotomy had not been created. Therefore, a TA based upon previous structures was accomplished (Chipman, Schraagen, & Shalin, 2000). Referencing the tasks and sub-tasks found within the TA, cognitive interview questions were formed and oriented around the three levels of Situational Awareness (SA; Endsley, 1995) being perception, comprehension, and projection. This provided SA information that could be later assessed using the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995) when evaluating the completed VR training tool. Other questions surrounded decision-making at key points within the osteotomy and questions pertaining to error prevention/management, and communication/teamwork/leadership were also included. Therefore, a custom-make questionnaire was created for semi-structured cognitive interviews.

Cognitive Interview and Cognitive Task Analysis

A cognitive interview is a knowledge elicitation method within Cognitive Task Analysis (CTA; Chipman et al, 2000). CTA was selected as the most appropriate knowledge elicitation technique to uncover the cognitive differences between experts and trainees in terms of their NTS. It can reveal the different underpinning cognitions in subject matter experts (SME) and trainees (Flin, Youngson, & Yule, 2015) and each participant's cognitions can be mapped out to provide an in-depth and insightful representative of their NTS. It has enabled surgical experts to articulate operative steps and cognitive decisions that are covert in nature, yet linked with overt task complexities (Smink et al, 2012). CTA has successfully been implemented in many other complex fields to exploit the skills possessed by experts and provide a way for trainees to acquire them. From this, learning and retention of technical skills in surgical procedures can be improved (Campbell et al, 2011; Luker, Sullivan, Peyre et al, 2008; Smink et al, 2012; Sullivan, Brown, Peyre, et al, 2007; Velmahos, Toutouzas, Sillin et al, 2004; Yates, Sullivan, Clark, 2012). Lastly, due to the time constraints and researcher experience CTA was selected as it is streamlined, and not as resource intensive as techniques previously used in expert knowledge elicitation (see Militello & Hutton, 2000).

Questions Measuring NTS Levels in Trainees

To measure the variability of NTS training within OMFS surgical post-graduates, three NTS questions were added to a technical skills questionnaire (used by the second collaborative PhD project), and was given to 80 1st, 2nd, and 3rd year post-graduates. These NTS questions were general, but entailed surgical situations that had various outcomes describing the behavioural response of the participant. Although the questions were basic, each outcome specifically pertained to a different type of decision-making technique. Each choice answer within a question pertained to either recognition primed/rule-based, analytical/creative response, or to the complete opt-out of decision making, to facilitate any participants whom felt they would pass on the decision-making task to a superior. These questions were included to establish a general measure of the variation of NTS levels in trainees, and potentially highlight that although similar in technical/hands-on skills their NTS levels vary within the cohort.

Virtual Reality Hardware and Software

The current project is in progress, once the results have been analysed, the desired NTS training program will be created using the latest virtual reality headset along with the relevant PC hardware (Oculus.com, 2017) and software (Unity.com, 2017). Both a PC compatible Xbox 360 controller and a Leap Motion controller will be used. As this current project is heavily dependent on the programming of software to create novel and customized training material, internal software programming staff will be approached to aid in the creation of the desired content. This content will then be added to the previously mentioned technical skills VR surgical training device (Pulijala et al, 2017; *see also* hud.ac.uk, 2017).

RESULTS

Initial findings

The interview results are in the analysis stage*, however from the collective themes emerging from a thematic analysis of participants' transcripts, the cognitive interviews have highlighted several apparent task complexities and collectively provide information as to how experts overcome such complexities. For example, the pterygomaxillary dysjunction was stated as the hardest part of the procedure by all participants. The necessity for understanding the cues, data, and changes within the situation was apparent, one participant stated the following:

> '…you can't see it, it's just feel. That is one of the reasons why the trainees feel it's very difficult when it comes to this area. They don't understand what I'm doing…, I see the trainees when they put one single tap, they tend to use a lot of force. You know, so it goes a little uncontrolled. So, when I tell them to do two taps they tend to use the force in a more durațious manner. So, that probably is the reason why I tell them to do two taps…'

Furthermore, several specific decision-points within the Le Fort 1 Osteotomy emerged, these decision-points revolve around surgeons' abilities to make checks before performing a high-risk task (e.g. checks within pterygomaxillary dysjunction and maxillary osteotomy), cues/sensory information associated with performing the high-risk task itself (downfracture of Maxilla), and communicating the best course of action prior/during/after error occurrence within these tasks. Situational Awareness involved in these decisions were embedded within their responses, for example one participant recalled their thoughts in a problematic event:

*the result will be complete by the start of the NDM conference, with on-going development of training content

> '… "how am I going to do this?", as soon I use the smith's spreader I knew it [the Maxilla] will just collapse. You're thinking ahead on how to manage this case and how to gain my result, how to get my result and how to maintain it'.

Projection of the situation was a technique frequently recalled, for example another participant stated:

> '…as I am controlling the bleeder, I will start thinking of which other area I need to address. [Interviewer: Right, so you are going back?] Yes, back to go forward'.

This example represented the typical utilisation of previous understanding of situations to readjust current understanding, thus re-establishing a more accurate projective situational awareness after unpredicted change had occurred. Such cognitions from the previous examples can be extracted, categorised, and included into the training program to aid trainees' understanding of higher level non-technical application. The emerging themes within the social elements of NTS, being communication and teamwork, and leadership revolve around the inter-dependence of staff with a shared understanding of the task goals, ways to reach the goals, and understanding of desired behaviour. Upon decompressing the results of the cognitive interviews, approximately four training categories will be created each pertaining to a specific problematic area. The final content included in the VR training programme will be standardized via referencing with the Le Fort 1 Osteotomy Task Analysis and review from the participant located in the UK, being identified as an expert from the previously mentioned criteria.

Three-dimensional virtual surgical scenarios will be created that provide the user with a first-person perspective. These controlled environments will enable the user to deliberate on their responses and decisions as if in real-life. The subsequent intervention will allow explanation, walkthrough, and interactivity to illustrate the differences between the user's cognitions/choices/actions and those of a representative expert. Depending on resources available, a more sophisticated program could be created.

DISCUSSION

Use of Non-Technical Skills Course Training

Once created and empirically investigated, this project hopes to provide subsequent support for the abilities of VR simulation to be a training and diagnostic tool in surgical students' NTS acquisition and performance. As with other domains such as aviation (Moroney & Moroney, 1999), simulation can aid in the determination/evaluation of progression of each student. Specifically, the use of VR NTS simulations may offer standardized scenarios to ensure equivalent training dynamics, procedures, and feedback with the ability to repeat task training until reaching specific performance criteria. Moreover, performing comparisons within large sets of students can ensure thorough training and correct staging of progression for all students. This facilitates the identification of skills/techniques that need more attention, if not performed to minimal criteria. Pragmatic evidence of this can be found in Yorkshire and the Humber region in the UK. Local to those areas, simulators were successfully implemented for performance-driven evaluative purposes as part of skills training programmes (National Health Service, 2016).

The Digital Surgeon

Postulating the factors which may increase VR simulation deployment, the compatibility of the intended users with surgical simulation should be considered. The current student generation has been theorised to be more appropriate for increased technologically orientated training when compared to previous generations. Prensky (2001) suggested that due to the rapid uptake of digital technology, Generation Y could have a different way of thinking and processing new information in comparison to the previous Generation X. This has meant that previous teaching methods pertaining to Generation X may be outdated. Although counterintuitive to introduce another piece of technology, the composition of the millennial generation (DiLullo, McGee, & Kriebel, 2011) as digital natives has meant new practicing surgeons benefit from simulated VR training (e.g., Seymour, 2008; Seymour et al., 2002). This has resulted in decreased operating times and improve technical performance (Gurusamy, Aggarwal, Loizidou, & Davidson, 2009). The results from evaluation of the NTS training tool created in the current project will determine the pragmatic applicability of this perspective.

Complex Workloads

Furthermore, such training may be of more use to current students than with former training programmes, due to the increase in more sophisticated technology that are dependent upon the surgeon, consequently amplifying their cognitive workload in the OR. The shifting pace of technology has amplified students' cognitive demand during training stages, also generating a greater learning curve. The addition of more complex equipment has facilitated an increase in research literature to help reduce the cognitive workload of practicing surgeons during procedures (Banerjee & Gavade, 2015). However, for students this has not occurred to the same extent (however *see* Pulijala et al, 2017). Upon analysis and discussion of the results, this study could highlight how current teaching methods and facilities should be adapted with Human Factors research to alleviate high frequency and dependency of excessive cognitive workloads.

Finding the 'best' Training Method

Indeed, approximately 43.1% of UK trainees have stated they have experienced negative events including but not limited to unreal knowledge expectations, and poor/inadequate experiences of learning (Chapman et al, 2013). Via more rapid NTS skills acquisition using VR technology, this project may also conclude that students' negative experiences in the OR can be reduced. This has been shown to speed up training time and faster reach the ultimate goal of enhancing surgical performance and patient safety (Stefanidis, Wang, Korndorffer, Dunne, & Scott, 2010; Pulijala et al, 2017). Largely, there is a plethora of evidence to suggest training of vital non-technically orientated skills improve intra-operative cognitions and behavioural performance markers. However, without a unified formal agenda on how to teach these skills more research is necessary to understand if the synergy of advanced display technologies and Human Factors incorporation is the optimal implementation over more traditional methods, when training tomorrow's surgeons.

REFERENCES

Agha, R. A., & Fowler, A. J. (2015). The role and validity of surgical simulation. *International Surgery*, *100*(2), 350-357. doi: 10.9738/INTSURG-D-14-00004.1

Aiken, M. P., & Berry, M. J. (2015). Posttraumatic stress disorder: possibilities for olfaction and virtual reality exposure therapy. *Virtual Reality*, *19*(2), 95-109. doi: 10.1007/s10055-015-0260-x

Aïm, F., Lonjon, G., Hannouche, D., & Nizard, R. (2016). Effectiveness of virtual reality training in orthopaedic surgery. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, *32*(1), 224-232. doi: http://dx.doi.org/10.1016/j.arthro.2015.07.023

Alaker, M., Wynn, G. R., & Arulampalam, T. (2016). Virtual reality training in laparoscopic surgery: a systematic review & meta-analysis. *International Journal of Surgery*, *29*, 85-94. doi: 10.1016/j.ijsu.2016.03.034

ASPiH, (2016). *Simulation-Based Education on Healthcare – Standard Framework and Guidance.* Staffordshire, UK: Association for Simulated Practice in Healthcare.

Aydin, A., Khan, R., Khan, M. S., Dasgupta, P., & Ahmed, K. (2015). Simulation-based training for prostate surgery. *BJU international*, *116*(4), 665-674. doi: 10.1111/bju.12721

Banerjee, A., & Gavade, S. (2015). Integration of Devices in Operation Room for Reduction in Cognitive Load during Surgical Procedures. In Computational Science and Computational Intelligence, *International Conference* (pp. 821-824). IEEE. doi: 10.1109/CSCI.2015.111

Blackmur, J. P., Clement, R. G., Brady, R. R., & Oliver, C. W. (2013). Surgical training 2.0: How contemporary developments in information technology can augment surgical training. *The surgeon*, *11*(2), 105-112. doi: http://dx.doi.org/10.1016/j.surge.2012.12.001

Bourrier, Y., Jambon, F., Garbay, C., & Luengo, V. (2016). *An Interaction Centred Approach to the teaching of Non-Technical Skills in a Virtual Environment*. (Doctoral dissertation, Laboratoire d'Informatique de Paris 6; Laboratoire d'Informatique de Grenoble). Retrieved from https://hal.archives-ouvertes.fr/hal-01335038/file/ARTICLE_HAL_Interaction_NTS.pdf. Accessed on: 9 April 2017.

Campbell, J., Tirapelle, L., Yates, K., Clark, R., Inaba, K., Green, D.… & Sullivan, M. (2011). The effectiveness of a cognitive task analysis informed curriculum to increase self-efficacy and improve performance for an open cricothyrotomy. *Journal of surgical education*, *68*(5), 403-407. doi: 10.1016/j.jsurg.2011.05.007

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (Eds.). (2000). *Cognitive task analysis*. Psychology Press.

Coffey-Zern, S., Calvi, R. M., Vorrasi, J. S., & Meara, D. J. (2015). Incorporating Simulation into Oral and Maxillofacial Surgery Residency Education and Training: Christiana Care's Method. *Journal of Oral and Maxillofacial Surgery*, *73*(7), 1244-1245.

Crispen, P., & Hoffman, R. R. (2016). How Many Experts? *IEEE Intelligent Systems*, *31*(6), 56-62. DOI: http://dx.doi.org/10.1016/j.joms.2015.02.011

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272-309. DOI: 10.1080/15213269.2015.1015740

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64. Retrieved from http://uwf.edu/skass/documents/HF.37.1995-Endsley-Theory.pdf . Accessed on: 9 April 2017.

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A. K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies*, *82*, 48-56. doi: http://doi.org/10.1016/j.ijhcs.2015.05.004

Fletcher, G. C. L., McGeorge, P., Flin, R. H., Glavin, R. J., & Maran, N. J. (2002). The role of non-technical skills in anaesthesia: a review of current literature. *British Journal of Anaesthesia*, *88*(3), 418-429. doi: https://doi.org/10.1093/bja/88.3.418

Flin, R., Martin, L., Goeters, K. M., Hormann, H. J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, *3*, 97-120. Retrieved from https://www.abdn.ac.uk/iprc/documents/NOTECHS%20HFAS%20proof%20copy.pdf. Accessed on: 9 April 2017.

Flin, R., O'Connor, P., & Mearns, K. (2002). Crew resource management: improving team work in high reliability industries. *Team performance management: an international journal*, *8*(3/4), 68-78. doi: 10.1108/13527590210433366

Flin, R., Youngson, G. G., & Yule, S. (Eds.). (2015). *Enhancing surgical performance: A primer in non-technical skills*. CRC Press.

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., & Youngson, G. (2007). Teaching surgeons about non-technical skills. *The Surgeon*, *5*(2), 86-89. Retrieved from http://fhs.mcmaster.ca/surgery/documents/Flinetal2008.pdf. Accessed on: 9 April 2017.

Flowers, M. G., & Aggarwal, R. (2014). Second Life™: a novel simulation platform for the training of surgical residents. doi: 10.1586/17434440.2014.863706

Gahm, G., Reger, G., Ingram, M. V., Reger, M., & Rizzo, A. (2015). *A multisite, randomized clinical trial of virtual reality and prolonged exposure therapy for active duty soldiers with PTSD*. Geneva foundation Tacoma WA. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA614117. Accessed on: 9 April 2017.

Graafland, M., Schraagen, J. M. C., Boermeester, M. A., Bemelman, W. A., & Schijven, M. P. (2015). Training situational awareness to reduce surgical errors in the operating room. *British Journal of Surgery*, *102*(1), 16-23. doi: 10.1002/bjs.9643

Hee.nhs.uk. (2015). *Health Education England*. Retrieved from https://hee.nhs.uk/hee-your-area/south-west/our-work/sw-research-innovation/human-factors-patient-safety.

Hud.ac.uk, (2017). 3D virtual reality training for trainee dental surgeon unveiled. Retrieved from https://www.hud.ac.uk/news/2017/march/3dvirtualrealitytrainingfortraineedentalsurgeonunveiled.php. Accessed on: 9 April 2017.

Hscic.gov.uk. (2016). *NHS workforce statistics in England, summary of staff in the NHS 2004-2014.* Retrieved from http://content.digital.nhs.uk/catalogue/PUB16973. Accessed on: 9 April 2017.

Kim, K., Rosenthal, M. Z., Zielinski, D., & Brady, R. (2012). Comparison of desktop, head mounted display, and six wall fully immersive systems using a stressful task. In *Virtual Reality Short Papers and Posters*, (pp. 143-144). IEEE. doi: 10.1109/VR.2012.6180922

Koo, B., & Choi, S. (2015). SSVEP response on Oculus Rift. In *Brain-Computer Interface (BCI), 2015 3rd International Winter Conference on Brain-Computer Interface* (pp. 1-4). IEEE. doi: 10.1109/IWW-BCI.2015.7073028

Lin, D. T., Park, J., Liebert, C. A., & Lau, J. N. (2015). Validity evidence for Surgical Improvement of Clinical Knowledge Ops: a novel gaming platform to assess surgical decision making. *The American Journal of Surgery*, *209*(1), 79-85. doi: 10.1016/j.amjsurg.2014.08.033

Luker, K. R., Sullivan, M. E., Peyre, S. E., Sherman, R., & Grunwald, T. (2008). The use of a cognitive task analysis–based multimedia program to teach surgical decision making in flexor tendon repair. *The American Journal of Surgery*, *195*(1), 11-15. doi: 10.1016/j.amjsurg.2007.08.052

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618-1641. doi: http://dx.doi.org/10.1080/001401398186108

Mizuno Y, Narimatsu H, Kodama Y, Matsumura T, Kami M (2014). Mid-career changes in the occupation or specialty among general surgeons, from youth to middle age, have accelerated the shortage of general surgeons in Japan. Surgery Today; 44(4), pp. 601- 606. doi: 10.1007/s00595-013-0613-6

Moroney, W.F. and Moroney, B.W. (1999). Flight Simulation, *in Handbook of Aviation Human Factors*, Garland, D.J., Wise, J.A.., and Hopkins, V.D., Eds., Mahwah, NJ: Lawrence Erlbaum Associates.

Moulton, C. A., Regehr, G., Lingard, L., Merritt, C., & MacRae, H. (2010). Slowing down to stay out of trouble in the operating room: remaining attentive in automaticity. *Academic Medicine*, *85*(10), 1571-1577. doi: 10.1097/ACM.0b013e3181f073dd.

National Health Service. (2016). *Health Education England - Training Facilities*. Retrieved from http://www.yorksandhumberdeanery.nhs.uk/surgery/training_facilities/. Accessed on: 9 April 2017.

Nori, R., Piccardi, L., Migliori, M., Guidazzoli, A., Frasca, F., De Luca, D., & Giusberti, F. (2015). The virtual reality walking Corsi test. Computers in Human Behavior, 48, 72-77. doi: http://doi.org/10.1016/j.chb.2015.01.035

Oculus.com (2017). Next-generation virtual reality just got better. Retrieved from https://www3.oculus.com/en-us/rift/

Olmos, H., Gómez, S., Alcañiz, M., Contero, M., Andrés-Sebastiá, M. P., & Martín-Dorta, N. (2015). Combining Virtual Reality and Relaxation Techniques to Improve Attention Levels in Students from an Initial Vocational Qualification Program. In *Design for Teaching and Learning in a Networked World* (pp. 613-616). Springer International Publishing. doi: 10.1007/978-3-319-24258-3_71

OMFS Intercollegiate Surgical Curriculum (2016) *Educating the surgeons of the future*. Retrieved from https://www.iscp.ac.uk/static/public/syllabus/syllabus_omfs_2016.pdf. Last accessed 9 April 2017.

Pugh, C. M., DaRosa, D. A., Santacaterina, S., & Clark, R. E. (2011). Faculty evaluation of simulation-based modules for assessment of intraoperative decision making. *Surgery*, *149*(4), 534-542. doi: 10.1016/j.surg.2010.10.010

Pulijala, Y., Ma, M., & Ayoub, A. (2017). VR Surgery: Interactive Virtual Reality Application for Training Oral and Maxillofacial Surgeons using Oculus Rift and Leap Motion. In *Serious Games and Edutainment Applications* (pp. 187-202). Springer International Publishing.

Rcseng.ac.uk (2016). Surgery and the NHS in numbers. Retrieved from https://www.rcseng.ac.uk/media/media-background-briefings-and-statistics/surgery-and-the-nhs-in-numbers. Accessed on: 9 April 2017.

Robb, M., & Miller, G. (2012). Human factors engineering in oil and gas-a review of industry guidance. *Work*, *41*(Supplement 1), 752-762. doi: 10.3233/WOR-2012-0236-752.

Salvendy, G. (2012). *Handbook of human factors and ergonomics*. John Wiley & Sons.

Sarker, S. K., Rehman, S., Ladwa, M., Chang, A., & Vincent, C. (2009). A decision-making learning and assessment tool in laparoscopic cholecystectomy. *Surgical endoscopy*, *23*(1), 197-203. doi: 10.1007/s00464-008-9774-6.

Schaafstal, A., Schraagen, J. M., & Van Berl, M. (2000). Cognitive task analysis and innovation of training: The case of structured troubleshooting. *Human Factors*, *42*(1), 75-86. doi: 10.1518/001872000779656570

Schaverien, M. V. (2010). Development of expertise in surgical training. *Journal of surgical education*, *67*(1), 37-43. doi: 10.1016/j.jsurg.2009.11.002

Servais, E. L., LaMorte, W. W., Agarwal, S., Moschetti, W., Mallipattu, S. K., & Moulton, S. L. (2006). Teaching surgical decision-making: an interactive, web-based approach. *Journal of Surgical Research*, *134*(1), 102-106. doi: 10.1016/j.jss.2005.11.583

Seymour, N. E. (2008). VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World journal of surgery*, *32*(2), 182-188. doi: 10.1007/s00268-007-9307-9

Shaw, L. A., Wünsche, B. C., Lutteroth, C., Marks, S., & Callies, R. (2015). Challenges in virtual reality exergame design. Retrieved from http://aut.researchgateway.ac.nz/bitstream/handle/10292/8787/CRPIT%20Vol%20162%20-%20AUIC%202015%20-%20Excerpt.pdf?sequence=2&isAllowed=y. Accessed on: 9 April 2017.

Shepherd, A., & Stammers, R. B. (2005). Task analysis. *Evaluation of human work*, 129-57. In Wilson and Corlett, Evaluation of Human Work.

Shiban, Y., Reichenberger, J., Neumann, I. D., & Mühlberger, A. (2015). Social conditioning and extinction paradigm: a translational study in virtual reality. *Frontiers in psychology*, *6*. doi: 10.3389/fpsyg.2015.00400

Smink, D. S., Peyre, S. E., Soybel, D. I., Tavakkolizadeh, A., Vernon, A. H., & Anastakis, D. J. (2012). Utilization of a cognitive task analysis for laparoscopic appendectomy to identify differentiated intraoperative teaching objectives. *The American Journal of Surgery*, *203*(4), 540-545. doi: 10.1016/j.amjsurg.2011.11.002

Stefanidis, D., Sevdalis, N., Paige, J., Zevin, B., Aggarwal, R., Grantcharov, T..., & Association for Surgical Education Simulation Committee. (2015). Simulation in surgery: what's needed next? *Annals of surgery*, *261*(5), 846-853. doi: 10.1097/SLA.0000000000000826.

Sullivan, M. E., Brown, C. V., Peyre, S. E., Salim, A., Martin, M., Towfigh, S., & Grunwald, T. (2007). The use of cognitive task analysis to improve the learning of percutaneous tracheostomy placement. *The American journal of surgery*, *193*(1), 96-99. doi: 10.1016/j.amjsurg.2006.09.005

Sullivan, M. E., Yates, K. A., Inaba, K., Lam, L., & Clark, R. E. (2014). The use of cognitive task analysis to reveal the instructional limitations of experts in the teaching of procedural skills. *Academic Medicine*, *89*(5), 811-816. doi: 10.1097/ACM.0000000000000224

Surgeons.org. (2017). *Non-Technical Skills for Surgeons*. Retrieved from http://www.surgeons.org/for-health-professionals/register-courses-events/professional-development/non-technical-skills-for-surgeons._Accessed on: 9 April 2017.

Tran, L. N., Gupta, P., Poniatowski, L. H., Alanee, S., Dall'Era, M. A., & Sweet, R. M. (2013). Validation study of a computer-based open surgical trainer: SimPraxis® simulation platform. *Advances in medical education and practice*, *4*, 23. doi: https://doi.org/10.2147/AMEP.S38422

Unity.com. (2017). *Everything you need to succeed in games and VR/AR*. Retrieved from https://unity3d.com/. Accessed on: 9 April 2017.

Vannaprathip, N., Haddawy, P., Suebnukarn, S., Sangsartra, P., Sasikhant, N., & Sangutai, S. (2016, March). Desitra: A Simulator for Teaching Situated Decision Making in Dental Surgery. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (pp. 397-401). ACM. doi: 10.1145/2856767.2856807

Velmahos, G. C., Toutouzas, K. G., Sillin, L. F., Chan, L., Clark, R. E., Theodorou, D., & Maupin, F. (2004). Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. *The American Journal of Surgery*, *187*(1), 114-119. doi: http://doi.org/10.1016/j.amjsurg.2002.12.005

Waterworth, J. A., Waterworth, E. L., Riva, G., & Mantovani, F. (2015). Presence: form, content and consciousness. In *Immersed in Media* (pp. 35-58). Springer International Publishing. Retrieved from http://dnulib.edu.vn:8080/dspace/bitstream/DNULIB_52011/7481/1/immersed_in_media.pdf#page=47. Accessed on: 9 April 2017.

Westmidlandsdeanery.nhs.uk. (2017). *Core surgical training teaching timetables.* Retrieved from http://www.westmidlandsdeanery.nhs.uk/Specialty-Schools/Postgraduate-School-of-Surgery/Core-Surgical-Training-Teaching-Timetables/ModuleID/9889/ItemID/2235/mctl/EventDetails. Accessed on: 9 April 2017.

Who.com, (2015). *Patient Safety: Human Factors*. Retrieved from http://www.who.int/patientsafety/research/methods_measures/human_factors/en/. Accessed on: 9 April 2017.

Yates, K., Sullivan, M., & Clark, R. (2012). Integrated studies on the use of cognitive task analysis to capture surgical expertise for central venous catheter placement and open cricothyrotomy. *The American Journal of Surgery*, *203*(1), 76-80. doi: 10.1016/j.amjsurg.2011.07.011

Yiasemidou, M., Glassman, D., Tomlinson, J., Song, D., & Gough, M. J. (2017). Perceptions about the Present and Future of Surgical Simulation: A National Study of Mixed Qualitative and Quantitative Methodology. *Journal of Surgical Education*, *74*(1), 108-116. doi: 10.1016/j.jsurg.2016.07.011.

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006). Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery*, *139*(2), 140-149. doi: 10.1016/j.surg.2005.06.017

The role of expertise and organisational information in managing patient flow in hospitals

Matthew WOODWARD[a], Julie GORE[b], Christos VASILAKIS[a]
[a]*Centre for Healthcare Innovation & Improvement (CHI[2]), School of Management, University of Bath*
[b]*School of Management, University of Bath*

ABSTRACT
This research concerns naturalistic decision making by clinicians and operational managers regarding patient flow in hospitals. Naturalistic decision making in the healthcare domain has typically focussed upon clinical decisions by doctors and nurses, however patient care is enabled by a system in which many other roles manage dynamic situations. This research focuses on operational managers and clinical coordinators who strive to interpret system state to make decisions to effectively, efficiently and safely move patients through care. Of particular interest is how the availability and representation of organisational information influences decision making processes.

KEYWORDS
Planning and prediction; health care; patient flow; simulation models; mental models

INTRODUCTION

The field of naturalistic decision making (NDM) has long held an interest in the domain of healthcare as a source of research. As Bogner (1997) notes, in healthcare the locus of time stress is usually outside of the decision maker and decisions are reactive to patient needs, resulting in a demanding environment. Secondary and tertiary hospital services in particular are dynamic, complex systems with multiple people and functions interacting in a time-pressured environment. This makes them fertile ground for the field of naturalistic decision making which has a track record of studying systems in which complexity and uncertainty are inherent (Gore, Flin, Stanton, & Wong, 2015) The local situation can change rapidly, information is often incomplete or ambiguous, expertise is important and social structures and team working are integral to the hospital's function. For example, in tertiary care settings both the state of the organisation and the state of the patient are dynamic. Emergency Departments (ED) can change from calm to busy within minutes, patients can deteriorate from relatively stable to critically ill over a period of hours. Some patients may be referred with a full medical history, others may present in an emergency situation with nothing but basic demographic information.

Given such potential the NDM approach has been applied in several studies of a number of healthcare domains. For example Crandall (1993) studied the cues used by neonatal nurses Phipps and Parker (2014) have used the NDM approach to examine anaesthetists rule-related behaviour, and Xiao et al (1997) studied the mental preparatory work utilised by anaesthetists.

However the literature has primarily focussed upon clinical decisions made by clinicians; those intervening with the patient at the point of care. Yet over the last few decades the increasing complexity and specialisation of medical care has resulted in healthcare that requires intervention from different teams involving many specialised professionals (Gawande, 2010). These professionals are formed into organisations which include not only clinicians but managers, technicians, coordinators and administrators who directly and indirectly impact the quality of care delivered. The development of frameworks such as the Systems Engineering Approach to Patient Safety (SEIPS) (Holden et al., 2008) recognises that healthcare is a sociotechnical system which as a whole has significant bearing upon the safety and quality of care delivered.

Thus this research takes a broader perspective of organisational participants by seeking to understand the decision processes of hospital operational managers and clinical coordinators in facilitating the flow of patients through a care service. That is those staff who seek to understand the current situation and plan ahead regarding the state of the system in terms of available beds, patient condition, patient location, staffing and more. Nemeth et al (2006) has taken a step in this direction by using a NDM approach to research how anaesthesia co-ordinators make decisions with an emphasis on understanding the artefacts used by such roles.

One approach to assist with planning is that of operational research (O.R.) , which closely relates to management science and business analytics (Liberatore & Luo, 2010). This discipline typically seeks to abstract elements of a service (healthcare or otherwise) into a representation/model and uses quantitative data to describe system behaviour. Some models have the scope to compute the most probable process outcomes and thus can be used to analyse and optimise system configurations at a tactical level. The use of computer simulation models in healthcare particularly at strategic and tactical levels is reasonably well established; academics and consultants have developed models to help senior managers make structural decisions about service design. For example a recent study developed a discrete event simulation to model the implementation of stroke thrombolysis. Having established a validated baseline model, different interventions (such as increasing drug treatment time windows) could be modelled and treatment rates predicted (Monks, Pearson, Pitt, Stein, & James, 2015).

However recent position papers and systematic reviews regarding the application of computer simulation modelling to healthcare (Pitt, Monks, Crowe, & Vasilakis, 2015; Sobolev, Sanchez, & Vasilakis, 2011) have raised the issue of an implementation gap between the outputs provided by modellers and the engagement with and implementation of results by healthcare staff. Other recent papers in operational research have identified the importance of the behavioural aspects of OR and have identified opportunities for research in this area (Hämäläinen, Luoma, & Saarinen, 2013). They argue that given O.R. is used to facilitate thinking and problem solving, an understanding of the human-model interaction is important for making the best use of such tools.

Thus this research seeks to develop a richer understanding of the context within which such decision support models are developed and utilised. As Gore et al (2006) note in their illustrations of organisational decision making (ODM) in healthcare, "Consideration of the incentives and penalties surrounding the decision space sits well with the remit of ODM, particularly insofar as the decision space may be highly ambiguous, involving multiple parties each with different degrees and levels of accountability." Similarly NDM and the field of human factors in general seeks to use empirical data about the interactions between humans and organisational information to develop "ecologically sensitive" solutions (Gore et al, 2006). There is opportunity therefore to use NDM methods to elicit the *mental* models ("mental representations of specific situations" (Lipshitz & Shaul, 1997)) held by those making decisions about patient flow. These could then be usefully applied to the development of explicit (rather than tacit) models/representations that have the potential for learning about and anticipating patient flow dynamics. The engagement of organisational actors in this way and the provision of feedback loops for organisational learning could help forge links across the reported implementation gap.

RESEARCH QUESTION

This research thus seeks to bring together these two approaches to the analysis of work systems: naturalistic decision making and systems modelling. The research question asks: How are decision-making processes influenced by the availability and representation of system state? The concept of distributed cognition (Hutchins, 1995) will be used to frame an investigation into how knowledge regarding system state is held and shared between human and technology actors. NDM methods (Crandall, Klein, & Hoffman, 2006) will be used to elicit the perceptual cues, organisational information and mental models used by those influencing patient flow. Then an explicit representation of system state, for example a computer simulation model using resource/demand data, will be developed and used to explore how these decision processes are influenced.

The work domain will be tertiary care hospitals and the decision makers studied will be those who manage the movement of patients through hospital services and facilities, namely doctors, nurses, clinical coordinators and operational managers. Decisions may include:

- *when* to transfer a patient from one hospital department to another, e.g. emergency department (ED) to medical assessment unit (MAU)
- *which patient* to prioritise for transfer (particularly from ED)
- *which physical location/bed* to transfer a patient to
- *when* to escalate a situation (i.e. when the service is starting to fail in delivering it's function)
- *how* to manage patients close to breach (i.e. of organisational/government targets).

More specific questions will be pursued depending upon how enquiry develops, for example:

- How does the availability, representation and reliability of information/models impact the timeliness of decisions (e.g. proactive versus reactive)?
- How does the availability, representation and reliability of information/models impact upon human-model interactions (e.g. frequency, data updates, communications)?
- Do information/model artefacts change team communication?
- How do decision-makers interpret trust in information/models?
- How do decision-makers interpret confidence in decisions made by themselves and others?

APPROACH

This research is at an early stage of development (first year doctoral study) and an agreement of the methods to be used is in the process of being finalised at the time of submission. The intention is to conduct a pilot study prior to the PhD transfer point (March 2018) to improve researcher skills with the methods and to gain a sense of the data content and structure.

The intention is to take an interpretative/neo-empirical stance as defined by Johnson et al (Johnson, Buehring, Cassell, & Symon, 2006). In this approach the qualitative and subjective views of the organisational actors are of primary interest yet the intention is to report these interpretations whilst striving to minimise the biases of the researcher. The methodological approach is mixed-methods as quantitative data is integral to the work system under scrutiny and thus quantitative methods will also be used to develop a model for the intervention phase.

The approach is particularly focussed upon understanding the role of expertise and experience and how this relates to the use of organisational information. Some NDM models (e.g. the recognition-primed decision making model Klein, 1998) emphasise the importance of recognising environmental cues to support situation pattern-matching. The representation/visualisation of the current state of a hospital service (e.g. Emergency department through to emergency general surgery) could be designed to help pick-up on these cues.

PROJECT DESIGN

Three phases will be undertaken to: 1) study and interpret current decision-making processes, 2) develop a representation of system state for example, with a simulation or analytical model and, 3) evaluate how the use of such a representation influences decision making. Stages 1) and 3) will collect empirical data in the field, in this case within a large tertiary care hospital in the UK.

Phase 1 will analyse current practice by engaging with stakeholders and describing the baseline work system using task and process analyses. An emphasis will be placed on establishing decision points important to patient flow. The criteria for “important” decisions will be those that influence process effectiveness, patient safety or elicit rich decision-making processes. An ethnographic approach will be taken to collect this data through periods observing and interacting with staff in the hospital. The rationale for such an approach is to capture the wide range of artefacts that are anticipated to contribute to distributed cognition and also to build working relationships with the staff under study. Participatory methods will be used to develop the work system description and gain support for the project.

Established qualitative NDM knowledge elicitation techniques will be used with individuals to further study the decision points of interest. For example the Critical Decision Method will be employed to elicit the important cues and criteria that trigger the recognition of certain situations (for example preparation to escalate operational situation). This method is planned to be used with a wide range of actors to capture both different levels of expertise and also different social perspectives (for example clinical, operational, safety). It is likely that the Recognition-Primed Decision making model (RPD, Klein 1998) will be used as a framework to structure interview questions.

An understanding of current process/organisational information sources will be developed through an audit of business intelligence data that is available or utilised (e.g. workplace dashboards, routine audits, operational data). The alignment of these sources with the decision requirements elicited earlier in Phase 1 will be evaluated with the involvement of healthcare staff. This phase will also concern data gathering for potential simulation model population, for example staffing level rotas, service demand profiles and task/process durations and distributions. Where possible this data would be collected retrospectively from existing NHS data records. Some data (for example specific task durations) may need to be collected prospectively through observation or updates to existing recording systems.

In the second phase the intention is to develop an artefact that will provide decisions makers with a representation of system state. The preferred option is to develop a computer simulation model that may be used to anticipate future service state given a set of input variables. These variables could mirror the current system state or could be pre-set to represent different scenarios (for example low staffing or bed closures). Tests of model validity and sensitivity will be conducted. It is expected that there will be iteration between data gathering and model development as a better understanding of the system is gained. If the pragmatic demand for simplicity prevails the artefact may instead provide a near real time representation (tabular or graphical) of organisational information, for example service bed state, patients in ED etc.

The third phase will make the artefact available for the participants to interact with, and for data on user interactions and interpretation to be collected. There are two options under consideration for the intervention. In the first the artefact is made available to the participants in situ with a view to supporting their operational decisions on a daily basis (for example at a bed managers planning meeting). An alternative option is to provide the intervention as training sessions away from the operational environment in which participants are able to explore the impact of different scenarios using the model as a way of gaining a sense of system behaviour.

During this period methods including observation, verbal protocol analysis and the critical decision method will be used to evaluate model use and understanding. In particular the evaluation will look to understand if the participants interpret the hospital service in a different way and whether this effects their decision-making processes. The questions outlined in the section above will be explored. Quantitative data derived partly from the model (e.g. time to decide following a new piece of information) *may* be used as a metric to help determine decision quality.

RISKS

A barrier to this research may be the time made available by busy clinicians and managers in the hospital to engage to allow qualitative methods to be employed with sufficient rigour. Obtaining reliable quantitative process data could also be a challenge due to issues with organisational data coding and availability. Both of these risks may be mitigated by making clear the potential benefit of the research to the health service early in the project, engaging those teams who are motivated to be involved and choosing services with reasonable existing data recording.

QUESTIONS ON METHODS

The doctoral consortium will provide a timely opportunity to discuss and gain feedback upon potential methodological approaches. The following methods are under consideration and are listed with some questions for discussion.

Ethnography

- How could one best capture decisions and information use in a timely manner when they are distributed across multiple physical and social spaces?
- Is it legitimate to enrol staff as data gatherers (for example keeping a record of where they look for information)?
- With the ethnographic method to what extent can questions be posed directly (for example about choice of information source, which may then alter the observed behaviour) as opposed to passive observation?

Interviews - Critical Decision Method (CDM)

- Are there different techniques for eliciting information about proactive versus reactive decisions? Is it helpful to make a distinction, for example a decision taken in response to an event (e.g. the arrival of patient to a minor injuries unit) may be interpreted as planning ahead (arguably a proactive decision)?
- In the use of CDM how does one account for hindsight (i.e. creating a narrative to construe the decision processes in light of the outcome)?
- How many different participants should be interviewed to gain a credible and dependable account of decision-making in this context?

Artefact/model interaction

- Is it valid/credible to use quantitative data (e.g. timings) to establish the quality of a decision?
- If study of model interaction is taken out of the working environment (into a training or experimental session) can it still be classed as naturalistic decision making?

CONCLUSION

This paper outlines the direction for first year management and psychology doctoral research. The intention is to bring together the approaches of naturalistic decision making and simulation modelling to understand how organisational information can be effectively represented to help manage hospital patient flow. At this early stage of research there are many questions to be addressed regarding approach and methodology. This paper seeks to elicit discussion and guidance from those attending the NDM doctoral consortium 2017.

REFERENCES

Bogner, M. (1997). NDM in Health Care. In G. Zsambok, C. E.; Klein (Ed.), *Naturalistic Decision Making - Proceeds of 1997 Conference* (pp. 61–69).

Crandall, B., Klein, G., & Hoffman, R. (2006). *Working Minds: a Practitioner's Guide to Cognitive Task Analysis*. *Perspectives in biology and medicine*.

Gore, J., Banks, A., Millward, L., & Kyriakidou, O. (2006). Naturalistic Decision Making and Organizations: Reviewing Pragmatic Science. *Organization Studies*, *27*, 925–942.

Gore, J., Flin, R., Stanton, N., & Wong, B. L. W. (2015). Applications for naturalistic decision-making. *Journal of Occupational and Organizational Psychology*, *88*(2), 223–230.

Hämäläinen, R. P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, *228*(3), 623–634.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Schoofs Hundt, A., Ozok, A. A., & Rivera-Rodriguez, A. J. (2008). SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*.

Hutchins, E. (1995). Cognition in the Wild. *MIT Press*.

Johnson, P., Buehring, A., Cassell, C., & Symon, G. (2006). Evaluating qualitative management research: Towards a contingent criteriology. *International Journal of Management Reviews*, *8*(3), 131–156.

Liberatore, M. J., & Luo, W. (2010). The analytics movement: Implications for operations research. *Interfaces*, *40*(4), 313–324.

Lipshitz, R., & Shaul, O. (1997). Schemata and Mental Models in Recognition-Primed Decision Making. In G. Zsambok, C; Klein (Ed.), *Naturalistic Decision Making - Proceeds of 1997 Conference* (pp. 293–303).

Monks, T., Pearson, M., Pitt, M., Stein, K., & James, M. A. (2015). Evaluating the impact of a simulation study in emergency stroke care. *Operations Research for Health Care*, *6*, 40–49.

Pitt, M., Monks, T., Crowe, S., & Vasilakis, C. (2015). Systems modelling and simulation in health service design, delivery and decision making. *BMJ Quality & Safety*, bmjqs-2015-004430.

Sobolev, B. G., Sanchez, V., & Vasilakis, C. (2011). Systematic review of the use of computer simulation modeling of patient flow in surgical care. *Journal of Medical Systems*, *35*, 1–16.

Xiao, Y., Milgram, P., & Doyle, J. (1997). Capturing and Modeling Planning Expertise in Anesthesiology: Results of a Field Study. In G. Zsambok, C. E; Klein (Ed.), *Naturalistic Decision Making - Proceeds of 1997 Conference* (pp. 197–205).

Bi-annual International Conference on Naturalistic Decision Making

NDM13

HISTORY OF NDM CONFERENCES

1989	Dayton	**NDM1** set the stage for expanding the study of problem solving and decision making, linking it to expertise studies, making it more pertinent to the needs of the applied community, and giving greater focus on national needs. This Conference served as a "call."	Klein, G., Orasanu, J., Calderwood R., & Zsambok, C. E. (Eds.), (1993*). Decision making in action: Models and methods.* New York: Ablex Publishing.
1994	Dayton	**NDM2** was more specific, dealing with a host of application areas and some tentative results from NDM work. Ideas for future directions were charted since NDM was still largely a promissory note.	Zsambok, C. E., & Klein, G. (Eds.), (1997). *Naturalistic Decision Making. NJ: Lawrence Erlbaum.*
1996	Aberdeen, Scotland	**NDM3** highlighted the interest in NDM on the part of European researchers, and served to integrate the ideas of NDM with the existing paradigms in the European community, such as Work Analysis.	Flin, R., & Salas, E. (Eds.). (1998). *Decision making under stress: Emerging themes and applications.* London: Ashgate
1998	Washington DC	**NDM4** represented some of the pay-off from the initial promissory notes. A host of research studies was presented on diverse topics. There was a healthy debate on the relation of NDM to other paradigms, including those of human factors and "cognition in the wild."	Salas, E., & Klein, G. (2001). *Linking expertise and naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.
2000	Stockholm, Sweden	**NDM5** was organized around a matrix combining methodology (Cognitive Task Analysis, Observational Methods, Microworld Techniques) and application areas (Distributed Decision Making, Decision Errors, Learning From Experience, Motivation and Emotion, and Situation Awareness and Training).	B. Brehmer, R. Lipshitz, & H. Montgomery (Eds.). *(2004). How professionals make decisions.* Mahwah, NJ: Lawrence Erlbaum Associates.
2003	Pensacola Beach, FL	**NDM6** addressed the issues that experts face in situations that fall outside 'the routine'. Other discussions included NDM and cognitive task analysis methodology, NDM and traditional lab-based DM, and microcognition to macrocognition.	Hoffman, R. R. (Ed.) (2006). *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making.* Mahwah, NJ: Erlbaum.
2005	Amsterdam, The Netherlands	**NDM7** emphasized five themes: adaptive decision support, cognitive ethnography, crime and decision making, crisis management, and medical decision making. In sessions, the NDM framework was applied to new and diverse domains, such as landmine detection, judgments in crime situations, and space exploration.	Schraagen, J.M., Militello, L., Ormerod, T. & R. Lipshitz (Eds.). *(2008). Naturalistic Decision Making and Macrocognition.* London: Ashgate.
2007	Monterey, California	**NDM8** represented the diversity of research within NDM including: knowledge management, applications to organisations and teams and military security operations. Debate centred upon the appropratemenss of the macro-cognition construct and the methodological challenges that continue to face the field.	Mosier, K. L. & Fischer, U. M. (Eds.) (2010). *Informed by Knowledge: Expert Performance in Complex Situations.* Sussex, UK: Taylor & Francis.
2009	London, UK	**NDM9** addressed the effect of modern computing technology on decision making that occurs in naturalistic settings such as medical diagnosis and treatment, command and control, financial markets, information analysis, team decision making and coordination.	Wong, B. L. W. and Stanton, N. A. (Eds.) *(2009) Naturalistic Decision Making and Computers: Proceedings of the 9th Bi-annual International Conference on Naturalistic Decision Making*, British Computer Society, London Stanton, N.A., Wong, W., Gore, J., Sevdalis, N. & Strub, M. (2011) Critical Thinking. *Theoretical issues in Egonomics Science.* 12 (3): 204-209
2011	Orlando, Florida	**NDM10** brought together researchers and practitioners from diverse domains who seek to understand and improve how people actually perform cognitively complex functions in demanding situations.	Fiore, S. M. & Harper-Sciarini, M. (2011*). Proceedings of the 10th Bi-Annual International Conference on Naturalistic Decision Making.* University of Central Florida. ISBN 978-1-4507-7872-5 *Harper, M. & Sciarini, L. (2011)* Selected Papers from the 10th Bi-Annual International Conferennce on Naturalistic Decision Making. *The International Journal Cognitive Technology 16: 2, 1-63*
2013	Marseilles, France	**NDM11** focussed on sensemaking, trust and uncertainty management and expertise interacting with technical systems across a wide range of operational domains.	*Chaudet, H., Pellegrin, L., Bonnardel (2015)* Special issue on the 11th conference on Naturalistic Decision Making, *Cognition, Technology & Work Volume 17: 3, 315–318*
2015	Washington DC	**NDM12** extended NDM thinking reaching across domains, disciplines and applications. Since the first 1989 NDM conference the NDM community of practice has grown worldwide extending well beyond the early fire ground commander studies hence an integration of multidisciplinary efforts to improve work in complex domains	*Mosier, K. & Militello, L. (2016)* Extending Naturalistic Decision Making: Reaching across domains, disciplines, and applications. *Journal of Cognitive Engineering and Decision. 10, (3), 227-228* *Adapted from Robert Hoffman, NDM 6 Organizer

UNIVERSITY OF BATH
SCHOOL OF MANAGEMENT

University of HUDDERSFIELD