Providing support for the use of analogies in demand forecasting tasks

Wing Yee Lee, Paul Goodwin, Robert Fildes,
Konstantinos Nikolopoulos, & Michael Lawrence

University of Bath
School of Management
Working Paper Series
2005.18

University of Bath School of Management
Working Paper Series

School of Management
Claverton Down
Bath
BA2 7AY
United Kingdom
Tel: +44 1225 826742
Fax: +44 1225 826473
http://www.bath.ac.uk/management/research/papers.htm

2005		
2005.01	Bruce A. Rayton	Specific Human Capital as an Additional Reason for Profit Sharing
2005.02	Catherine Pardo, Stephan C. Henneberg, Stefanos Mouzas and Peter Naudè	Unpicking the Meaning of Value in Key Account Management
2005.03	Andrew Pettigrew and Stephan C. Henneberg (Editors)	Funding Gap or Leadership Gap – A Panel Discussion on Entrepreneurship and Innovation
2005.04	Robert Heath & Agnes Nairn	Measuring Affective Advertising: Implications of Low Attention Processing on Recall
2005.05	Juani Swart	Identifying the sub-components of intellectual capital: a literature review and development of measures
2005.06	Juani Swart, John Purcell and Nick Kinnie	Knowledge work and new organisational forms: the HRM challenge
2005.07	Niki Panteli, Ioanna Tsiourva and Soy Modelly	Intra-organizational Connectivity and Interactivity with Intranets: The case of a Pharmaceutical Company
2005.08	Stefanos Mouzas, Stephan Henneberg and Peter Naudé	Amalgamating strategic possibilities
2005.09	Abed Al-Nasser Abdallah	Cross-Listing, Investor Protection, and Disclosure: Does It Make a Difference: The Case of Cross-Listed Versus Non-Cross-Listed firms
2005.10	Richard Fairchild and Sasanee Lovisuth	Strategic Financing Decisions in a Spatial Model of Product Market Competition.
2005.11	Richard Fairchild	Persuasive advertising and welfare in a Hotelling market.
2005.12	Stephan C. Henneberg, Catherine Pardo, Stefanos Mouzas and Peter Naudé	Dyadic 'Key Relationship Programmes': Value dimensions and strategies.
2005.13	Felicia Fai and Jing-Lin Duanmu	Knowledge transfers, organizational governance and knowledge utilization: the case of electrical supplier firms in Wuxi, PRC

2005.14	Yvonne Ward and Professor Andrew Graves	Through-life Management: The Provision of Integrated Customer Solutions By Aerospace Manufacturers
2005.15	Mark Ginnever, Andy McKechnie & Niki Panteli	A Model for Sustaining Relationships in IT Outsourcing with Small IT Vendors
2005.16	John Purcell	Business strategies and human resource management: uneasy bedfellows or strategic partners?
2005.17	Richard Fairchild	Managerial Overconfidence, Moral Hazard, and Financing and Investment Decisions
2005.18	Wing Yee Lee, Paul Goodwin, Robert Fildes, Konstantinos Nikolopoulos, & Michael Lawrence	Providing support for the use of analogies in demand forecasting tasks

Providing support for the use of analogies in demand forecasting tasks

Wing Yee Lee[1*], Paul Goodwin[1],

Robert Fildes[2], Konstantinos Nikolopoulos[2],

Michael Lawrence[3],

[1] School of Management, University of Bath, Bath, BA2 7AY, United Kingdom.

**email:* mnpwyl@bath.ac.uk

[2] Department of Management Science, Lancaster University Management School, LA1 4YX, United Kingdom.

[3] School of Information Systems, Technology and Management, University of New South Wales, Sydney 2052, Australia

November 2005 v2.4

* Corresponding author

ABSTRACT

Management judgment is widely used to adjust statistical forecasts in order to take into account special events, such as sales promotions. There is evidence that forecasters often use analogous past events to estimate the effects of a future special event. However, the unaided forecaster using such an approach may suffer from errors in recall, difficulties in making judgments about similarity and difficulties in adapting the analogous cases to the specific attributes associated with the future case. An experiment was carried out to investigate whether a forecasting support system which provides users with guidance on similarity judgments and support for adaptation judgments could lead to more accurate forecasting of the effects of sales promotions. The experiment suggested that a simple, and easily implemented, form of adaptation support could significantly improve forecast accuracy under some conditions. The results also indicated that the support would be likely to be acceptable to potential users of the system.

Key words

Judgmental forecasting, forecasting by analogy, forecasting support system, sales promotions

Acknowledgements

This research was supported by Engineering and Physical Sciences Research Council (EPSRC) grants GR/60198/01 and GR/60181/01.

INTRODUCTION

Research over the last twenty-five years has shown that judgmental adjustments to statistical forecasts can improve accuracy under appropriate conditions (Goodwin, 2001). In particular, judgment can be valuable when the forecaster has access to important information about a forthcoming event that is not available to the statistical method (Sanders, 2001). A typical event in sales forecasting would be a sales promotion campaign. Useful data on the effects of such campaigns might be scarce because of their infrequency or diverse nature and, as such, would preclude the use of advanced statistical methods to estimate their effects. In addition to improvements in accuracy, judgmental adjustments may also give forecasters a sense of ownership of the forecasting process and they may value the opportunity to apply their expertise. This is likely to increase the acceptability of the adjusted forecasts relative to those derived from an arcane statistical method (Taylor and Thomas 1982; Lawrence, Goodwin et al. 2002).

There have been a number of papers which have investigated the conditions favouring judgmental adjustment (e.g. Willemain 1989; Willemain 1991) and methods for encouraging the forecaster to restrict adjustments to occasions when these conditions apply (e.g. Goodwin, 2000). However, the question of how forecasters can be supported in their task of estimating the *size of adjustments* has largely been unexplored by researchers (three exceptions are the papers by Wolfe and Flores (1990) and Flores et al (1992) and Webby (1993) which all explored the merits of decomposing the judgmental adjustment task). An intuitively reasonable principle of any judgmental support method is that it is more likely to be acceptable if it is congruent with the natural thinking processes of forecasters. One approach to estimation that people appear to use naturally when making estimates is the use of analogies. For example, in an experiment which required people to forecast the demand for a product in periods when it was being promoted with a given level of expenditure, Goodwin and Fildes (1999) found evidence that the forecasters looked back for the past promotion that had the most similar expenditure and used the sales associated with this promotion as a basis for their forecast. Hoch and Schkade (1996) found similar ‘pattern matching’ tendencies in a forecasting task involving credit ratings of loan applicants, while, McIntyre, Achabel and Miller (1993), in a field study, observed the use of analogies by expert buyers at specialty and department stores when they forecast the effects of promotions. In addition, our own observations of sales forecasting meetings in supply chain companies has found that forecasters commonly search for analogous past circumstances to those which will prevail in the forecast period in order to establish a basis for their judgmental forecasts. For instance, the forecasters of a beverage company used the sales of a previous similar sport tournament as a reference for the forecast of the sales in the upcoming football championship.

Despite its apparent widespread use by forecasters, the use of analogies may be problematical for several reasons. First, the forecaster may have to use memory to recall the most similar cases (e.g. the most similar promotion campaigns). This places not only a burden on the forecaster’s memory, but also requires them to make demanding judgments about the similarity of the current case to past cases. Any errors associated with this process are likely to be to the detriment of forecast accuracy. Moreover, limitations in human information processing capacity may mean that reliance is placed on a single recalled case. The noise associated with the outcome of this case may reduce its reliability as a predictor for future similar cases. Finally, recalled cases are unlikely to be identical to the case that is being predicted. The forecaster will therefore have to carry out the potentially difficult task of adapting the outcome of the past case(s) to take into account the aspects of the future case that are different. All of this suggests that a forecasting support system may be effective in improving the accuracy of judgments if it includes facilities which: 1) reduce the demands on memory, 2) provide guidance on similarity and 3) provide information to support the adaptation judgments.

In this paper we report on an experiment that investigated the effectiveness of providing various levels of support for the use of analogies in sales forecasting. Analogies, in the form of similar promotion campaigns, were available to help forecasters to determine how much they should adjust statistical sales forecasts to take into account the effects of forthcoming promotions. The paper is organised as follows. We first discuss the cognitive process underlying Forecasting by Analogy (FBA) and why the approach is of interest. Next, we review previous work related to the support of FBA in sales forecasting. On the basis of this review, we describe and explain the design of the forecasting support system that was developed. This is followed by a description of the experiment carried out and its results and analysis. Finally, we consider the limitations of this research and its implications for future research.

Literature Review

The concept of applying analogues to reasoning has been explored in many research fields including psychology, artificial intelligence and decision support systems. Labels for the concept have varied according to the research tradition and the applications of the analogues. In psychology, it is termed 'pattern matching' and is found to be a basic component of many human cognitive models (Lindsay and Norman 1977). In artificial intelligence, it is known as Case-based Reasoning (CBR) and many expert systems have been built to emulate and/or assist people like chefs (to create recipes) and architects (to design buildings) (Kolodner 1993). Kolodner has argued that CBR is likely to be useful "when knowledge is incomplete and/or evidence is sparse", when the domains "are not completely understood by the reasoner", and when "no algorithmic method is available for evaluation" (Kolodner 1993, pp24-25).

Surprisingly, relatively few applications of analogies in forecasting have been reported. Duncan et al (2001) applied the approach to time series forecasting, while Nikolopoulos et al (2005) used it to forecast TV audience ratings (in this application the process was referred to as nearest neighbour analysis). McIntyre et al (1993) constructed an expert system, based on the use of analogies by buyers in specialty and department stores, to forecast the effects of sales promotions. However, in all these studies either an algorithm was used to derive forecasts from the analogies or the judgmental process was formalised in an expert system and the resulting forecasts were obtained without interaction with the forecaster. Only Green (2005), who investigated analogies in forecasting the outcomes of conflicts and Hoch and Schkade (1996), who applied it to loan applicant credit rating forecasts, considered how the judgmental process associated with analogies could be supported. As we indicated earlier, in some environments it is necessary and/or more acceptable to apply judgments to forecasting. However, judgments are often undermined by human biases, which may be mitigated by the provision of some support.

What are likely to be the support needs of forecasters who employ analogies? The process of using analogies to produce forecasts is likely to involve three stages: i) recall, ii) similarity judgments and iii) adaptation judgments. At each of these stages cognitive limitations may reduce the accuracy of the judgments that are made. When confronted with a forecasting task, the forecaster first needs to recall some past cases. Then, he/she needs to judge how suitable each of these cases is as a reference for the current forecasting task. The suitability is likely to be determined by how similar the past case is to the current one. Finally, when one or several cases are selected, the forecaster needs to adapt the outcomes of these cases according to the differences between them and the target case, before determining the final forecast value.

A number of issues relate to the unaided forecaster's ability to recall analogous past cases. First, human memory limitations may mean that only a small sample of past cases may be recalled. Secondly, the details of these cases may be recalled incorrectly. Our memory is organised based on differences between incidents and norms (Schank 1982); the more 'unusual' the case is, the more likely we remember, and hence recall it. The rest of the incidents may become blended and turned into more abstract rules. Such phenomenon might be particularly common for decision makers with years of experience, as shown in the review by Klein and Calderwood (1988). Similarly, as a forecaster has probably dealt with many past cases, it is unlikely that he/she can remember every single one correctly to search for the most useful analogues. Klein and Calderwood also argued that old cases are often recalled serially, as opposed to concurrently, based on familiarity. The goal is to satisfice, that is, being good enough, rather than to optimise. If this also applies to sales forecasting using FBA, it would mean that forecasters might not apply the most appropriate cases. One obvious way of providing support to overcome these limitations of memory is to provide a database of past cases. However, when Hoch and Schkade (1996) provided a database of 50 past cases they found that it was only effective, relative to no support, when the noise associated with the variable to be forecast was low – it led to relatively inaccurate forecasts when the noise level was high. Thus, on its own, memory support may not be sufficient to improve forecast accuracy. This suggests that the provision of more extensive support which embraces the later stages of FBA is worth investigating.

Even when people remember all the relevant cases they may use inappropriate ones due to reliance on irrelevant associations (Gilovich 1981). Psychology experiments have shown that people are not good at recalling useful analogues and tend to rely on cases that are easier to bring to mind (Gentner 1989). Alternatively, they may require hints to apply the suitable analogues (Holyoak 1985; Gentner 1989). In his research into forecasting in conflict situations Green (2005) provided support for the similarity judgments by using what he referred to as structured analogies. This required forecasters to formally describe the similarities and differences between the past and future and then provide an overall similarity rating for each past analogous case. Arguably, the characteristics of conflicts are much less homogenous than those relating to the sales promotions experienced by a single company where variations in sales may be largely explained by a relatively small number of well defined variables. This suggests that for sales promotions that there may be benefits in designing a support system that automatically highlights the most similar cases in the database.

A number of potential biases may be associated with the final stage of FBA, the adaptation judgment. First, there is the issue of how much weight to give to the outcomes of the past cases to establish a starting estimate of the future outcome. These past outcomes are likely to differ from the future outcome because of noise and because of the imperfect match between the attributes of the past and future cases. If only a single past case is selected (Goodwin and Fildes, 1999) noise associated with its outcome may result in a poor starting estimate. If several past cases are selected (McIntyre, Achabal et al. 1993), the problem of noise may be reduced but the difficult judgment has to be made about the relative value of these cases to the forecast and some of the cases identified may have undue influence (Kolodner 1991). A simple heuristic would involve weighting the outcomes of the cases equally by taking their mean but this might lead to a starting estimate that was more distant from the future outcome if some of the cases are less similar than others. Having arrived at a starting estimate, it has to be adapted to take into account the perceived dissimilarities between the past cases and the conditions that apply to the future case. This is a difficult judgment since the effects of changes in the values of the independent variables (i.e. the promotion attributes) will need to be estimated. The unaided forecaster may make this estimate by comparing just one or two past cases and using mental arithmetic to estimate the effects. In addition, Tversky and Kahneman's (1974) work suggests that any estimated adaptations from past cases are likely to be insufficient since they will be biased towards the starting estimate which will act as an anchor. All of this suggests two possible support mechanisms for the adaptation judgment. First a ranking of several analogous past cases in terms of their similarity may help the forecaster to obtain a more reliable starting estimate of the promotion effects. Secondly, a device that allows the forecaster to estimate the effect of changes in the values of the independent variables, albeit from a small sample of past cases, may improve the reliability of the adaptation.

RESEARCH HYPOTHESES

The forgoing discussion suggests that the following hypotheses are appropriate.

H1: Providing support for similarity judgments in addition to memory support will lead to more accurate forecasts than providing memory support on its own.

H2: Providing adaptation + similarity + memory support will lead to more accurate forecasts than just providing similarity + memory support

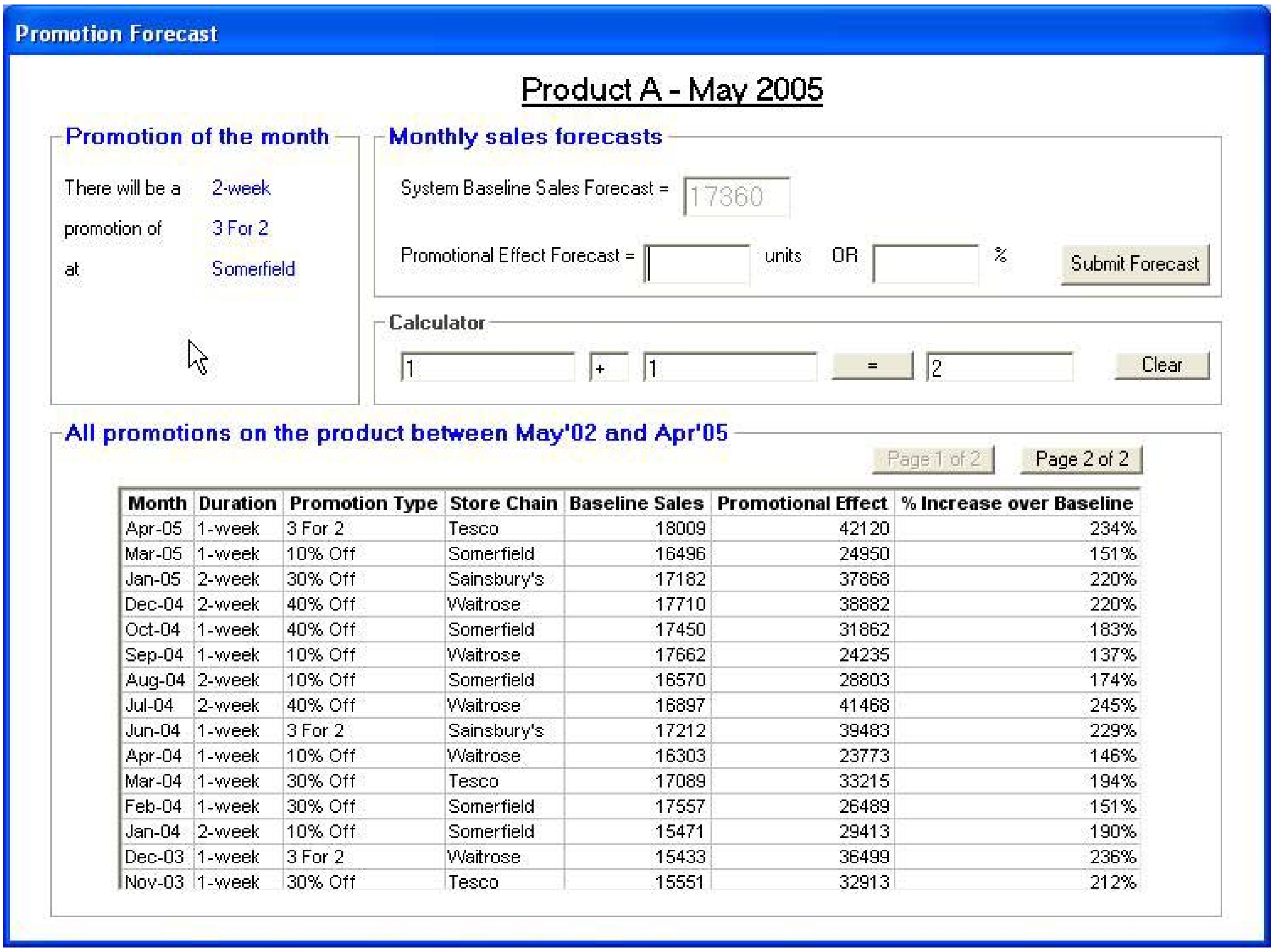
Clearly, the effectiveness of any form of support will be dependent on its acceptability to forecasters. To achieve acceptability the support is likely to need to be transparent, intuitively reasonable and to be seen as relevant to the task (Yates, Veinott et al. 2003). The perceived relevance to the task is likely to be greater if more stages in the cognitive process are supported. This leads to the following hypothesis.

H3: The greater the level of support provided then the more acceptable that support will be to forecasters

RESEARCH DESIGN

A laboratory experiment was conducted to test these hypotheses. It simulated a sales promotion forecasting task undertaken by manufacturers who distribute their products to supermarkets who are running promotion campaigns. The forecasts required were the estimated additional sales at the supermarkets in the month during which a promotion took place (i.e. the sales that were extra to the 'baseline' sales that would have applied had there not been a promotion). Real British store names were used for the supermarkets to increase the realism of the task. The forecasters were 54 undergraduate and postgraduate students from the Management schools at the Universities of Bath and Lancaster. Each participant received a £5 as a reward for taking part and the 50% of students who made the most accurate forecasts in their groups, were given an additional £5. Remus (1986) found students to be good proxies for real managers in experiments.

The participants produced their forecasts using a computerised forecasting support system (FSS) which had certain facilities available depending on the experimental treatment. Figure 1 shows an interface for one of the treatments. The following features were common to all treatments. First, for each forecast, there was a description of the forthcoming promotion. Second, the baseline 'forecast' for the product in the forecast period was presented and the participants were informed that their 'experience' suggested that this forecast was a reliable indicator of baseline sales. The third common feature was a pair of input textboxes where the participant entered his/her forecast for the additional sales resulting from the promotion. One of the input boxes was in units, and the other in percentages (i.e. the additional sales as a percentage of the baseline 'forecast'). The participants could enter their forecast in either format, and the corresponding value in the other format would be presented in the other textbox. This feature was suggested by several participants in a pilot study. The final common feature was a simple calculator that could handle addition, difference, multiplication and division. This was built in to the system so that participants' cognitive load was reduced, and because the forecasters at the companies we visited often used calculators along with their FSSs.



Month	Duration	Promotion Type	Store Chain	Baseline Sales	Promotional Effect	% Increase over Baseline
Apr-05	1-week	3 For 2	Tesco	18009	42120	234%
Mar-05	1-week	10% Off	Somerfield	16496	24950	151%
Jan-05	2-week	30% Off	Sainsbury's	17182	37868	220%
Dec-04	2-week	40% Off	Waitrose	17710	38882	220%
Oct-04	1-week	40% Off	Somerfield	17450	31862	183%
Sep-04	1-week	10% Off	Waitrose	17662	24235	137%
Aug-04	2-week	10% Off	Somerfield	16570	28803	174%
Jul-04	2-week	40% Off	Waitrose	16897	41468	245%
Jun-04	1-week	3 For 2	Sainsbury's	17212	39483	229%
Apr-04	1-week	10% Off	Waitrose	16303	23773	146%
Mar-04	1-week	30% Off	Tesco	17089	33215	194%
Feb-04	1-week	30% Off	Somerfield	17557	26489	151%
Jan-04	2-week	10% Off	Somerfield	15471	29413	190%
Dec-03	1-week	3 For 2	Waitrose	15433	36499	236%
Nov-03	1-week	30% Off	Tesco	15551	32913	212%

Figure 1 Interface for memory support (level 1)

The experiment was a 3 (support level) × 2 (noise level) × 2 (promotional model type) factorial design. The subjects were randomly assigned to one of the three support levels. The noise and promotion-type factors were within-subject variables.

Support Levels Subjects in Level 1 received only memory support in the form of a database of 30 past promotions for a given product (as in Hoch and Schkade's (1996) study) (see figure 1). This was presented in two 'pages' and provided information on the month when the promotion took place, its duration and type, the store that was running it, the estimated baseline sales for the month when it took place and the estimated extra sales that it achieved expressed both as an absolute value and as a percentage of the baseline sales.

Level 2 subjects received support for both memory and similarity judgments. The interface was the same as that for Level 1, but additionally included the presentation of a table giving details of the three promotions in the database that were most similar to the promotion for which a prediction was required (see figure 2). These were presented in order of similarity.

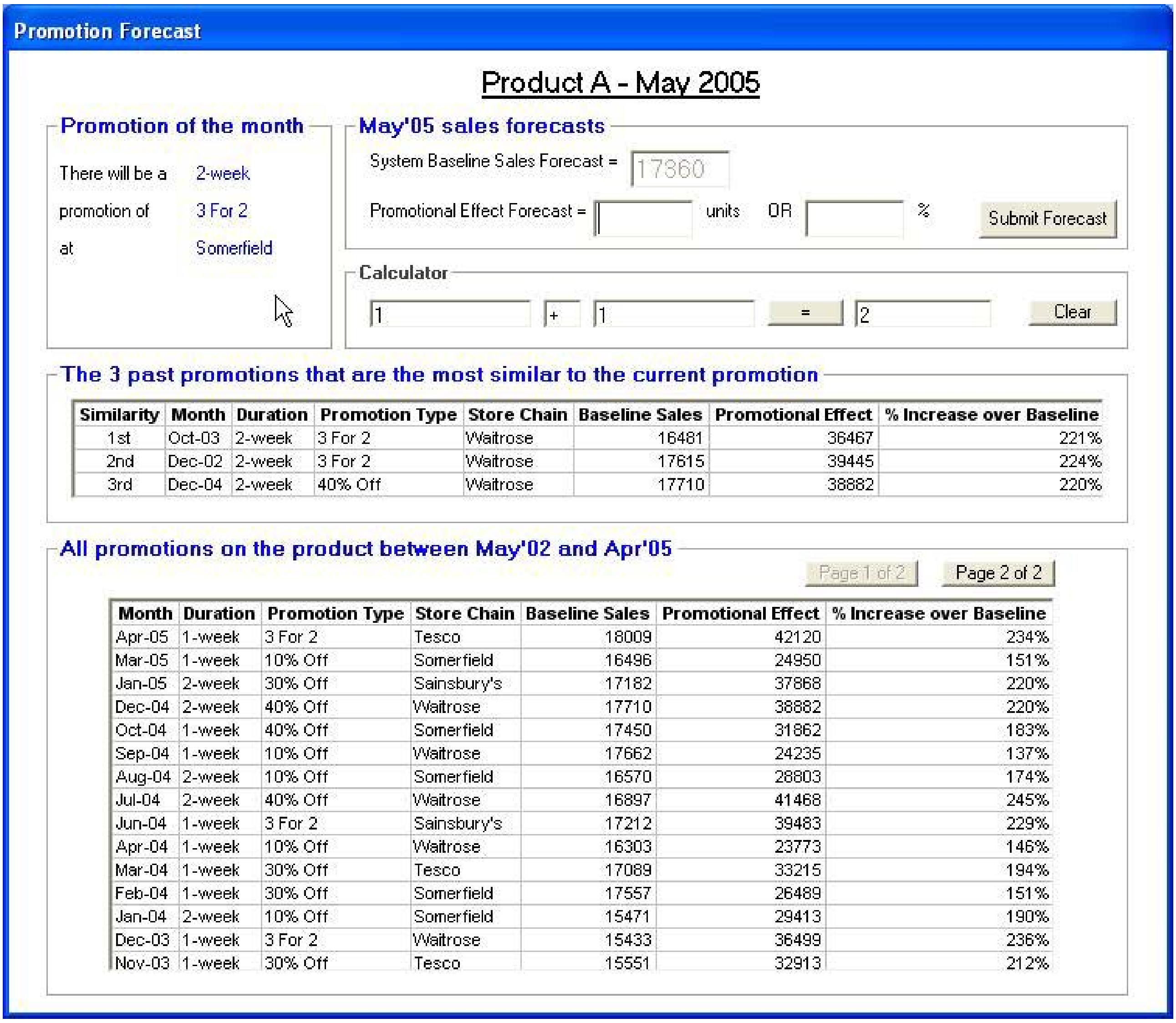


Figure 2 Interface for memory & similarity support (level 2)

To identify the most similar promotions, in other words the past cases that were most likely to have the closest promotion effects, a hierarchical rule system was applied. First of all, the number of matching promotion attributes (i.e. duration, promotion type and store) was established between each past case and the target case (i.e. the upcoming promotion).

Similar attribute alternatives, such as 'Tesco' and 'Sainsbury's' for store, were used as proxies for each other. The greater the number of matching attributes, the more similar the cases were seen to be. When there were ties in this number more weight was placed in the selection of cases on the similarity of the promotion type, followed by similarity of store and then similarity of duration. This was because promotion type had the most varied alternatives (see Table 1), hence was more likely to be the key to distinguish the similarity level between cases. For instance, when two of the three attributes of several cases were the same as those of a target case, those with the same promotion type and store would usually be selected over those with the same promotion type and duration, followed by those with the same store and duration. Of course, because of noise the most similar case might not turn out to have as close a promotion effect to the forecast promotion as say the second or third most similar case.

Promotion attributes	**Attribute alternatives**	**Multiplicative model effects**	**Additive model effects (sales units)**
Duration (D)	1-week	0.25	4532
	2-week	0.375	6800
Type of promotion (T)	10% Off	1.6 [= ln5]	4791
	30% Off	2.3 [= ln15]	6888
	40% Off	3.0 [= ln20]	8984
	3 For 2	3.5	10481
	Buy One Get One Free	4	11979
Store (S)	Somerfield	1.4	2241
	Waitrose	1.5	2401
	Sainsbury's	2.5	4003
	Tesco	2.6	4163

Table 1 – Effects used in promotion generating models

Level 3 subjects received support for memory, similarity judgments *and* adaptation judgments. To provide the adaptation support the computer searched for pairs of cases in the relevant product's database that differed in only one aspect (e.g. duration). For each pair it calculated the ratio of the percentage promotion effects of the two cases. Finally, it calculated the mean ratio for all pairs of promotions that just differed on this aspect. For example, it might find that for a given product, on average, all else remaining equal, 2-week promotions had an effect 1.5 times greater than 1-week promotions. The user could then use a simple interface (see figure 3) to interrogate the computer to discover these ratios. This support was designed to be easily understood in terms of both how the values were derived and how they might be applied, so that the feature would be acceptable to users.

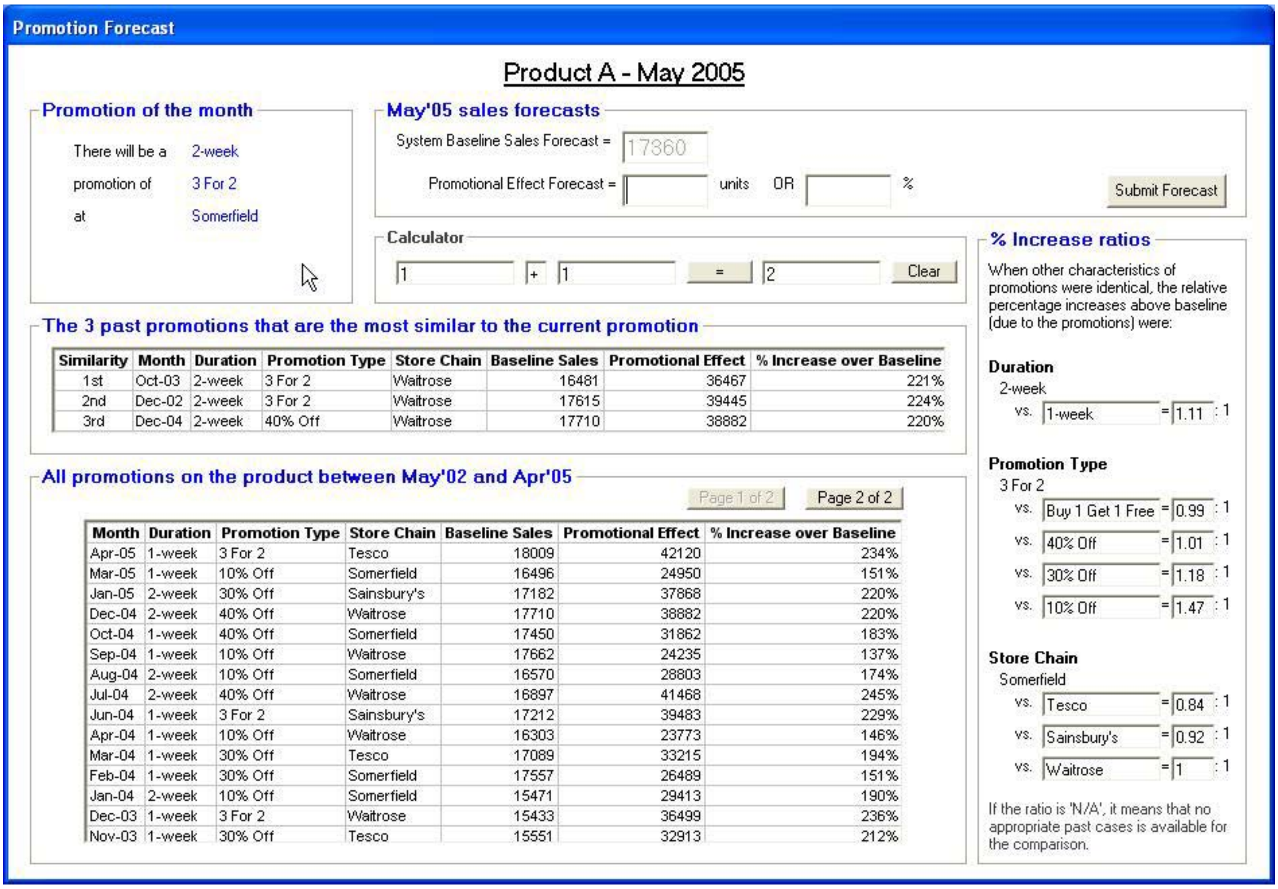


Figure 3 Interface for memory, similarity & adaptation support (level 3)

Noise levels One of two levels of noise was added to the promotion effects (recall that Hoch and Schkade (1996) found that noise level was a determinant of the effectiveness of the support that they provided). The expected absolute noise (or mean absolute deviation) for low-noise and high-noise products was 5% and 15% respectively. A procedure described by Goodwin and Fildes (1999) was used to generate the noise.

Promotional models. The evidence and theories of the effects of sales promotions in the marketing literature are sketchy and conflicting. Some are at conceptual levels, hence no mathematical formulae are suggested (e.g. Dommermuth 1989; Grewal, Krishnan et al. 1998; Laroche, Pons et al. 2003). Others propose certain promotional models but they are too complex or specific and require too much data or too many assumptions (e.g. McIntyre, Achabal et al. 1993; Cooper, Baron et al. 1999) to be applied in this study. Therefore, the independent variables in the promotion mix were based on some company data and relevant papers in the marketing literature (e.g. Raghubir, Inman et al. 2004). In order to broaden the extent to which the findings could be generalised, the promotion effects were generated for individual products from the independent variables using either an additive or a multiplicative model. The key difference between the two models is that the promotion attributes of the latter interact with one another, for instance, the effect of a promotion being 'Buy One Get One Free' is magnified more by being at Tesco than at Somerfield. The basic structures of the models are shown below:

Multiplicative models:	$E = [A \times D \times (S \times T - 1)] + \varepsilon$
Additive models:	$E = [(D + S + T)/8000 \times A] + \varepsilon$

where:

E	Promotion effect
A	Average sales, i.e. the stem sales level from which baseline sales were derived, so that baseline sales = average sales + noise
D	Effect of number of weeks in promotion duration
S	Effect of Store
T	Effect of Promotion type
ε	Noise on promotion effects

The stem sales levels of the products were determined by first, generating 16 random numbers from a normal distribution with a mean of 0 and standard deviation of 1. These values were than multiplied by 10000 and added to 30000 so that no stem value was negative. These steps ensured that the stem sales levels were similar to the sales levels of the products of the companies visited by the research team. Table 1 shows the values of D, S and T that were used in the models. The relative effects between alternatives of a promotion attribute were the same regardless of the promotional model type. For instance, in the case of store, Tesco and

Sainsbury's had a similar effect for a given promotion in both multiplicative and additive promotional models. The effects for the additive model shown in Table 1 are based on an average level of sales (i.e. A) of 8000. Hence, for a product with different average sales, they were adjusted so that the promotion effects remained proportionate to the average sales.

Experimental procedure Before carrying out any forecasting task, the participants were asked to read an introductory document to familiarise themselves with their roles and the assumptions used in the experiment. This contained a description of the relative effects of the different promotion attributes. With the help of this information, the relative importance of the attributes in the promotion mix (e.g. promotion type was more influential) should have been derivable from the database. This ensured that even the participants receiving the level 1 treatment had the opportunity to assess the similarity of past promotions to the target case. After a practice session that involved making forecasts for four trial products, the forecasters proceeded to the main experiment which involved making forecasts of the promotion effects for 12 products. Six of the products involved effects generated by the additive model, while six had effects generated by the multiplicative model. For each model there was an equal split between products which had effects subject to high noise and those subject to low noise.

On completion of the forecasting tasks, participants responded to a questionnaire that was designed to obtain a more in-depth understanding of their reasoning processes and their attitude to the support they received. The first four questions, which required participants to rate the program on four 7-level scales, were adopted from Yang and Yoo (2004) who found cognitive attitude measures to be reliable predictors to the usage of an information system. All four semantic pairs were based on the seven cognitive word pairs proposed by Crites et al (1994). Yang and Yoo only applied the first three in their study; the fourth pair was added in this research because they also observed that 'perceived usefulness' was another good predictor of information system use. These were:

1. I think that the program is a **wise ↔ foolish** instrument in performing the task

2. I think that the program is a **beneficial** ↔ **harmful** instrument in performing the task
3. I think that the program is a **valuable** ↔ **worthless** instrument in performing the task
4. I think that the program is very **useful** ↔ **useless** to the task

There were also three open questions in the questionnaire: one related to the strategy the forecaster employed to arrive at the forecasts, the other two asked whether there was anything about the program the forecaster found useful or otherwise. They were:

5. Did you have a strategy/method to make your forecasts? If so, what was it and why? A detail answer to this question would be extremely useful to my analysis.
6. What features of the program did you find obstructive/annoying in carrying out your task?
7. What features of the program did you find useful in carrying out your task?

The whole process took around 40 minutes. On some occasions, further clarifications from participants were requested either verbally or immediately after their completion of the questionnaires or via emails.

RESULTS AND ANALYSIS

Forecast Accuracy Subjects' performance was assessed using the median absolute percentage error (MdAPE) which is suitable for comparing forecasting methods across different time series. Also, unlike the mean absolute percentage error (MAPE) this measure is not distorted by extreme percentage errors. As the participants were not expected to forecast noise (Harvey and Bolger 1996; Goodwin and Fildes 1999), the MdAPE was computed by comparing forecasts with signals (i.e. the effect of the promotion that was being predicted, excluding noise (ε)).

The results of 6 participants were discounted for various reasons. For example, two participants conferred during the experiment, while another participant's forecasts constituted an extreme outlier which was likely to be a result of incorrect entry. As a result, there were 17 sets of observations for Level 1, 15 for Level 2 and 16 for Level 3.

When all the products were considered, that is regardless of the promotional model type and the noise level, it was found to be appropriate to assume that the MdAPEs of the three supports had homogenous variances (Levene's Test: $p = 0.870$). One-way ANOVA was therefore applied to test for the differences between the three levels of support. This revealed that there was a significant difference between the three levels ($p=0.025$). Multiple comparisons using the Scheffé Test were applied to determine the differences between the supports. The Scheffé Test is robust for unequal numbers of observations among groups (Bryman and Cramer 2001). It showed that, while Level 3 was significantly more accurate than Level 1; Level 2 did not lead to significant improvements over Level 1. H1 was therefore rejected, but there was partial support for H2 in that memory + similarity + adaptation support led to significantly greater accuracy than memory support alone ($p = 0.026$). Indeed level 3 support led to the lowest mean APE (with the mean taken across subjects for a given product) for 8 of the 12 products.

To what extent was the performance of the three levels of support conditional on the nature of the promotion effect (i.e., whether it was generated using an additive or multiplicative model and whether it was subject to low or high noise)? Further analysis revealed that that there was a significant interaction between Model, Noise and Level of support ($p=0.001$). Post-hoc pairwise comparisons showed that Level 3 support led to forecasts that were significantly more accurate than those produced under the other two levels when the promotion effect was either i) additive and subject to high noise (level 3 v level 1, $p <0.001$, level 3 v level 2, $p = 0.041$) or ii) multiplicative and subject to low noise (level 3 v level 1, $p <0.001$, level 3 v level 2, $p = 0.015$). For products subject to the other promotion effects no significant differences were found between the levels of support. Thus providing memory + similarity + adaptation support appears to carry no risks in that it will not harm forecasting accuracy and under some conditions it appears to produce significantly more accurate forecasts.

User acceptance Subjects' responses to the four rating questions in the post-experiment questionnaire were used to test H3 that greater level of support would increase the likelihood that the support system would be accepted by users. As the responses were highly skewed, even after logarithmic transformation, the nonparametric Kruskal-Wallis Test, was applied to see if their were significant differences in the attitudes of respondents depending on the level of support they received. Significant differences (at $p < 0.05$) were found for all four ratings. The median test revealed that a higher proportion of participants receiving Level 3 support rated their support as being wiser and more valuable. Relatively more Level 2 participants rated their support less beneficial and valuable than those of the other two supports. No significant correlation was found between forecast accuracy and attitude ratings. The findings therefore only partly support H3 in that the higher cognitive attitude scores of Level 3 indicated that it was more likely to be acceptable than the other two support levels. However, Level 2 participants found that their support even less acceptable than those of Mode 1. Nevertheless, it should be noted that the median ratings for the different attributes were relatively high for all the levels of support suggesting that none of them would be unacceptable to potential users.

DISCUSSION

Four main results emerge from this experiment. First, providing memory + similarity + adaptation (level 3) support for the use of analogies appears, in general, to be beneficial to forecast accuracy. Second, providing only memory + similarity (level 2) support does not appear to be beneficial. Third, the effectiveness of level 3 support appears to be conditional on the nature of the promotion effect. Finally, level 3 support appears to be highly acceptable to forecasters.

Why did providing memory + similarity support not lead to any incremental benefits over providing memory support alone? The most likely explanation is that many of the subjects did not use the similarity support in the way in which it was designed to be used, that is, to only consider the three listed cases as the starting point for their forecasts. Indeed it appears from the responses to the post-experiment questionnaire that over half of the Level 2 subjects were in this category. For example, five subjects established criteria which determined whether the cases in the similarity table should be used, such as only cases with exactly the same store as the target case were considered. This strategy appears to be similar to the lexicographic strategy employed in multiattribute decision making in that the user's focus was on a single attribute of the promotion (Goodwin and Wright 1998).

These deviations from the intended usage of the support might have resulted from a belief that other information was also relevant (other than those provided by the 3-case table), and/or the belief that the most similar cases chosen were not as similar as they claimed to be. Indeed, the order of similarity presented in the table was widely ignored or little reliance was placed on it. The support system assessed the similarity of the cases on the basis of how the promotion attributes related to the resulting promotion effect. Many of the subjects appear to have assessed similarity on the basis of the attributes alone. For example, while participants may have perceived a 2-week '30% Off' promotion at Tesco as more similar to 1-week '30% Off' promotion at Tesco, the support system regarded a 1-week '40% Off' promotion at Sainsbury's to be more similar. This difference in similarity judgment might be because the participants believed that store and promotion type were more important attributes than duration, and that exact matches were required. Alternatively, the participants might have perceived duration as a characteristic that was easier to adapt than store or promotion type, as there were only two alternatives, as opposed to four or five. Most of the Level 3 participants also did not use the similarity support as intended so the presence of adaptation support (ratios) probably accounted for most of the benefits of this support level, though there was also the possibility of synergistic effects arising from the similarity and adaptation support.

These findings are consistent with the views of Yates et al (2003) who suggested that people would not rely on an aid if the information provided to them was not seen as relevant to the decision they are making This suggests that it would be worth investigating a more interactive similarity judgment support that was also more transparent, comprehensible and flexible. For example, this might include a facility to enable users to sort cases in the database. Subjects therefore could search for the most similar cases in terms of their own definitions of similarity, with their cognitive effort supported by the system.

Finally, why was the effectiveness of Level 3 support contingent on the nature of the promotion effect? This may have been a combination of task difficulty and the accuracy of the ratios that were provided to support the adaptation phase – this accuracy was affected by both the noise level and the type of promotional model. The mean MdAPEs achieved by subjects who received only memory support (level 1) (see
Table 2) suggest that the forecasting difficulty of the promotion effects can be ranked from the easiest to the most difficult as follows: 1) additive + low noise, 2) additive + high noise, 3) multiplicative + low noise and 4) multiplicative + high noise. In the easiest tasks there was less scope for improvements and any potentially small advantage of the support was likely to be negated by the (relatively small) sampling errors associated with the ratio estimates. For the most difficult tasks (multiplicative + high noise) the ratios provided were relatively inaccurate estimates of the true ratios so the potential benefits of the support were lost. Thus the support was effective in the two 'medium difficulty' tasks where there were substantial potential improvements to be gained that were not undermined by inaccurate ratio estimation.

Type of promotional model	Additive	Additive	Multiplicative	Multiplicative
Noise level	**Low**	**High**	**Low**	**High**
Means of level 1 participants' MdAPEs	9.83%	15.97%	20.56%	27.33%

Table 2 – Level 1 mean MdAPEs according to the nature of the promotion effect

CONCLUSIONS

The use of analogies to forecast the effects of special events appears to be a natural and widespread phenomenon in company forecasting, but to date there has been an absence of studies in the literature which have investigated how the approach might be supported through appropriate software design. The experiment reported in this paper has shown that a simple support method can help to improve the accuracy of forecasts under some conditions, without carrying any associated risk that it will lead to poorer forecasts in others. The support was not only intuitive and transparent, and hence likely to be acceptable to forecasters, but it was also easy to create. Our results suggest that it may well be worth including the method in future forecasting software. Indeed, most current software systems used in companies provides little incentive for the creation of a centralised record of past special events and their results. Many packages do include note functions where users may input details of the reasons for adjusting a certain forecast. However, these functions are often merely textboxes where no structure is provided. This makes the search for the details of past events cumbersome.

Of course, the results reported here are subject to a number of caveats. Simulated promotion effects were used in order to allow the experiment to be run under controlled conditions where the true values of the underlying 'signal' were known. The approach clearly needs to be tested in real time on real data in commercial environments. Further research is also need to investigate potential ways in which the support may be developed and enhanced. For instance, would allowing users to select the most similar cases, albeit with some guidance provided by the system, be more effective than providing them automatically? Would information on the reliability of the ratios provided for adaptation support encourage users to identify conditions where the use of the ratios is more likely to be beneficial? Future studies will address these questions and will also assess the benefits of the approach in practical contexts.

References

Bryman, A. and D. Cramer (2001). Quantitative Data Analysis with SPSS Release 10 for Windows. Hove, Routledge.

Cooper, L. G., P. Baron, et al. (1999). PromoCast Trademark: A new forecasting method for promotion planning. Marketing Science, INFORMS: Institute for Operations Research. **18:** 301.

Crites, S. L., L. R. Fabrigar, et al. (1994). "Measuring the Affective and Cognitive Properties of Attitudes - Conceptual and Methodological Issues." Personality and Social Psychology Bulletin **20**(6): 619-634.

Dommermuth, W. P. (1989). Promotion analysis, creativity and strategy. Boston, PWS-Kent.

Flores, B. E., D. L. Olson, et al. (1992). "Judgmental adjustments of forecasts: A comparison of methods." International Journal of Forecasting
International Journal of Forecasting J1 - International Journal of Forecasting **7**(4): 421-433.

Gentner, D. (1989). Finding the needle: Accessing and reasoning from prior cases. Proceedings of the DARPA Workshop on Case-Based Reasoning. K. J. Hammond. San Mateo, California, Morgan Kaufmann. **2:** 137-143.

Gilovich, T. (1981). "Seeing the Past in the Present - the Effect of Associations to Familiar Events on Judgments and Decisions." Journal of Personality and Social Psychology **40**(5): 797-808.

Goodwin, P. and R. Fildes (1999). "Judgmental forecasts of times series affected by special events: does providing a statistical forecast improve accuracy?" Journal of Behavioral Decision Making **12**: 37-53.

Goodwin, P. and G. Wright (1998). Decision Analysis for Management Judgment. Chichester, Wiley.

Green, K. C. (2005). Role thinking: Does standing in the other guy's shoes improve forecast accuracy? 25th International Symposium on Forecasting, San Antonio, Texas.

Grewal, D., R. Krishnan, et al. (1998). The effect of store name, brand name and price discounts on consumers' evaluations and purchase... Journal of Retailing, Elsevier Science Publishing Company, Inc. **74:** 331.

Harvey, N. and F. Bolger (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. International Journal of Forecasting. **12:** 119.

Hoch, S. J. and D. A. Schkade (1996). A psychological approach to decision support systems. Management Science, INFORMS: Institute for Operations Research. **42:** 51.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. The Psychology of Learning and Motivation. G. H. Bower. Orlando, Florida, Academic Press**:** 59-87.

Klein, G. A. and R. Calderwood (1988). How do people use analogues to make decisions? Proceedings of a Workshop on Case-Based Reasoning: Holiday Inn, Clearwater Beach, Florida, May 10-13, 1988.

Kolodner, J. L. (1991). "Improving human decision making through case-based decision aiding." AI Magazine **12**(2): 52-68.

Kolodner, J. L. (1993). Case-Based Reasoning. San Mateo, California, Morgan Kaufmann.

Laroche, M., F. Pons, et al. (2003). A model of consumer response to two retail sales promotion techniques. Journal of Business Research, Elsevier Science Publishing Company, Inc. **56:** 513.
Lawrence, M., P. Goodwin, et al. (2002). Influence of user participation on DSS use and decision accuracy. Omega (Oxford). **30:** 381.
Lindsay, P. H. and D. A. Norman (1977). Human Information Processing: An Introduction to Psychology. London, Academic Press.
McIntyre, S. H., D. D. Achabal, et al. (1993). Applying Case-Based Reasoning to Forecasting Retail Sales. Journal of Retailing, Elsevier Science Publishing Company, Inc. **69:** 372.
Nikolopoulos, K., P. Goodwin, et al. (2005). "Forecasting with cue information: a comparison of mulitple regression with alternative forecasting approaches." Lancaster University working paper.
Raghubir, P., J. J. Inman, et al. (2004). The Three Faces of Consumer Promotions. California Management Review, California Management Review. **46:** 23-42.
Remus, W. (1986). Graduate Students as Surrogates for Managers in Experiments on Business Decision Making. Journal of Business Research, Elsevier Science Publishing Company, Inc. **14:** 19-25.
Schank, R. C. (1982). Dynamic Theory: A Theory of Reminding and Learning in Computers and People. Cambridge, Cambridge University Press.
Taylor, P. F. and M. E. Thomas (1982). "Short term forecasting : horses for courses." Journal of the Operational Research Society **33**: 685-694.
Webby, R. (1993). Graphical Support for the Integration of Event Information into Time Series Forecasting: An Empirical Investigation. School of Information Systems. Sydney, University of New South Wales**:** 273.
Willemain, T. R. (1989). "Graphical adjustment of statistical forecasts." International Journal of Forecasting **5**(2): 179-185.
Willemain, T. R. (1991). "The effect of graphical adjustment on forecast accuracy." International Journal of Forecasting **7**(2): 151-154.
Wolfe, C. and B. Flores (1990). "Judgemental Adjustment of Earnings Forecasts." Journal of Forecasting
Journal of Forecasting J1 - Journal of Forecasting **9**(4): 389-405.
Yang, H.-d. and Y. Yoo (2004). "It's all about attitude: revisiting the technology acceptance model." Decision Support Systems **38**(1): 19-31.
Yates, J. F., E. S. Veinott, et al. (2003). Hard decisions, bad decisions: on decision quality and decision aiding. Emerging Perspectives on Judgment and Decision Research. J. Shanteau. New York, Cambridge University Press: 13-63.

University of Bath School of Management
Working Paper Series
Past Papers

School of Management
Claverton Down
Bath
BA2 7AY
United Kingdom
Tel: +44 1225 826742
Fax: +44 1225 826473
http://www.bath.ac.uk/management/research/papers.htm

2004		
2004.01	Stephan C. M. Henneberg	Political Marketing Theory: Hendiadyoin or Oxymoron
2004.02	Yi-Ling Chen & Stephan C. Henneberg	Political Pulling Power. Celebrity Political Endorsement and Campaign Management for the Taipei City Councillor Election 2002
2004.03	Stephan C. Henneberg , Stefanos Mouzas & Pete Naudé	Network Pictures – A Concept of Managers' Cognitive Maps in Networks
2004.04	Peter Reason	Education for Ecology: Science, Aesthetics, Spirit and Ceremony
2004.05	Yvonne Ward & Andrew Graves	A New Cost Management & Accounting Approach For Lean Enterprises
2004.06	Jing Lin Duanmu & Felicia Fai	Assessing the context, nature, and extent of MNEs' Backward knowledge transfer to Chinese suppliers
2004.07	Richard Fairchild	The effect of product differentiation on strategic financing decisions.
2004.08	Richard Fairchild	Behavioral Finance in a Principal-agent Model of Capital Budgeting
2004.09	Patrick Stacey & Joe Nandhakumar	Managing the Development Process in a Games Factory: A Temporal Perspective
2004.10	Stephan C. M. Henneberg	Operationalising a Multi-faceted Concept of Strategic Postures of Political Marketing
2004.11	Felicia Fai	Technological Diversification, its Relation to Product Diversification and the Organisation of the Firm.
2004.12	Richard Fairchild and Ganggang Zhang	Investor Irrationality and Optimal Open-market Share Repurchasing
2004.13	Bruce A. Rayton and Suwina Cheng	Corporate governance in the United Kingdom: changes to the regulatory template and company practice from 1998-2002

2004.14	Bruce A. Rayton	Examining the interconnection of job satisfaction and organizational commitment: An application of the bivariate probit model