ELSEVIER

The effects of structural characteristics of explanations on use of a DSS

M. Sinan Gönül [a], Dilek Önkal [a,*], Michael Lawrence [b]

[a] Faculty of Business Administration, Bilkent University, 06800 Ankara, Turkey
[b] School of Information Systems, University of New South Wales, Sydney 2052, Australia

Abstract

Research in the field of expert systems has shown that providing supporting explanations may influence effective use of system developed advice. However, despite many studies showing the less than optimal use made of DSS prepared advice, almost no research has been undertaken to study if the provision of explanations enhances the users' ability to wisely accept DSS advice. This study outlines an experiment to examine the effects of structural characteristics of explanations provided within a forecasting DSS context. In particular, the effects of explanation length (short vs. long) and the conveyed confidence level (weak vs. strong confidence) are examined. Strongly confident and long explanations are found to be more effective in participants' acceptance of interval forecasts. In addition, explanations with higher information value are more effective than those with low information value and thus are persuasive tools in the presentation of advice to users.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Explanation; Forecast; Judgment; Adjustment

1. Introduction

The last few decades have witnessed a significant increase in the availability and accessibility of information providers that target the decision makers in a variety of fields ranging from medical to financial sectors [5,14]. Decision makers routinely seek various forms of external information assistance to support their decision making processes with forecasts constituting one of the most widely used forms of such external assistance. However, decision makers choose to trust and use these forecasts only if they believe these predictions are justifiable, relevant, and valuable in effectively manag-

ing the uncertainties about the future [8]. What makes an individual use an external forecast, then, is a direct function of his/her perceptions of its acceptability. A provided forecast may be considered accurate, justifiable and informative from a provider's perspective; however, its utilization is totally dependent on whether the user is persuaded that this is the case.

This acceptability issue is a major concern especially for developers of decision support systems (DSS). What makes a DSS successful is not simply the accuracy of its results, but the acceptance of these results by its users. However, research evidence suggests reluctance by decision makers to trust the advice provided by a DSS [2,17,22]. In the field of sales forecasting, despite ready availability of excellent software, surveys continue to show that many organisations develop their forecasts using only management judgement [32], and when

---

* Corresponding author. Tel.: +90 312 290 1596; fax: +90 312 266 4958.

E-mail address: onkal@bilkent.edu.tr (D. Önkal).

decision support software is used, the advice is frequently adjusted to produce a lower quality decision than if no adjustment was made [6]. Davis and Kotteman [2] in summarising the literature concluded, "despite the broad effectiveness of decision rules, decision makers have been notoriously reluctant to use them" [pg 145]. This reluctance to make use of a DSS does not seem to stem from lack of familiarity with computer systems. Davis and Kotteman [2] in an experiment using a production planning DSS and Lim and O'Connor [21], using a forecasting DSS showed that students trained in computer use and playing the role of subjects, continued to ignore the good advice given by the DSS even when they could see the DSS was outperforming their own efforts and when real money was being paid for performance. It seems clear then that a key element in DSS design is the need to focus on effectively communicating to users the quality of decision support provided.

Provision of an explanation may constitute one of the most effective methods for improving the acceptance of forecasts or other decision support advice [19]. However, not all types of explanations may be equally effective in persuading the user that the advice is acceptable: the underlying structural characteristics of explanations could be directly related to their persuasive effectiveness. The current study focuses on exploring the effects of structural characteristics of explanations on the acceptance of DSS advice. In particular, the effects of explanation length and the effects of the level of confidence conveyed in the explanation are investigated via an experimental study. The study involves presenting graphs of time series data along with DSS provided advice in the form of point and interval forecasts, accompanied by explanations that differ in their length and conveyed confidence. The subjects are asked to make a "final forecast" using the presented cues. The level of acceptance of the DSS advice is assumed to be directly related to the difference between the provided forecast and the subject's final forecast. We reason that decision makers accept the DSS advice if they believe it accurate and reliable but will adjust away from it if they perceive it as unacceptable. Therefore, an explanation that is more effective and persuasive will lead to a lower propensity to adjust the provided forecast, while the situation will be the reverse for a less acceptable/effective explanation.

This paper concludes that explanations can play a role in influencing a user's acceptance of the DSS advice and suggests, for a forecasting application, what structure of explanations seems most appropriate. Given the evidence of low acceptance of DSS advice, this is an important contribution suggesting that explanations may be a means of improving DSS success.

The organization of this paper is as follows. In the next section, a brief literature review is provided, along with the implications for current research. Section 3 details the research methodology used in the experiment with results of statistical analysis presented in Section 4. Discussion and limitations are outlined in Section 5. We conclude by providing implications of the findings and directions for future research.

2. Literature review and implications for current research

The development of explanation tools has attracted attention since the introduction of MYCIN—a medical expert system incorporating an explanation facility developed during the 1970s [28]. However this explanation research has had very little impact within the DSS field where explanations are almost unknown. In this section, a brief overview of explanation research provides answers to questions such as what are the general types and use of explanations in the wider class of intelligent systems that may apply to a DSS and why are explanations provided. The term 'intelligent system' is generally applied to the class of systems providing advice to a decision maker and includes Expert Systems and DSS (for a complete review of related literature, please see Refs. [3,10,34]). Almost all the research in explanations has taken place within the realm of expert systems, where these explanations serve to communicate the rationale underlying the advice. Expert Systems (ES) differ from DSS in a number of ways[1] (see Ref. [30] for a full coverage of the differences), although both are built to assist a decision maker by providing decision advice.

2.1. Classes of explanations

Within the intelligent system domain, explanations have been classified along the three dimensions of (i) the content of explanations, (ii) the provision mechanism of explanations, and (iii) the presentation format of explanations [3,10,16,24].

Within dimension (i), the content of explanations, there are three main categories. The first category can be called *trace* or *line of reasoning* explanations. This kind of explanation provides a perspective about the internal

---

[1] A key distinction is that an ES manipulates a knowledge base to derive managerial advice while a DSS typically manipulates a model to develop the advice.

outcome generation procedure that is used by the decision support system. It presents insight about the information that is utilized and the rules, processes or steps that are used by the system to reach the recommendation or the outcome it has generated for a particular case. Thus a trace explanation usually answers the question "how". Moreover, they are the simplest to design, since they can be generated as a by-product in the system during the generation of outcome; these make them the most commonly used explanation type [24].

The second category of explanations within dimension (i) is the *justification* explanations. These explanations often tap into the deep knowledge that is behind a generated outcome. They justify the outcome or recommendation they have given by providing the theory or causal model through which that outcome was generated. They provide the rationale behind a decision. In this respect, they generally answer the question "why". Justification type explanations have been found to be the most effective in increasing the acceptance of conclusions derived by an ES, when compared to the other types of explanations [34].

The last category of explanations within the content dimension is the *strategic* explanation. These explanations provide an insight about the overall problem-solving strategy used by the system. They try to form a global perspective on the operation of the system by tapping into the planning, strategy and higher level goals in the system [16]. This kind of explanation is less commonly used, due to the difficulties in its generation and presentation.

The second dimension of classification of explanations describes the provision mechanism of explanations; this involves how the users of a system can access the explanation. Explanations can be *user-invoked*, *automatic*, or *intelligent*. The *user-invoked* explanations are accessed entirely at the request of the user. Accessing an explanation may involve pressing a key, writing a command, a hypertext or any other available mechanism. In contrast, *automatic* explanations are provided to the user at all times and, naturally, the user has no control over this type of provision mechanism. Finally, the *intelligent* explanations are often provided to the users whenever the system judges that there appears to be a need. These explanations are often complex and may require complicated AI decisions.

The third dimension, the presentation format of an explanation refers to how an explanation is provided to the user. In general, it can be either text-based or in a multimedia format. The text-based explanation is the simplest form of presentation. A decision support system can provide either predetermined sentences (i.e., "canned" text) as explanation or it can combine a few sentences according to certain integrated rules to form an explanation. For the latter case, it has been found that sentences that resemble natural language are effective in increasing the transparency of a system's evaluation, leading to better perceptions about the system [26]. The multimedia type of explanation involves more than simple text, making use of graphs, pictures, and animations if applicable. This type of explanation is harder to develop, and can be expensive. In return, it may bring more persuasive power to the system and increase confidence in the output of the system.

2.2. The role of explanations

An explanation can assist in meeting three possible goals—(i) to explain a perceived anomaly, (ii) to supply extra knowledge, and (iii) to facilitate learning from the system. When people encounter a perceived anomaly in the advice they are given, they typically demand an explanation. Providing an explanation may help users understand the nonconforming advice. It can either verify that the advice does in fact match the users' expectations, or it may try to resolve the contradiction between the users and the system [24]. Another reason for requesting an explanation may be the users' need for extra knowledge that will enable them to participate effectively in a problem-solving task. Individuals often require additional and prompt information when they are solving a problem, and an explanation facility may provide this effectively. It has been found that there is a positive relationship between explanation use and performance in a cooperative problem-solving task [9]. For a problem-solving task, the need for information is generally short term. On the other hand, there may be long-term learning-related needs. Users of a decision support system may be able to learn about the DSS and the problem domain from the information provided to them and so enhance their effectiveness and efficiency in future tasks. In this regard, users may request an explanation to facilitate learning from the system. Specifically, Mao and Benbasat [24] and Papamichail and French [26] argue that the cognitive effort perspective suggests that people will request an explanation as long as the benefit gained from the explanation outweighs the mental effort spent in the absence of it.

2.3. The value of explanations

In examining the implications of explanations for intelligent systems, Dhaliwal and Benbasat [3] provide a

strong framework for the effects of explanation use based on cognitive learning theory. In a broad sense, it is suggested that the use of explanations often leads to learning, which brings increased performance and better user perceptions of the system, resulting in the acceptance of the system. Improved understanding and learning are argued to have two important implications. The first is the improvement in decision making performance, in terms of both accuracy and speed. The second implication is the improvement in the perceptions of the user toward the system, which is directly related to intended future use. This improvement involves the improvements in perceptions of ease of use, of usefulness, of satisfaction, and of trust. Dhaliwal and Benbasat suggest that these modified perceptions coupled with the increase in performance will eventually lead to acceptance of the system. Another cited effect of explanation use stresses the increased transparency of the system [16,26]. The explanations (especially the *traces*) provide information about how the system operates and how it generates an outcome. This increase in visibility facilitates the participants' acceptance of the system as being logical, and thus, acceptance of its advice as being justified and useful.

Research on advice taking identifies three main reasons for accepting a piece of advice [12]. First, decision makers wish to improve their judgments through using the advice provided. The advisors are generally chosen so that they have more knowledge and expertise than the decision maker in a particular area of interest. Therefore, the decision makers assume that the advice is provided to them by a comparatively more knowledgeable and experienced person so that using such advice will improve the accuracy of their own judgments. Second, the decision makers try to share the responsibility of a decision with someone else, especially when the risk associated with an error is high; the advisors are very suitable for this task. It is psychologically relaxing for a decision maker to know that they can always blame their advisors if a decision turns out to be wrong. The decision maker who uses advice shares the burden of decision making. The last reason for taking advice is that the decision makers are generally reluctant to completely reject the help offered to them.

Previous studies reveal a significant difference between the use of advice and assessments of its quality [13]. Although decision makers are not very effective in using and combining the advice they receive, they can competently assess the quality of advice received from different sources, allocating weights to distinguish a good advisor from a bad one in terms of relevance and accuracy [11,13]. This is supported in time series forecasting, where accurate and reliable information leads to improvements in the quality of judgments and forecasts [27], even though individuals still tend to favour their own judgments over other information [22]. After apparently assessing the quality of advice, a decision maker usually forms a judgment regarding the reputation of an advisor based on the accuracy of advice provided over repeated occasions. It is also suggested that the reputations formed about advisors are directly related to the level of 'discounting' of their advice [33]. Discounting in this context refers to the decision makers' giving more weight to their self estimates or opinions than the advisors' recommendations, thus partially ignoring the advice they have received while explicitly favoring their own judgment. As expected, a good reputation will lead to a decrease in the amount of advice discounting. It is found that advisors' reputations are formed rapidly in just a few trials and later adjusted in an asymmetrical way (i.e., it is easier for advisors to lose a good reputation they have established than to recover from a bad reputation) [33].

Related to the assessments of advisors' reputation are the issues of trust, confidence, and expertise. It is found that cues to expertise, like confidence of an advisor in his/her recommendations, are important in building trust, especially when other information about that person is not available [29]. The same study also points out that a decision maker's trust in an advisor is directly linked to the acceptance of the advice that is received from that advisor. Extending this framework, Wærn and Ramberg [31] examine the formation of trust when the provided advice is in the form of an explanation from a computer or a human advisor. It is found that for easy tasks, humans are trusted more than computers; for harder tasks in which the subjects are rarely accurate, the trust in computers seems to exceed the trust in human advisors. Investigating the effects of extra information on decision making, it is found that decision makers are often eager to use any available information, even if they are only rumors or predictions that may worsen accuracy [4]. Accordingly, it may be assumed that when people are given advice, information or explanation, they will not be dismissive and will attempt to use all sources at least partially.

Within the specific DSS domain, using a forecasting task setting, Lawrence et al. [19] provide the only study that examines the effects of presenting explanations on DSS advice acceptance. The authors divided explanations into two types: technical explanations (a "trace" type explanation—the technique used for making the forecast and the reasons for its adoption) and managerial

explanations (a "justification" type explanation focused on what the forecast means given the time series in question); they report that providing an explanation (in either format) increased the acceptance of the advice and had positive effects on decision-maker confidence.

2.4. The current study

Given the practical relevance of justification type explanations within a DSS framework and the relative absence of research in this area, the current study focuses on the relationship between the use made of advice from a forecasting decision support system and the characteristics of the explanation accompanying that advice. We have chosen, for obvious reasons given the forecasting task setting, to frame the explanations in the classification categories of *justification/automatic/text* using the scheme presented in Section 2.1. In particular, this study aims to examine the influence of structural characteristics of justification type explanations on DSS advice acceptance (as conveyed via adjustments made to provided advice). From the literature on advice-taking research, we have chosen to examine the two structural characteristics of explanation length and level of confidence conveyed in the explanations. Explanation length may constitute an important factor for explanation effectiveness. The same thrust of information could be given in a short, brief and to-the-point format, or it can be provided in a long and detailed manner. Possible effects of the length of explanations are not clear. Longer explanations may be advantageous, since they can express information clearly and in more detail, with no short cuts. However, they can be boring, and may result in incomplete reading. Brief explanations may be advantageous, since they can be easily absorbed, although lacking the clarity and comprehensibility of the longer ones.

We hypothesise that more information in the explanation enhances its value, hence:

H1. Long explanations are more persuasive than short explanations.

Regardless of its length, an explanation can bear either a strong/precise tone, or convey vagueness/uncertainty. It may be argued that the level of confidence conveyed in an explanation may also elicit mixed reactions. Strong and precise language used in an explanation may convey a higher level of confidence and may therefore be more likely to persuade a decision maker to accept the DSS advice. Similarly, an explanation phrased in vague and uncertain terms may carry less persuasive power and may invite less

confidence in the associated forecast; there could then be a greater likelihood of rejecting the advice and modifying the forecast. However, an explanation presented with confident and strong wording carries also the possibility of backfiring given that forecasting, due to its nature, involves no certainties. Therefore, the more strong and precise the style gets, the more unrealistic it may become to users. Decision makers receiving a strong and confident explanation may think that the explanation is "too confident", unnatural, or that it is unwarranted.

We hypothesise that confident explanations are more persuasive, hence:

H2. Strongly confident explanations are more persuasive than weakly confident explanations.

3. Research methodology

3.1. Subjects

The participants were 116 3rd year business students at Bilkent University who were taking a business forecasting course offered in three sections. No monetary incentives were given, but participation in the study earned the students extra credit in their final course grades.

3.2. Design

In a pencil and paper task, participants were presented with 30 time series plots followed by one-period ahead point and 95% interval forecasts for each series. External forecasts were accompanied by explanations. While all the subjects received the same set of time series (with a different randomized ordering for each subject) and forecasts, the explanations they received were manipulated experimentally.

Two independent variables were manipulated regarding the structural characteristics of explanations. The first independent variable was the length of explanations with the levels being "short" vs. "long". The level of confidence conveyed was the second independent variable with "strong" vs. "weak" confidence as its two levels. Thus, a 2×2 factorial design led to four groups, with the following number of participants in each group:

1st Group	Short explanation, weak confidence	($n=29$)
2nd Group	Long explanation, weak confidence	($n=30$)
3rd Group	Short explanation, strong confidence	($n=29$)
4th Group	Long explanation, strong confidence	($n=28$)

In manipulating explanation length, the short explanations were developed by summarising the long explanations and omitting detail judged to be of lesser importance. Short explanations were between 1 and 1.5 lines long, while the long explanations were between 4 and 5 lines long. Manipulating conveyed confidence consisted of changing the words used in the explanations. For strong confidence, words such as "definitely, surely, etc." were used, while words such as "probably, possibly, etc." were utilized to denote weak confidence (an example of the explanations used in the experiment is provided in Appendix A).

3.3. Generation of the time series and forecasts

A total of 30 artificially generated time series was used in the study. These series were presented as real stock prices, with the names of stocks and time periods concealed in order to prevent potential biases and extraneous information effects. The same set of 30 time series was presented to each subject (with a different randomized sequencing of the series to avoid any ordering effects).

The series were constructed using three levels of trend and two levels of variability, producing six groups. Each group contained five series for a total of 30 series presented to each participant. Trend levels used were 2% increasing, 2% decreasing and zero trend. The trend level of 2% was found by ordinary least squares regression conducted on the ISE-100 (Istanbul Stock Exchange) index for the previous year. The noise added to the trended data was normally generated with zero mean and two levels of variability. The low variability series had a standard deviation of 5%, while the high variability series had a standard deviation of 15%. These variability levels are realistic for the ISE-100, and have been used in previous work [25].

External forecasts were constructed via Holt's Linear Exponential Smoothing technique. Interval forecasts were estimated using the variance of the exponential smoothing forecast error and assuming the error distribution followed a normal distribution. The participants were not given any specific information on how the forecasts were generated.

3.4. Task and procedure

Each subject was presented with a booklet including the instructions, the forms, and a wrap-up questionnaire. Instructions were discussed and detailed examples were given in the beginning of the experimental session.

In particular, subjects were requested to perform the following task for each of the 30 time series (please see Appendix B for a sample form given to participants): They were to (1) examine the provided time series, (2) study the given point and interval forecasts, (3) carefully read and think about the provided explanations. Participants were then asked whether they were satisfied with the provided point and interval forecasts. If they were not satisfied (with one or both formats), they were asked to use their judgment to modify the given point forecast or interval forecast or both.

All participants were required to do the same task in the experiment. But in one of the sections of the course (taken by 47 students), participants were also requested to complete the additional task of rating the information value of the provided explanations (i.e., the perceived usefulness of that specific explanation) using a 7-point rating scale. The 7-point scale ranged from "very misleading" to "very helpful" with the mid-scale degree being "no value".

4. Findings

The findings are outlined under four categories: percentage of forecasts adjusted; size of adjustments made; accuracy of adjusted forecasts; and the perceived information value of the explanations.

4.1. Percentage of forecasts adjusted

The simplest measure to capture the influence of explanations involves examining whether a subject has accepted, without change, the DSS advice (the forecast suggested by the system). We reason that the more persuasive the explanation, the more likely it is that the user will want to accept the advice and not make any changes. We adopt the complement measure, 'fraction of

Table 1
Percentage of point and interval forecasts adjusted

	Point forecasts		Interval forecasts	
	Weak confidence	Strong confidence	Weak confidence	Strong confidence
Short	41.95% (365/870)	42.26% (355/840)	50.69% (441/870)	57.14% (480/840)
Long	44.71% (389/870)	41.33% (372/900)	56.32% (490/870)	54.22% (488/900)

advice modified by the user', to preserve similarity with Refs. [19] and [20]. Thus we count, for each subject, the frequency of adjustments made to the point and interval forecasts. For the interval forecasts we included any adjustment made to the interval (i.e., if any modification was made on any of the bounds, it was counted as a single adjustment). The percentage of forecasts adjusted is given in Table 1, along with the count of adjustments made over the total number of data points in each of the groups.

For the adjustments made on the point forecasts, no significant differences are found among the four groups ($F_{3, 3476} = 0.79$, $p = 0.501$), with no significant main or interaction effects for explanation length and the level of confidence conveyed in the explanations. This suggests that the number of adjustments made on the point forecasts is not affected by the manipulations made through the explanations.

However, this is not the case for interval forecasts; the difference among the groups is in fact significant ($F_{3, 3476} = 2.89$, $p = 0.03$). Analysis of variance results (as summarized in Table 2) shows no significant main effects for the length and style (conveyed confidence) variables. However, the interaction effect is significant, as can be observed from Fig. 1.

Manipulating length or style (conveyed confidence) alone seems not to have a significant influence on the number of adjustments made to the interval forecasts. However, long explanations combined with weak confidence appear to lead to a higher percentage of forecasts adjusted then short explanations with weak confidence. Also, short but strongly confident explanations lead to more adjustments than long but strongly confident explanations. However the effect sizes are quite low. In summary, the results suggest that neither H1 nor H2 is supported by this study.

It can also be observed from Table 1 that subjects made adjustments to interval forecasts more than to point forecasts, regardless of the group ($t_{3479} = 12.97$, $p < 0.001$). This may be due to the fact that there is only a single point to adjust in the point forecasts, while there are two bounds that can be adjusted for the interval forecasts. In addition there is evidence that people prefer asymmetric bounds for interval forecasts and would thus tend to adjust the symmetric intervals provided [18]. Hence, for both
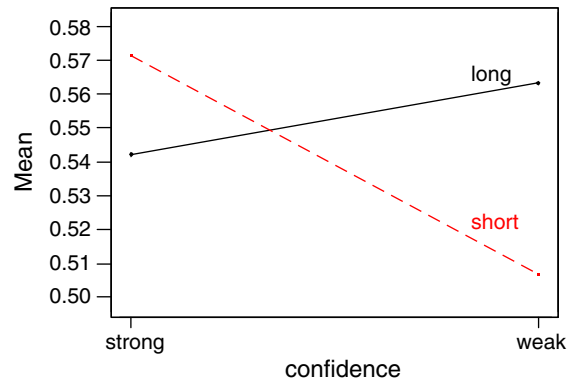

Fig. 1. Interaction effect for percentage of interval forecasts adjusted.

these reasons, interval forecasts may be more likely candidates for adjustment than point forecasts.

4.2. Size of adjustments made

In addition to examining the percentage of forecasts adjusted, we can also investigate the size of adjustments actually made to the forecasts. While the former measure shows the presence/absence of adjustments, the latter measure will reveal the magnitude of adjustments thus providing another view of the influence of the provided explanation. To capture the size of adjustments made to the point and interval forecasts, two ratios are utilized: absolute percentage adjustment in point forecasts (APAP) and absolute percentage adjustment in interval width (APAI), where

$$\text{APAP} = \frac{|\text{adjusted point forecast} - \text{provided point forecast}|}{\text{provided point forecast}} \times 100 \qquad (1)$$

$$\text{APAI} = \frac{|\text{adjusted width} - \text{provided width}|}{\text{provided width}} \times 100. \qquad (2)$$

If no adjustments are made to the provided forecasts, these ratios automatically assume the lowest possible score (0%). The further away the adjusted predictions are from the provided values, the higher are the APAP and APAI scores.

Table 2
ANOVA results for percentage of interval forecasts adjusted

Terms	Coefficient (%)	$F_{1, 3476}$	p
Length	−0.68	0.65	0.422
Confidence	−1.09	1.66	0.197
Length * confidence	−2.14	6.42	0.011

Table 3
Mean values for absolute percentage adjustment in point forecasts (APAP) and in interval width (APAI)

	APAP		APAI	
	Weak confidence (%)	Strong confidence (%)	Weak confidence (%)	Strong confidence (%)
Short	4.72	4.11	15.85	15.98
Long	4.25	4.14	16.89	13.68

Table 4
ANOVA results for size of adjustments in interval forecasts

Terms	Coefficient (%)	$F_{1,3476}$	p
Length	0.31	0.66	0.417
Confidence	0.77	3.92	0.048
Length * confidence	−0.83	4.64	0.031

As can be observed from Table 3, the average size of adjustments made to the point forecasts (measured by APAP) is not statistically different among the four groups ($F_{3,3476}=1.07$, $p=0.362$), and neither main effects nor interaction effects are found to be significant. It seems that the size of adjustments subjects introduced to the point forecasts is unaffected by different explanation manipulations. It is also clear from Table 3 that the size of adjustments made to the point forecasts is much less than the magnitude of adjustments made to the interval width, with the difference being statistically significant ($t_{3479}=28.64$, $p<0.001$).

As in the case of percentage of forecasts adjusted, adjustments are very different for the case of interval forecasts in comparison to the adjustments for point forecasts. There exists a significant difference among the groups for APAI ($F_{3,3476}=3.14$, $p=0.02$) with the least adjustment made in the long and strongly confident explanation group. Table 4 shows the significant effects of conveyed confidence (i.e., style) and the interaction between explanation length and conveyed confidence.

Fig. 2 displays the size of adjustments made in interval forecasts and parallels the findings for the percentage of forecasts adjusted. When a long explanation is accompanied by strong confidence, adjustments are much smaller than all the other possible combinations. On the other hand, when a long explanation is communicated in a weakly confident style, individuals tend to make the most adjustments. There seems to be no
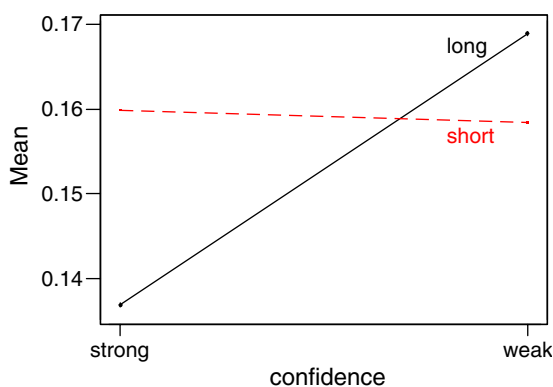

Fig. 2. Interaction effect for size of adjustments made in interval forecasts.

Table 5
Average information value

	Average information value	
	Weak confidence	Strong confidence
Short	2.43	2.33
Long	2.29	2.38

difference due to strong vs. weak confidence when a given explanation is short. In summary, the results do not provide support for H1 or H2.

4.3. Perceived information value of explanations

As mentioned previously, a subgroup of participants (i.e., 47 students in a given section of the forecasting course) was asked to evaluate the information value of each explanation by using a 7-point scale (with 1 = "very misleading"; 4 = "no value"; 7 = "very helpful"). Such ratings of participants' perceptions may provide insights about their decision to adjust the system provided forecasts, as well as the size of adjustments they actually make.

For this analysis, we grouped the data into a 3-point scale. In the recoded version, the levels became 1 = "misleading" (initial ratings of 1 and 2 combined), 2 = "no real value" (initial ratings of 3, 4 and 5 combined), and 3 = "helpful" (initial ratings of 6 and 7 combined). The average information value attributed to the provided explanations is 2.36 across all groups, which is significantly greater than 2 which was associated with "no real value" ($t_{1401}=20.91$, $p<0.001$). Hence, confirming the participants' verbal comments at the end of the study, we may conclude that the subjects on average found the provided explanations helpful to some degree. Table 5 provides the average information value within each group. Interestingly, explanation length and confidence appear to have no impact on perceived information value. However a significant difference between the groups was found in the perceived information value of explanations ($F_{3,1398}=3.07$, $p=0.03$) due to a significant interaction term as depicted in Table 6 and Fig. 3. This result suggests that the characteristics of length and confidence have a low manipulation effect on the

Table 6
ANOVA results for information value

Terms	Coefficient	$F_{1,1398}$	p
Length	0.023	1.77	0.183
Confidence	0.006	0.12	0.739
Length * confidence	0.047	7.44	0.006

Table 7
Differences in adjustment and accuracy measures with respect to information value

Measures	Misleading (%)	No real value (%)	Helpful (%)	$F_{2, 1399}$	p
% of point forecasts adjusted	54.8	41.3	26.9	26.49	<0.001
% of interval forecasts adjusted	63.0	50.8	41.1	13.38	<0.001
APAP	8.0	4.4	3.1	21.80	<0.001
APAI	18.5	13.5	11.7	6.02	0.002
MAPE	26.6	19.8	16.0	8.50	<0.001
Hit rate	71.8	80.4	84.8	6.84	0.001

subjects and this may explain the lack of significant influence of these characteristics. The persuasive influence of an explanation may be based on its information value, not its characteristics of length and confidence. We explore this in the following analysis, where Table 7 presents the fraction of forecasts changed and the size of the change grouped according to the perceived information value of the explanations.

Table 7 shows the very strong persuasive influence of perceived information value. All the adjustment scores (i.e., percentage of point forecasts adjusted, percentage of interval forecasts adjusted, APAP, APAI) decrease as the attributed information value of an explanation increases. Furthermore we also include in this table the forecast accuracy[2] for point forecasts and the hit rate[3] for interval forecasts as measures of decision making skill. This shows that the accuracy of the adjusted forecasts steadily improves as the perceived information value increases (i.e., MAPE decreases and hit rate increases at the same time). These findings confirm the previous results in that if a participant finds an explanation to be helpful in understanding the provided forecasts and time series, he/she tends to trust and accept the provided forecasts. The explanation is persuasive if it has high perceived information value and this results in smaller adjustments and higher accuracy in the adjusted final forecasts. On the other hand, if an explanation is deemed "misleading", the provided forecasts lack credibility, leading to a higher rate of adjustment and reduced accuracy.

5. Discussion and conclusions

The findings of the current study suggest that providing an explanation for a given interval forecast has an influence on the adjustment and therefore on the acceptance of that forecast, a finding not replicated for point forecasts. However, these results appear to be contingent on the structural characteristics of explanations. In particular, long explanations expressed in a strongly confident style seem to be most influential compared to others, leading to less adjustment. A long and confident explanation leads forecast users to introduce judgmental adjustments less often, and the magnitudes of the adjustments appear to be smaller as well. On the other hand, if short explanations are required, it seems better to present them in weakly confident wording. The perceptions of usefulness of such explanations (i.e., their perceived information value) also seem to be higher.

Interestingly, all of the above arguments are supported only for interval forecasts. The provided explanations have no statistically significant impact on point forecasts. A provided explanation for a specific data set seems to affect the interval forecast adjustment on that specific data set; and yet, adjustment of the point forecast for that specific data set appears to be independent from that explanation. A potential reason lies in the experimental manipulations conducted; they perhaps were not particularly suited to produce any direct effects on point forecasts. Point forecasts are single-shot numbers meant to contain a variety of information, both internal and external; it is generally very difficult to incorporate all the information within a single number. In this regard, trying to reach an accurate point forecast becomes a heavy judgmental burden, especially since the point predictions appear to convey a false sense of certainty. People may find it difficult to incorporate the cues of trust or confidence they have received from the explanations into their point forecasts. They may have simply adjusted the point forecasts according to some mental heuristics or shortcuts, making them less responsive to the manipulations conducted through the framing of explanations. It is found that a lower percentage of point forecasts is adjusted relative to the interval predictions. The size of adjustments made is also less. In short, it is highly likely that point forecasts may respond to different kinds of manipulations than interval forecasts.

---

[2] The measure used is mean absolute percentage error (MAPE) defined as the mean value of the absolute value of the forecast error divided by the actual value and expressed as a percentage.

[3] The hit rate is the percentage of times the actual value falls within the interval forecast.
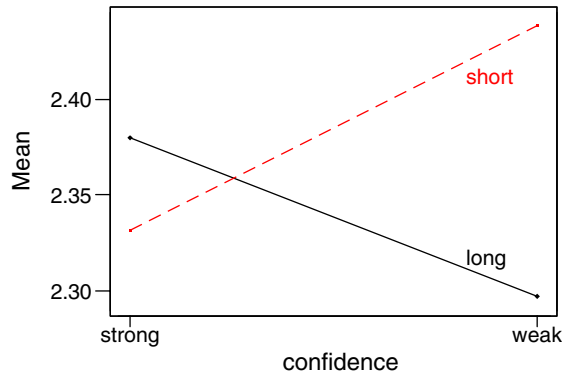
Fig. 3. Interaction effect for information value of explanations.

We found little evidence that the explanation characteristics of length or confidence exercise a persuasive influence on acceptance of DSS advice; this may be because our manipulations of length and confidence did not achieve the desired impact on the subjects. However a very significant impact was achieved through the perceived information value of the explanations. Those explanations rated as of high information value were much more persuasive than those of low value. This suggests that the critical element in designing explanations is not length or confidence, but information value.

This study was conducted with business students who were completing a forecasting course. The use of student participants is common in judgmental forecasting research and indeed in the wider field of experimentally based decision making and experimental economics research [7,32]. This practice is based on findings showing students as adequate surrogates for practitioners in decision making [1] and other business contexts [15,23]. Future studies with managers will prove valuable in extending the current work into investigating the potential effects of varying levels of expertise on the acceptance of explanations. In addition, current findings might be confounded by the complexity of the forecasting task given (i.e., stock price forecasting). Future work may utilize tasks of varying complexity to examine the possible interactions and confounding factors involved in multimedia systems to highlight comparisons with the text-based, trace type, automatically invoked explanations.

The results of this study have many practical implications in a variety of fields. The most apparent impact would be on institutions providing professional consulting, investment advice, and forecasting. Supplying explanations carries a special significance for the financial interests of those firms whose success is mostly dependent on their ability to convince their customers/users of the value of the information support they are providing. If the users are not satisfied with the presented advice/forecasts, they may switch to competing information providers.

In conclusion, our findings have immediate repercussions for the explanation facilities incorporated into decision support systems. Although there has been extensive research on designing these facilities, considerably less attention has been given to the structural characteristics of the explanations generated by these facilities. The results of the current work suggest directions for promising future work in this area.

Appendix A. Sample explanations

A.1. Short explanation, strong confidence

New government subsidy is certainly responsible for the upward trend starting from week 20. We strongly believe that this is excessive and there will be a drop.

A.2. Short explanation, weak confidence

New government subsidy may have led to the upward trend starting from week 20. This seems to be excessive and there may be a drop.

### A.3. Long explanation, strong confidence

The government has decided to provide a new subsidy for some of the company's products. This was announced in week 20, and it immediately caused the upward trend in the stock prices starting from that week. This upward trend can be distinctly observed from the time series data. However, we find this increase to be very excessive. We, therefore, strongly believe that there will be a drop in the stock prices in the 26th week.

A.4. Long explanation, weak confidence

The government has decided to provide a new subsidy for some of the company's products. This was announced in week 20, and it may have led to the upward trend in the stock prices starting from that week. This upward trend may be observed from the time series data. However, this increase may appear to be excessive so that there may be a drop in the stock prices in the 26th week.

Appendix B. Sample form given to participants for each time series

Week	Stock price
25	4276
24	3857
23	3597
22	3308
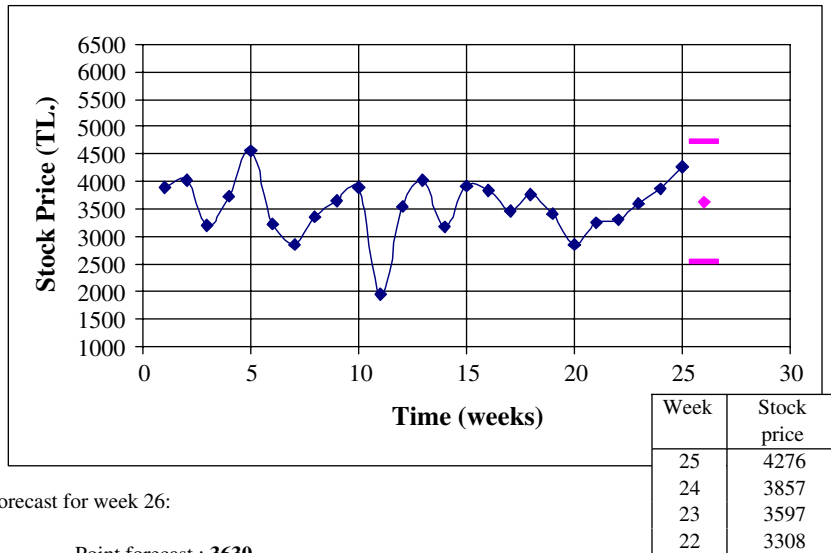
Our forecast for week 26:

Point forecast : **3630**

95% interval forecast: **[2530 – 4730]**

Our explanation :
New government subsidy is certainly responsible for the upward trend starting from week 20.We strongly believe that this is excessive and there will be a drop.

Would you like to modify the given point forecast?   Yes ☐   No ☐

If yes,
Your point forecast: _____

Would you like to modify the given interval forecast?   Yes ☐   No ☐

If yes,
Your 95% interval forecast: **[_____ – _____]**

References

[1] R.H. Ashton, S.S. Kramer, Students as surrogates in behavioral accounting research: some evidence, Journal of Accounting Research 18 (1980) 1–15.

[2] F. Davis, J. Kottemann, Determinants of decision rule use in a production planning task, Organizational Behavior and Human Decision Processes 63 (1995) 145–157.

[3] J.S. Dhaliwal, I. Benbasat, The use and effects of knowledge-based system explanations: theoretical foundations and a framework for empirical evaluation, Information Systems Research 7 (3) (1996 (September)) 342–362.

[4] N. DiFonzo, Cueing rumor and prediction: making sense (but losing dollars) in the stock market, Organizational Behavior and Human Decision Processes 71 (3) (1997) 329–353.

[5] J. Durkin, Expert systems: a view of the field, IEEE Expert (1996 (April)) 56–63.

[6] R. Fildes, P. Goodwin, M. Lawrence, The design features of forecasting support systems and their effectiveness, Decision Support Systems, in press.

[7] P. Goodwin, Enhancing judgmental sales forecasting: the role of laboratory research, in: G. Wright, P. Goodwin (Eds.), Forecasting with Judgment, John Wiley and Sons, Chichester, 1998, pp. 91–111.

[8] P. Goodwin, D. Önkal-Atay, M.E. Thomson, A.C. Pollock, A. Macaulay, Feedback-labelling synergies in judgmental stock price forecasting, Decision Support Systems 37 (2004) 175–186.

[9] S. Gregor, Explanations from knowledge-based systems and cooperative problem solving: an empirical study, International Journal of Human–Computer Studies 54 (2001) 81–105.

[10] S. Gregor, I. Benbasat, Explanations from intelligent systems: theoretical foundations and implications for practice, MIS Quarterly 23 (4) (1999 (December)) 497–530.

[11] C. Harries, N. Harvey, Taking advice, using information and knowing what you are doing, Acta Psychologica 104 (2000) 399–416.

[12] N. Harvey, I. Fischer, Accepting help, improving judgment, and sharing responsibility, Organizational Behavior and Human Decision Processes 70 (2) (1997) 117–133.

[13] N. Harvey, C. Harries, I. Fischer, Using advice and assessing its quality, Organizational Behavior and Human Decision Processes 81 (2) (2000) 252–273.

[14] F. Hayes-Roth, N. Jacobstein, The state of knowledge-based systems, Communications of the ACM 37 (3) (1994) 27–39.

[15] K.A. Houghton, J.J.F. Hronsky, The sharing of meaning between accounting students and members of the accounting profession, Accounting and Finance (1993 (November)) 131–147.

[16] H. Irandoust, Attitudes for achieving user acceptance explaining, arguing, critiquing, 7th International Command and Control Research and Technology Symposium, Quebec City–Canada (2002).

[17] B. Kleinmuntz, Why we still use our heads instead of formulas: toward an integrative approach, Psychological Bulletin 107 (1990) 296–310.

[18] M. Lawrence, M. O'Connor, Judgmental forecasting in the presence of loss functions, International Journal of Forecasting 21 (1) (2005 (January–March)) 3–14.

[19] M. Lawrence, L. Davies, M. O'Connor, P. Goodwin, Improving forecast utilization by providing explanations, 21st International Symposium on Forecasting, Atlanta–USA (2001).

[20] M. Lawrence, P. Goodwin, R. Fildes, Influence of user participation on DSS use and decision accuracy, Omega 30 (5) (2002 (October)) 381–392.

[21] J.S. Lim, M.J. O'Connor, Judgmental adjustment of initial forecasts, Journal of Behavioral Decision Making 8 (1995) 149–168.

[22] J.S. Lim, M. O'Connor, Judgmental forecasting with time series and causal information, International Journal of Forecasting 12 (1996) 139–153.

[23] E.A. Locke, Generalizing from Laboratory to Field Settings, Lexington Books, 1986.

[24] J.Y. Mao, I. Benbasat, The use of explanations in knowledge-based systems: cognitive perspectives and a process-tracing analysis, Journal of Management Information Systems 17 (2) (2000 (Fall)) 153–179.

[25] M. O'Connor, W. Remus, K. Griggs, Judgmental forecasting in times of change, International Journal of Forecasting 9 (1993) 163–172.

[26] K.N. Papamichail, S. French, Explaining and justifying the advice of a decision support system: a natural language generation approach, Expert Systems with Applications 24 (2003) 35–48.

[27] W. Remus, M. O'Connor, K. Griggs, Does reliable information improve the accuracy of judgmental forecasts? International Journal of Forecasting 11 (1995) 283–293.

[28] E.H. Shortliffe, Computer-Based Medical Consultations: MYCIN, Elsevier Computer Science Library, New York, 1976.

[29] J.A. Sniezek, L.M. Swol, Trust, confidence, and expertise in a judge–advisor system, Organizational Behavior and Human Decision Processes 84 (2) (2001) 288–307.

[30] E. Turban, J.E. Aronson, Decision Support Systems and Intelligent Systems, 6th edPrentice Hall, Upper Saddle River, NJ, 2001.

[31] Y. Wærn, R. Ramberg, People's perception of human and computer advice, Computers in Human Behavior 12 (1) (1996) 17–27.

[32] R. Webby, M. O'Connor, Judgmental and statistical time series forecasting: a review of the literature, International Journal of Forecasting 12 (1996) 91–118.

[33] I. Yaniv, E. Kleinberger, Advice taking in decision making: egocentric discounting and reputation formation, Organizational Behavior and Human Decision Processes 83 (2) (2000) 260–281.

[34] L.R. Ye, P.E. Johnson, The impact of explanation facilities on user acceptance of expert systems advice, MIS Quarterly 19 (2) (1995 (June)) 157–172.

M. Sinan Gönül received his B.S. in Electrical and Electronics Engineering from Middle East Technical University and his MBA from Bilkent University, Turkey. He participated in the International Management Program at McGill University, Canada. He is currently pursuing a Ph.D. degree in Operations Management and Decision Sciences. His research interests focus on judgment and decision making, decision support systems, judgmental forecasting, and econometric forecasting.

Dilek Önkal is a Professor of Decision Sciences in the Faculty of Business Administration at Bilkent University, Turkey. She received a Ph.D. in Decision Sciences from the University of Minnesota, and is doing research on judgmental forecasting, probabilistic financial forecasting, judgment and decision making, decision support systems, risk perception and risk communication. She is an Associate Editor of the *International Journal of Forecasting*. Her work has appeared in several book chapters and journals such as *Organizational Behavior and Human Decision Processes, Omega, Decision Support Systems, International Journal of Forecasting, Journal of Behavioral Decision Making, Risk Analysis, Journal of Forecasting, International Federation of Technical Analysts Journal, International Forum on Information and Documentation, Risk Management: An International Journal,* and *European Journal of Operational Research*.

Michael Lawrence is Emeritus Professor of Information Systems in the Commerce and Economics Faculty at the University of New South Wales, Sydney, Australia. Before joining the University he worked for Ciba-Geigy Corporation and Corning Glass Works in the USA. He has held visiting positions at Lancaster University, England; London Business School and Imperial College, London; Insead, Fontainebleau, France, and University of Maryland, USA. He is past President of the International Institute of Forecasters (IIF), the major professional and academic body committed to improving the state of the art of forecasting and is currently editor of the *International Journal of Forecasting* (an Elsevier journal) and serves on the Board of Directors of the IIF. He is also an associate editor of *Managerial and Decision Economics* (a Wiley journal) and Emeritus Associate Editor of *Omega* (an Elsevier journal). He has a PhD from the University of California, Berkeley in Operations Research. His research interests centre generally on how to best support decision making where a significant component of the decision involves management judgment, with specific interest in how to support the task of short term forecasting.