CHAPTER 3 : EXPERIMENT 1

Initial investigations into UI utility

3.1 INTRODUCTION

After reviewing the literature a contingent model describing the factors that may influence UI utility and utilisation was proposed. This model was developed by (a) reviewing areas of decision-making literature which are relevant to UI utility, (b) through examining studies which have previously attempted to directly investigate UI utility, and (c) by considering how all these elements might interact to impact UI utility. Due to the inconclusive nature of the evidence, the proposed research model is speculative.

This chapter describes the first of three experiments that each investigates UI utility. The contingent model of UI utility guides the design for each experiment. The first experiment contains initial investigations into UI utility, and in particular focuses on defining a task where UI is of normative value.

This chapter presents the first experiment by outlining the research hypotheses (section 3.2), defining the experimental design and procedure (section 3.3), detailing the main findings (section 3.4) and concluding with a discussion of the results (section 3.5).

3.2 HYPOTHESES

The literature review highlighted the limited and inconclusive findings from prior investigations into UI utility. Past studies have not been able to determine whether UI has an overall impact on either decision performance (whether those with UI make better decisions than those without) or even decision behaviour (whether those with UI make different decisions to those without). The most basic questions regarding UI utility are yet to be unequivocally answered. As a result, this study’s initial investigations into UI utility must address fundamental questions, which will help demonstrate the contingent model of UI utility outlined in the literature review. In particular, this study will look to understand whether the provision of UI impacts the simplest of decision related metrics: decision behaviour and decision performance.

Decision behaviour is a measure of how a decision-maker reacts to a certain situation, without regard to whether the reaction is correct or not. For example, decision behaviour could measure the proportion of investment opportunities a decision-maker decided to invest in, without considering whether the investments were a success. In contrast, decision performance is a measure of how correct the judgments of a decision-maker are. For example, decision performance could measure the proportion of successful investments made by a decision-maker. These metrics will be used throughout this study to measure the impact of providing UI.

As mentioned, previous studies into UI utility have failed to find conclusive results supporting the value of UI in decision-making. Two factors that may have heavily influenced this are the availability of proxies for UI, and the use of a task in which the UI did not have significant diagnostic or decision-making value.

When decision-makers are presented with a task that has multiple information cues, there may be numerous ways to determine the level of uncertainty in the task. Even if explicit UI is provided, decision-makers may choose to assess uncertainty using other available information cues, such as historical data. The provision of historical data in Oliver (1972) and Keys (1978) may have reduced the information value of UI and influenced the lack of significant results in those studies. As a result, this study will minimise the availability of UI proxies.

The selection of an appropriate task is of the utmost importance if UI is to be of utility. By definition, UI (such as in the form of a confidence level) indicates the likelihood of an event occurring. Those provided with UI should be better able to identify when an event is likely to occur. A typical business decision-making task that would benefit from the provision of UI is a capital investment decision, which is a part of the capital budgeting process.

"Capital budgeting is a formal means of making long-range decisions for investments. Examples include plant locations, equipment purchases, additions of product lines, and territorial expansions (Horngren et al., 1992, p. 1183)". Such significant investments must be given careful consideration, since they represent significant sums of money, and organisations generally have limited resources and therefore limited investment options. Only those investments with a high likelihood of success must be pursued. Accordingly, significant effort has been put into analysing and evaluating investment alternatives. Several different techniques have been developed to compare investment alternatives, including traditional approaches (for example, Accounting Rate of Return, or Payback Period) and the currently more favoured discounted cash flow techniques (such as Internal Rate of Return, or Net Present Value).[1]

---

[1] The details of these measures can be found in most introductory accounting textbooks (for example, Horngren et al., 1992), however a deep understanding is not required for this discussion.

Regardless of what techniques are used to evaluate investment alternatives, the process of selecting an investment within the capital budgeting process boils down to whether the investment is likely to achieve some predetermined target level. Common target levels appearing in the capital budgeting process include ‘cost of capital’, ‘hurdle rate’, ‘discount rate’ or ‘target rate of return’. Investment alternatives unlikely to meet these target levels are rejected. For example, an organisation may typically require capital investments above a certain size (for example $1 million) to generate at least a 15% return on a discounted basis (that is, calculating the return by using discounted cash flows).

Estimates of the likely return from capital investment projects have been found to be uncertain (for example, see Sihler, 1964; Berg, 1965; Aharoni, 1965; King, 1975) so the fact that the estimated return of a potential capital investment (for example, 16%) is higher than the hurdle rate for an organisation (for example, 15%) may not be enough for the project to get the green light. Management may also require a particular level of certainty that the return will be at least equal to the hurdle rate. Given the long-term, strategic nature of capital investment projects, management may desire to be at least 75% sure (a value greater than 50%, but arbitrarily selected) that the hurdle rate of return will be achieved. By providing an indication of the likelihood of a target level being achieved UI should add significant value to the capital budgeting process.

Capital Budgeting Rules for Example Company	
Hurdle Rate	**15%**
Required level of confidence that hurdle rate will be achieved	**75%**
Example Project	
Estimated Return	**16%**
75% Confidence Level for estimated return	**13%**
Project Evaluation	
Project rejected: 75% confidence level (13%) < hurdle rate (15%) NB. Without the availability of a confidence level, the project would have been accepted {Estimated Return (16%) > Hurdle Rate (15%)}	

Figure 3-1: How confidence levels (UI) can change capital budgeting decisions.

To illustrate, consider a business analyst assessing whether the production capacity of a manufacturing plant should be increased (see Figure 3-1). The organisation has a hurdle rate of return equal to 15%, and they require capital investments to have at least a 75% chance of achieving the hurdle rate of return. When calculating the estimated return for such a project an employee could be required to provide not only a ‘best estimate’ for the return (for example, 16% return which we will assume had a 50% chance of being exceeded and a 50% chance of being at most equalled) but also the level of return which the particular investment alternative had a 75% chance of exceeding. In this example, assume it was calculated the plant expansion had a 75% chance of delivering at least a 13% return. The project would be rejected despite the predicted return of 16% being higher than the hurdle rate of 15%, because there was a less than 75% chance that the hurdle rate would be achieved.

By comparing a ‘confidence level’, which indicates the point (13% return in the above illustration) at which a prescribed level of confidence will be achieved (75% in the above illustration), with some form of target level (for example, a 15% hurdle rate) better decisions should be made. Without the confidence level and with only the best estimate of return to compare to the hurdle rate, it is more likely that erroneous decisions will eventuate. It is the comparison of the hurdle rate (or target level) with the confidence level (the UI), highlighting any difference, which creates the utility for the UI.

Given that the difference between the target level and 75% confidence level is an indication of the likelihood of success of a project, it is pertinent to investigate the extent to which people are sensitive to the sign and magnitude of this difference. The sign of the difference between confidence levels and target levels should strongly influence decision behaviour. Normatively, when the difference is positive (confidence level greater than or equal to target level) investments will be accepted, while a negative difference (confidence level less than target level) will see an investment opportunity rejected.

The sign of the difference between a confidence level and a target level defines the decision rule for capital budgeting decisions. Changing the sign of the difference should change the decision behaviour of those with UI more than those without, since the confidence level directly tells them whether to invest or not (see Figure 3-2 for a definition of the sign of the difference between a confidence level and target level). However, while changing the sign of the difference should impact decision behaviour (that is, whether decision responses change, without looking at if the response was correct or not), decision performance (that is, whether a decision response is correct or not) may not be impacted. The UI should enable the ability to equally identify both good and bad investments (positive and negative differences). A change in the sign of this difference changes responses, but does not alter the final outcome of the investment. Those without UI will not be aware of the change in sign of the difference between a confidence level and target level, and therefore any change in sign should not affect their decision behaviour as abruptly.

	Project A	Project B
Hurdle Rate (target level)	**15%**	**15%**
Estimated Return	**21%**	**21%**
75% Confidence Level for estimated return	**13%**	**16%**
Confidence Level – Hurdle Rate	**-2%**	**1%**
Sign of difference between Confidence Level and Hurdle Rate	**Negative**	**Positive**

Figure 3-2: An Illustration defining the sign of the difference between a confidence level and target level.

There is a significant body of literature investigating the impact of target levels on decision-making, and the relevance of this work needs to be considered.[2] Several authors have observed that business executives define risk as the prospect of not being able to achieve a target level (for example, Mao, 1970; Conrath, 1973; Greer, 1974; Laughhunn, Payne & Crum, 1980; Crum, Laughhunn & Payne, 1981; Ho & Vera-Muñoz, 2001; March & Shapira, 1987). When a target level is threatened, decision behaviour changes. This could occur, for example, when a sales forecast is only just above the target level.

This finding has relevance to capital investment decisions. It tells us that the distance between a target level (hurdle rate), set by organisations for their capital investments, and any estimate of the likelihood of success (such as a 75% confidence level) of a capital investment opportunity should influence the risk associated with the decision. For example, for a hurdle rate of 15% a project with a 75% chance of achieving a16% return may be treated differently to a project with a 75% chance of achieving a 19% return, even though both are above the hurdle rate. As outlined in the contingent model of UI utility (See Figure 2-10 in the Literature Review) jeopardising the achievement of the target level is seen as a risk and may influence UI utility.

The results of other studies investigating the achievement of target levels would lead us to believe that when the magnitude of the difference is small, those with UI will not have an advantage over those without UI. When the 75% confidence level and target level are close together, in a task where a decision-

---

[2] A considerable amount of research has explored the impact of using planning budgets as target levels, and how that can act as both a motivator and de-motivator depending upon the ability of the individual to achieve the target level. This literature is not relevant to capital investment decision-making, since the target level (the hurdle rate) is likely to be a constant, financial yardstick used throughout the organisation, rather than a motivator for a specific individual.

maker needs to be at least 75% sure that the target level will be achieved, there will be a perception that achievement of the decision objective is being threatened, but the decision-maker may be indifferent to choosing to invest or not to invest. For example, in Project D in Figure 3-3 where the 75% confidence level lies on the hurdle rate either an “invest” or “don’t invest” decision are both equal in merit. For those with UI, it will be unclear whether the UI (75% confidence level) can be completely trusted, since it may have some degree of error. As a result, the decision performance of those with UI may be affected. Decision-makers may also be indifferent to this situation, since the difference between a correct and incorrect answer may be negligible (see Figure 3-3 for a definition of the magnitude of the difference between a confidence level and target level).

	Project A	**Project B**	**Project C**	**Project D**
Hurdle Rate (target level)	**15%**	**15%**	**15%**	**15%**
Estimated Return	**21%**	**21%**	**21%**	**21%**
75% Confidence Level for estimated return	**13%**	**16%**	**19%**	**15%**
Confidence Level – Hurdle Rate	**-2%**	**1%**	**3%**	**0%**
Magnitude of difference between Confidence Level and Hurdle Rate	**2%**	**1%**	**3%**	**0%**

Figure 3-3: An Illustration defining the magnitude of the difference between a confidence level and target level.

As the magnitude of the difference between the 75% confidence level and target level increases, the provision of UI should improve decision performance.

As the target level moves far enough away from the confidence level to the point where any perceived error in the confidence level does not impact decisions, the UI will be trusted more, and the decision performance of those with UI will improve relative to those without UI. This improvement in decision performance coincides with when it is most important that a correct decision be made: when there is a large difference between a correct and incorrect decision. For example, when the target level is far above the 75% confidence level, and the correct decision would be to reject the investment if the decision objective is to be at least 75% sure, the potential loss of saying ‘yes’ is much greater. The potential loss is greater, because the further above the confidence level the target level gets, the more wrong a ‘yes’ answer is.

Based on the preceding discussion, the hypotheses for Experiment 1 are:[3]

H_1 UI will improve decision performance.[4]

H_2 UI will provide greater relative decision performance as separation of a target level and UI (a 75% confidence level) increases.

---

[3] Note that the hypotheses are presented in non-null form, because the aim is to test the contingent model of UI utility developed in the literature review.

[4] The overall effect of UI on behaviour (that is, whether decision responses change, without looking at if the response was correct or not) will not be investigated. Changes in behaviour at an overall level are likely to be random. However, changes in behaviour due to the interaction of an experimental factor and the provision of UI will be investigated in order to better explain changes in decision performance.

These hypotheses will be investigated under the following restricted set of conditions:

- A capital investment task where decision performance is defined by the objective that participants must be at least 75% sure that a target level (or hurdle rate) will be achieved;
- UI is provided in the form of a 75% confidence level, which indicates the point that has a 75% chance of being exceeded;
- Proxies for UI are not present; and
- Explicit, objective UI has diagnostic or decision-making value.

The first hypothesis will be investigated by looking at overall decision performance, where participants must be at least 75% sure that a target level is achieved. Further analysis into how the sign of the difference between the 75% confidence level and target level changes decision behaviour will help explain differences in overall decision performance. The second hypothesis will be operationalised by looking at how the magnitude of the difference between a 75% confidence level and target level impacts decision performance.[5]

---

[5] While the magnitude of the difference between the confidence level and target level should influence both decision behaviour (responses) and decision performance (the correctness of responses), only decision performance will be investigated. The significance of the impact should be the same regardless of whether the dependent variable is decision performance or decision behaviour (since their derivations are related). We are most interested in performance, since it will better identify the benefits or disadvantages of UI, and therefore the investigation into the influence of magnitude on UI utility is limited to decision performance only.

How decision performance (correctness) will be measured is addressed later in this chapter, after the research instrument has been introduced. However, it is worth noting at this stage that the metric for determining decision performance will be asymmetric by requiring subjects to make investments when they are 75% sure of success.

The first experiment will vary the sign and magnitude of the difference between a target level and a 75% confidence level in order to investigate how the utility of UI is impacted when the UI indicates a greater or lesser chance of achieving a target level. The influence of target levels on UI utility has been researched in earlier studies, but only to a limited extent. Herman, Ornstein & Bahrick (1964) and Ho, Keller & Keltyka (2002) demonstrated the important role that the location of target levels plays in UI utility. They both showed how the target level provides a point of comparison for the UI, enabling the UI to directly indicate the likelihood of achieving the target level. Having target levels as part of the decision task increased the need to understand and utilise UI. Targets were of central importance to the search and attack game used in Herman, Ornstein & Bahrick (1964), however they did not further explore the importance of target levels since they were investigating UI format. The target levels were not directly manipulated and were in the centre of the UI (which was a circular confidence interval, gradually reducing in size), providing less opportunity for comparison with the UI. Ho, Keller & Keltyka (2002) explored the influence of target levels, specifically examining how the proximity of the target level relative to a forecast and ambiguity range (centred around the forecast) affected behaviour.[6] When the forecast was above the target level (a choice framed as a gain) individuals were outcome ambiguity avoiding (they believed the forecast

---

[6] Ho, Keller & Keltyka (2002) carried out two experiments on framing effects and target levels, one investigated probabilistic ambiguity while the other studied outcome ambiguity. The authors' definition of probabilistic ambiguity agrees with traditional definitions of ambiguity, where choices are made between options with varying degrees of ambiguous probabilities. Their definition of outcome ambiguity appears to be more like uncertainty rather than ambiguity, since it's the outcomes that are in question, not the probabilities associated with the outcomes. For example, they use a range of possible outcomes to illustrate outcome ambiguity. As a result, it is appropriate to use the results of their outcome ambiguity experiment as part of this discussion.

more than the range) whereas when the forecast was below the target level (a choice framed as a loss) individuals were ambiguity seeking (believing in the range rather than the forecast). Ho, Keller & Keltyka (2002) manipulated the target levels, but only with respect to the task's framing, in other words with respect to the forecast rather than the range.

Hypotheses for the first experiment have been designed to address two goals. The goal of the first hypothesis (H_1) is to help understand the overall effect of UI on decision-making in an appropriate task. The second hypothesis (H_2), aims to understand the influence on UI utility of the important element of target levels. By providing a point of comparison, target levels should improve UI utility. However, they may also confound UI utility, when they threaten the decision objective (for example, March & Shapira, 1987), such as being 75% successful. The next section (section 3.3) looks at how these hypotheses are operationalised in the research design for Experiment 1.

3.3 RESEARCH DESIGN

A paper-based decision-making experiment with a 3(x4) factorial design was devised to investigate the contingent utility of UI. The experiment's decision task required the evaluation of several capital investment opportunities. Participants were told to assume the role of a management consultant, advising a manufacturing company on whether they should increase the production capacity of each of twenty production plants, all of which produced the same product (widgets). The decision had to be based upon whether the participant thought that a given sales target level would be exceeded within the next two years.[7] The decision period was set as the second period in the forecast horizon in order to slightly reduce the impact of the historical data (i.e. the naïve forecast) and increase the need for extrapolation.

The experiment consisted of three between-subject treatment groups, with a within-subject factor that varied across four levels, thus giving a 3(x4) factorial design. Only one of the treatment groups received UI. All three treatment groups received minimal historical data (the last two years' sales figures for the widget market that the particular production plant supplied) and a target level that sales had to exceed two years into the future if a recommendation to increase production capacity was to be made. The first treatment group (Control group) received no further information, and therefore received the minimum amount of information required to complete the capital investment task. The second group (Forecast group) received an additional statistical forecast. This treatment group was included because, as mentioned in the literature review, the provision of a point forecast is commonly used as an indication of future performance in business. Finally, the third group (UI/F group) received the statistical forecast as well as UI in the form of a 75% confidence level. Thus the three treatment groups included a control (Control group), a treatment representing prevalent business practice (Forecast group), and an experimental group for evaluating the utility of UI (UI/F group). All information was displayed graphically.

---

[7] This task can be thought of as a capital budgeting sub-task. The estimation of sales would be a typical capital budgeting sub-task. It was felt that the evaluation of sales for a particular project would be a simpler and less abstract task for participants, rather than getting them to compare a project's estimated return with a hurdle rate. The target level for sales to be used in this task can be thought of as the level of sales required to achieve the company's capital investment hurdle rate.

In addition to the three treatment groups varying between subjects, the position of the target level was varied as a within-subject factor along the distribution of possible outcomes, and therefore moved relative to the position of the forecast (the mean of the distribution of possible outcomes) and the 75% confidence level (positioned at the lower quartile mark).

To complete the experiment, participants simply needed to make twenty yes-no decisions for each of the production plants. It was explained that these investments were of a magnitude large enough to be of significant importance (and risk) to the organisation that participants were advising. Accordingly, in all cases participants were told to recommend increasing production capacity only when they were “very sure” (specifically, at least 75% sure) that future sales would exceed the target level.

A copy of the instructions provided to participants (Appendix A) and an example from the research instrument (Appendix B) are available in the appendices.[8] The remainder of this section is divided into two parts. Section 3.3.1 provides more details on how the research instrument was designed, while section 3.3.2 describes the logistics of how the experiment was carried out.

3.3.1 Research Instrument

The design of the research instrument can be thought of in terms of physical design and informational design. First the physical features of the

---

[8] A more complex dependent variable was abandoned, because participants were unable to fully comprehend it. This discarded loss function is mentioned in Appendix A and B.

research instrument will be briefly summarised, then the design and development of the information cues will be discussed in greater detail.

One criticism of some of the prior studies into UI utility is their use of abstract tasks, lacking external validity. Significant effort was put into making this experiment's decision-task realistic, with high ecological validity. This required the combination of a realistic scenario (described in section 3.3) and presenting the task in a manner likely to be encountered in practice. Several decisions were made to determine the appropriate physical design of the research instrument, specifically the format for presenting the experiment's information cues, and the vehicle for delivering the research instrument.

The presence and availability of graphical information in business has been greatly increased by the ease with which it can be produced by modern personal computers. In short, it is common for graphical information to be provided and requested in business and therefore a graphical format was chosen to add realism to the decision-task. In addition to practical matters, there are also good theoretical reasons for utilising a graphical format. Graphical displays have been shown to provide judgment advantages for data with trend (Harvey & Bolger, 1996), for data that is non-seasonal and has less noise (Lawrence, Edmundson & O'Connor, 1985), and when judgments are being made over the short-term (Lawrence, Edmundson & O'Connor, 1985, 1986). It will shortly be explained that the data and task in this experiment conform to all of these characteristics.

It was decided that a paper-based handout would be used to collect responses, rather than a computerised interface. There were two primary reasons for this. Firstly, even though computers are used to produce much information in business, inevitably individuals analysing a printed report make the decisions. The use of a paper-based instrument therefore adds to the realism of the task. Secondly, a paper-based instrument allows the capture of unstructured information that may provide clues as to how individuals completed the task. For example, subjects may draw on a paper-based diagram when thinking about how to complete a decision-task, and their doodling may convey more than a verbal explanation of their decision-strategy.

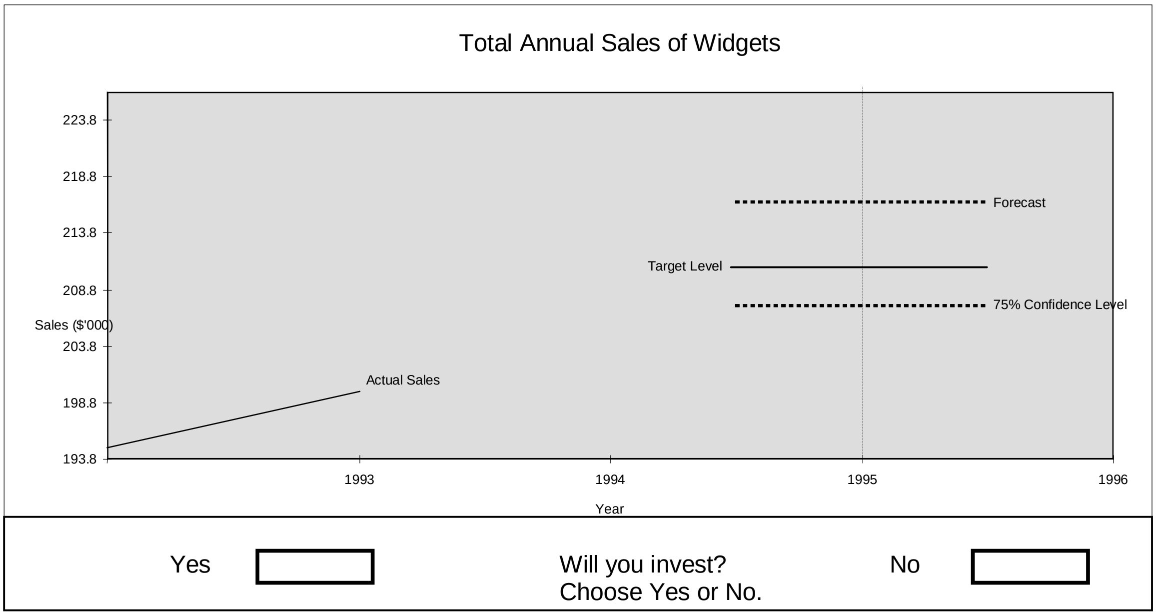
By selecting a graphical, paper-based instrument the physical design for Experiment 1 was complete. Informational design was more complex, and will be explained in greater detail. First, an example of the research instrument will be presented and described, and then each element of the instrument will be further explained.

Figure 3-4 shows an illustration of the information cues used by the UI/F treatment group, the only group to receive all information cues (an exact replica of the instrument is provided in Appendix B). It can be seen that there are four information cues: 'Actual Sales', 'Target Level', 'Forecast' and '75% Confidence Level'.[9] 'Actual Sales' is an historical information cue while the remaining information cues all relate to the forecast horizon. The experiment took place in 1994, explaining why the historical data only went up to 1993. The forecast horizon was the following year, 1995.

---

[9] The 75% confidence level was inadvertently labelled '75% Confidence Interval' in Experiment 1 (see Appendix A and Appendix B). This was corrected in Experiments 2 and 3, and results indicate that this did not influence the findings of this study.

Table 3-1 outlines the distribution of information cues between the three treatment groups. Regardless of treatment group the decision-task remained the same: deciding whether to invest in increased production capacity by determining whether ‘Actual Sales’ would end up above the ‘Target Level’.



Yes ☐ Will you invest? Choose Yes or No. No ☐

Figure 3-4: An illustration of the research instrument used in Experiment 1 for the UI/F treatment group.

All of the information cues were derived from the same set of data drawn from the M-Competition time series (Makridakis et al., 1982). Real time series from the M-Competition were used to maximise the external validity of the experiment. Only quarterly time series were used, and those with seasonal movements were deseasonalised. Quarterly series were used because they had adequate series length to help calculate the UI. There are no issues with presenting these M-Competition quarterly series as annual series in the experiment since the quarterly series were non-seasonal or deseasonalised.

	Information Cues Provided			
Treatment group	**Actual Sales**	**Target Level**	**Forecast**	**75% Confidence Level**
Control	✓	✓		
Forecast	✓	✓	✓	
UI/F	✓	✓	✓	✓

Table 3-1: Distribution of information cues across treatment groups.

A sequence of steps was followed to develop the information cues and select the final twenty time series used in this experiment (see Table 3-2). Information cue development commenced with selecting all quarterly time series from the M-Competition, which had at least 30 history points (necessary for the development of the confidence level) and had a final segment with positive trend. Positive trend was desired because previous studies have found people to have difficulties with downward-sloping series. For example O'Connor, Remus & Griggs (1997) found a tendency to anticipate the reversal of downward series, but not of upward series.

The historical data for each series was truncated at 30 data points. All seasonal time series were deseasonalised. Only those series with a positive last segment slope were selected to provide a higher level of task consistency. Previous research shows that this last segment slope heavily influences decision-makers (Lawrence & O'Connor, 1992). The last segment slope was not necessarily indicative of the trend in the time series' preceding history.

Development of Information Cues			
Step	**Information Cue**	**Definition**	**Number of possible values**
Step 1	Actuals	Select all quarterly series from M-Competition with at least 30 data points, and a final segment with positive trend.	
Step 2	Actuals	Truncate all time series to 30 data points.	
Step 3	Actuals	Deseasonalise all seasonal data.	
Step 4	Forecast	Forecast produced for the second period in the forecast horizon for all selected time series using Holt's Linear Exponential Smoothing	
Step 5	75% Confidence Level	75% confidence level produced from the distribution of forecast errors.	
Step 6	Target Level	Four target levels were defined at the .03, .19, .32 and .39 positions on the distribution of possible outcomes. NB. Forecast is at the .50 position and 75% confidence level is at the .25 position.	4
Step 7	Actuals	The last two points in each time series were used as Actual Sales.	
Step 8	Actuals	A selection of 20 time series that were without significant structural change.	20
			= 80 permutations

Table 3-2: The sequence of steps used to generate the Experiment 1 information cues.

Statistical forecasts were produced for the second period in the forecast horizon for all selected time series using Holt's Linear Exponential Smoothing.

The forecast errors generated by the forecasting technique were ranked and used to statistically develop the 75% confidence level. A confidence level was chosen as the medium for communicating UI since it lends itself towards comparing with a target level. An interval is unnecessary in this situation, and it is known that people prefer point estimates (Einhorn & Hogarth, 1985; Yaniv & Foster, 1997; Kennedy, Mitchell & Sefcik, 1998; Hirst, Koonce & Miller 1999)[10]. Lower confidence levels are a relatively common and robust measure of uncertainty. For example, they were one of the risk measures evaluated by Cooley, Roenfeldt & Modani (1977), and were shown to have high substitutability with other commonly used risk measures including range, standard deviation, semi-interquartile deviation, mean absolute deviation and semi-variance.

Each time series had four target levels developed so that the impact of, and difference between, the 75% confidence level and target level (CL-TL difference) could be investigated. The four target levels were distributed around the confidence level, with two above and two below, and their positions were also based on the distribution of forecast errors. All were below the forecast (which can be thought of as a 50% confidence level). This is depicted in Figure 3-5, which shows an illustration of the research instrument including the location

---

[10] More recently, and contrary to the ideas of researchers such as Yaniv & Foster (1997), Önkal & Bolger (2004) have found forecasters to prefer 95% prediction intervals most and point forecasts least, regardless of whether the forecasting role was provider or user. It may be that by allowing the forecasters to be involved in first the provision and then the utilisation of several forecasting formats, considerable learning took place in Önkal & Bolger (2004), which increased comfort with the prediction intervals. If Önkal & Bolger (2004) had started with the forecasters being users of forecasts, perhaps results would have been different.

of the four possible target level positions. Figure 3-6 (p. 177) defines where the four target levels are positioned on the distribution of actuals.

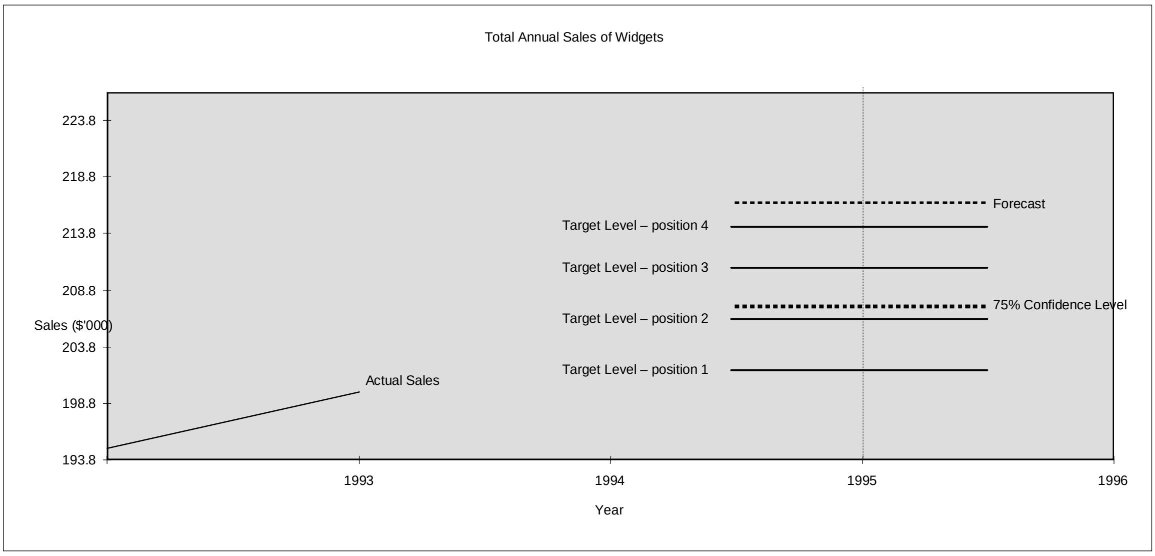


Figure 3-5: Manipulation of the difference between confidence level and target level.

The target levels could be placed at one of four positions on the distribution of actuals. With the forecast being the mean of the distribution (0.5) and the 75% confidence level equivalent to the lower quartile (0.25) the four possible target levels were positioned at .03, .19, .32 and .39 on the distribution of possible outcomes (Figure 3-6). When the target level was below the confidence level the sign of the CL-TL difference was positive, otherwise the sign was negative. The four target level locations also altered the magnitude of the CL-TL difference, thus providing an understanding of the impact of changing the magnitude of this difference from small (the target level being close to the confidence level) and large (the target level being further way from the confidence level).

Having this distribution of information cues created situations where the forecast and confidence level would be conveying conflicting information to those individuals in the UI/F group. When the forecast was above the target level and the confidence level was below the target level it would be difficult for the UI/F group to decide which information cue to trust.

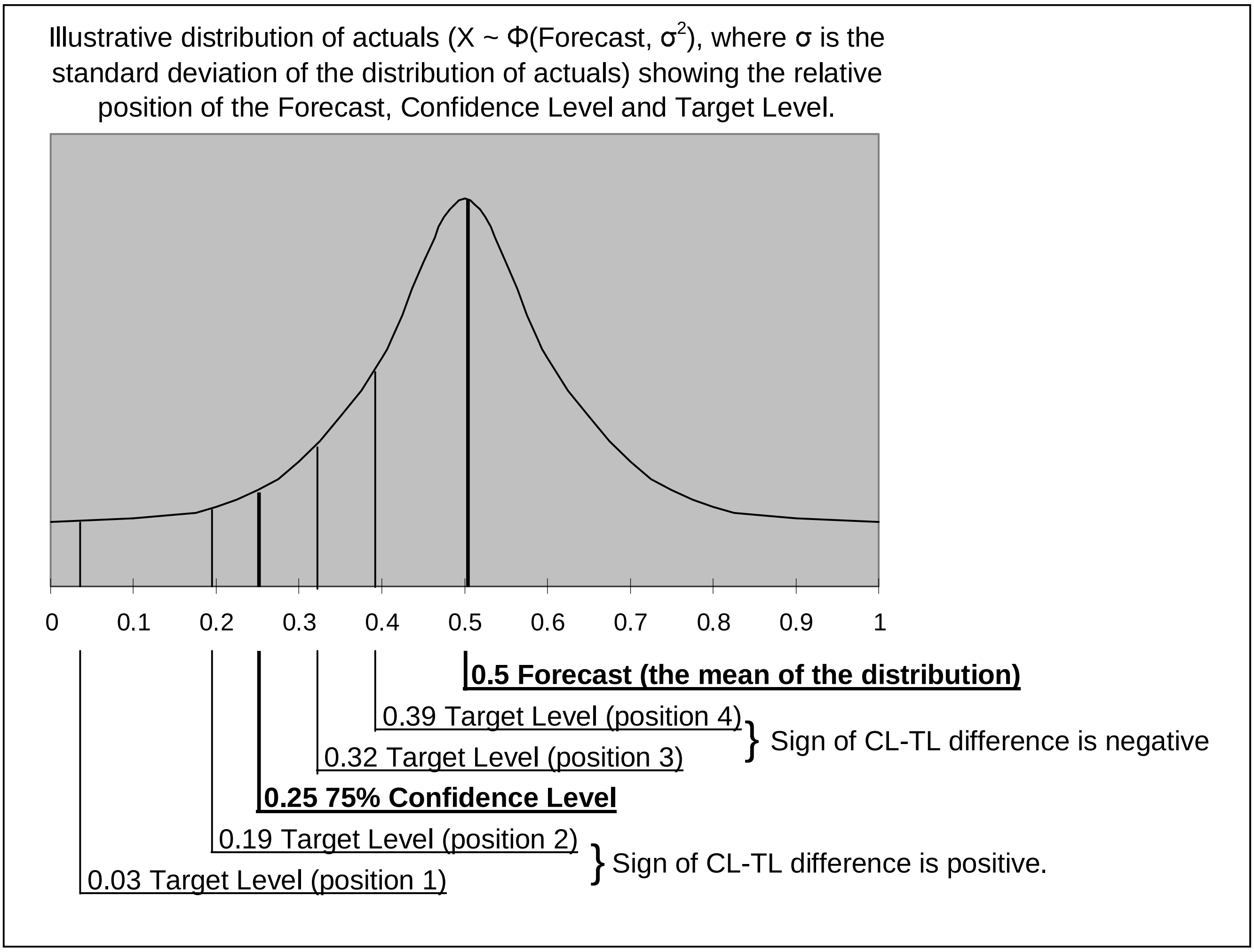


Figure 3-6: The definition of the target level information cue, relative to an illustrative distribution of actuals.[11]

---

[11] Note that Φ is the Normal cumulative density function, and σ is the standard deviation of the forecast errors.

The remaining information cue is ‘Actual Sales’. The last two history points in each time series defined this. Having a limited amount of historical data is typical of many types of business decisions, including capital investment decisions. Two points of history would be thought of as a luxury in many cases, since often no data is available. Limiting the availability of historical data also provided methodological benefits. Providing only two points minimised the availability of a potentially information-poor UI proxy, otherwise participants may have attempted to extract and estimate UI from the historical data rather than utilise the objectively prepared confidence level.

The final twenty time series were selected, based on an eyeball examination, to select those without significant structural change. The aim was to test people under relatively stable conditions to understand their judgments in relation to UI. However, the use of real time series, which are subject to minor changes in underlying structure, increases the ability to generalise the results of this experiment. Although manufactured time series would provide considerable benefits in terms of the internal consistency of the information cues, it is very difficult for them to capture the nuances of real time series. The massaging of the M-Competition data (Makridakis et al., 1982) reduces the benefits of using real time series, however they still contain some of the changing structure found in real life data. The final selection included eleven non-seasonal time series and nine deseasonalised time series.

3.3.2 Experimental Procedure

This section outlines how the first experiment was administered. Key features relating to the participants, feedback, incentives and other experimental characteristics will now be summarised.

Seventy-four graduate business students completed twenty trials of a capital investment decision task. Each trial required the participant to respond ‘yes or ‘no’ to the question of whether the capacity of a particular production plant should be increased, given sales information for the market that that plant supplied.[12]

Post-graduate students studying for a Masters of Commerce at the University of New South Wales were used as participants in the first experiment. Students have been shown to be appropriate surrogates for managers in behavioural decision-making research, since even when there are differences in the decision-making behaviour of managers and students, it appears that the difference is generally in the magnitude of the effect rather than the direction of the effect (Zelditch & Evan, 1962; Ashton & Kramer, 1980; Remus, 1986). Post-graduate students are preferred to under-graduate students due to their greater maturity, and less erratic behaviour (Remus, 1996; Whitecotton, 1996b; Whyte & Sebenius, 1997). Five classes of students studying Information Systems participated in the experiment.

---

[12] An additional eighteen students were used to pilot the task (UI/F treatment group only). The emphasis on instruction changed dramatically between this pilot and the experiment proper. For this reason the pilot’s results were excluded from analysis. It’s interesting to note that the performance of the pilot UI/F group was significantly worse than the experimental UI/F group, reinforcing the importance of training in UI as specified in the contingent model of UI utility.

As mentioned previously, the experiment was run during class, and was paper-based. No time constraints were placed on the completion of the experiment, however the duration of the class presumably acted as a natural constraint. All participants completed the task within approximately 40 minutes. 10 minutes of reading time was provided at the beginning of the task. During this time the instructions included in Appendix A were reviewed. A presentation was then given by the author for a further 10 minutes to emphasise particular aspects of the task. Finally, the research instrument was handed out and the task, answering 20 capital investment decisions, was completed in 10 to 20 minutes. Significant emphasis was placed on explanation, with the instructional part of the task taking up at least 50% of the time.

Only one information treatment was applied within a single class (Control, Forecast or UI/F) so that all participants completed identical tasks to their classmates.[13] However, participants did receive slightly different handouts to their neighbours. There were four different forms of the experimental materials handout. As mentioned previously, twenty time series were selected for the twenty decisions to be made as part of the task. Each time series had four possible target level positions, giving eighty possible time series-target level combinations (see Table 3-2, p. 174). Rather than making participants answer eighty questions, they received only one of the four potential target level positions for each series. This was done to shorten the experiment.

---

[13] Since only one information treatment group was applied to each class, this experiment did not have a 'clean' or 'pure' randomisation of subjects. It was assumed students randomly allocated themselves to classes. There was no indication that there was a difference in ability between classes, and therefore this experiment can be considered quasi-randomised.

The four possible target levels and twenty time series were distributed evenly amongst the four different forms of the experimental handouts, so that participants received each of the time series once, with one of the four possible target levels for each time series, and an equal number of each target level position distributed throughout the handout. That is, each of the four target level positions appeared five times within the twenty time series contained in each of the handouts.

The handouts were distributed to the audience so that neighbouring students had different handouts. Participants were made aware of this as a deterrent against copying. Participants were required to complete the experiment on their own and, while they could ask questions, there was a limit to how much they could be told so as not to influence the decision strategy they chose.

During the instructional presentation several important points were emphasised to participants. The importance of each and every decision to the firm was emphasised, explaining that ruinous losses might result if an unneeded investment in production capacity was carried out. The risk-averse, and asymmetric, nature of the exercise was clearly articulated: participants were told to increase production capacity only if they were “very sure” that sales would exceed the target level. “Very sure” was further specified as being at least 75% sure that the target level would be achieved. Those understanding the confidence level would have immediately seen its potential, however no possible decision strategies were discussed with participants: the determination of their decision strategy was left up to them.

Each of the information cues, relevant to the treatment group in a particular class, was also thoroughly explained. Participants were told: that the limited historical information was due to the product being new; that target levels had been set by senior management; and that forecasts and confidence levels had been calculated by a trusted expert.

As mentioned in the literature review, one of the criticisms of some of the previous investigations into UI utility was the UI’s lack of relevance to the task. Those in the UI/F treatment group in this study had a noticeable potential advantage over other treatment groups, being able to recognise situations where there was a strong chance that the target level would be achieved: that is, when the confidence level was above the target level and therefore there was at least a 75% chance of sales surpassing the target. However, in order to gain this advantage, those provided with UI needed to understand it. As a result, most explanation was reserved for the confidence level, provided only to classes in the UI/F treatment group. The confidence level was the information cue least familiar to the participants, so it was explained graphically by comparing the 75% confidence level to the forecast (see Figure 3-7). It was explained that the forecast could be thought of as a 50% confidence level (assuming a symmetric distribution of forecast errors), with a 50% chance of the actual being above the forecast and a 50% chance of the actual being below the forecast. In contrast, there was a 75% chance that the actual would be above the 75% confidence level and a 25% chance that the actual would be below the 75% confidence level.

Ensuring participants understood the task and the information cues available to them prior to commencing the task was very important, because no feedback was to be provided during completion of this experiment. Davis & Kottemann (1995) showed that describing a decision rule could significantly improve rule following and decision performance when no outcome feedback was available. Herman, Ornstein & Bahrick (1964) in their investigation into UI utility found training to significantly improve decision performance.

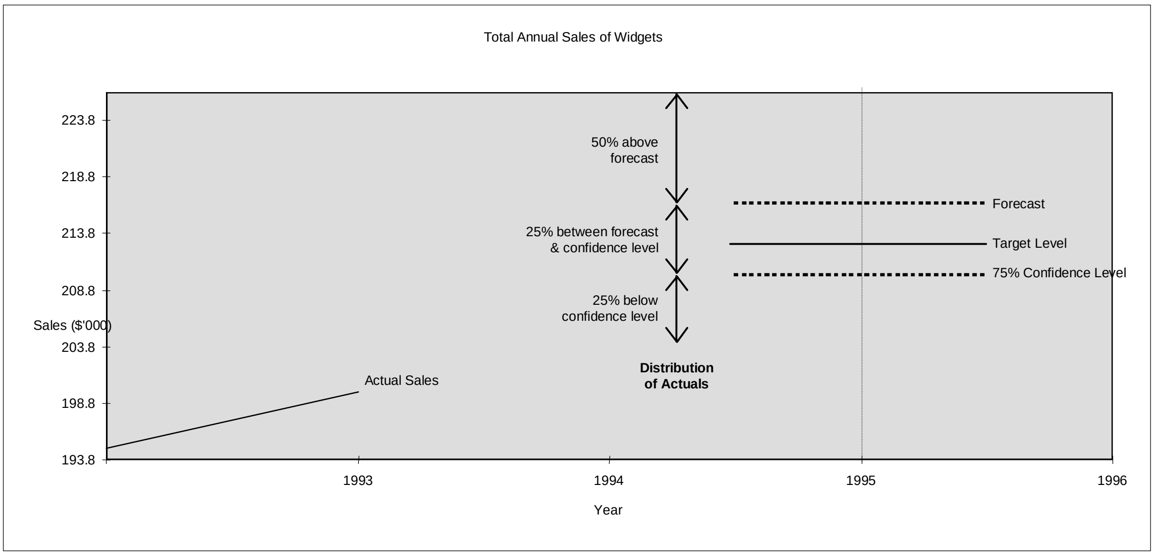


Figure 3-7: Distribution of actual values in relation to information cues.

No feedback was provided to participants during the experiment to prevent learning from experience, and to emphasise the one-off, strategic nature of the task. Such decisions have attracted minimal research (Kofler & Zweifel, 1993).

Feedback is not always beneficial. Sniezek & Reeves (1986) found that the provision of criterion feedback invariably led to overconfidence in judgment, without any increase in performance. Outcome feedback can create inconsistencies in judgment behaviour (Brehmer, 1980). By not providing feedback in this experiment, it was hoped that results would be more consistent (Hagafors & Brehmer, 1983; Hammond & Summers, 1972) with subjects not altering their behaviour, but rather using their own intuition. There is also some evidence that feedback may not be effective in probabilistic inference tasks because the function linking information cues to performance may be too complex (for example, Brehmer, 1974, 1976, 1980; Brehmer & Kuylenstierna, 1978; Hammerton, 1973).

One of the goals of this research was to see whether subjects were able to use UI when their own estimate of uncertainty was flawed or non-existent. This might occur if participants have some biased notion of what a likely range of uncertainty might be. If the task has uncertainty outside the expected range, then the UI will be beneficial. Inaccurate estimates of uncertainty might result from an inability to learn from past mistakes (Johnson, 1982). The task in this research is such a situation, where the one-off nature of the decisions makes it impossible to learn from experience. Such one-off scenarios are typical of strategic business decisions.

Given that there was some need for participants to be trained in order to successfully carry out the task, incentives were used to encourage attentiveness. Monetary incentives were used to encourage diligent performance, while marks for class participation acted as another motivating factor for participating in the experiment. Stone & Ziebart (1995) found performance-contingent monetary incentives to produce better decision strategies, leading to more accurate choices. Foong (1992, 2003) argues that UI utility is wholly contingent upon the availability of performance-based incentives. Although some studies have found incentives to reduce calibration in probability judgments and increase overconfidence (Henry & Sniezek, 1993; Wallsten, Budescu & Zwick, 1993), the meta-analysis reviewing the impact of financial incentives on psychological experiments completed by Camerer & Hogarth (1999) points to the need for financial incentives when investigating the utility of UI, particularly given the results of studies investigating probability judgment where incentives were found to help in the majority of studies. For this experiment the best three performers in each class received prize money, first place receiving AU$50, second place AU$20 and third place AU$10. With less than 30 students in each class the participants had a better than 1 in 10 chance of winning a prize. The prize money was heavily advertised while explaining the task.

This section has detailed a task and research instrument that defines a three (treatment group – between-subject) by four (CL-TL difference – within-subject) factorial design experiment. The task and research instrument were created to hold in check any factors that may have jeopardised the utility of UI in previous research. Many of the factors listed in the contingent model of UI utility (see Figure 2-10 in the literature review) were directly addressed as part of the experimental design. How they are addressed is summarised in Table 3-3. In the next section the results of the first experiment will be presented.

Factors influencing the contingent utility of UI	Importance	Lessons Learnt	Our Understanding	How addressed in Experiment 1
Task				
• UI format	◕	Midpoint of a range can be used.	◑	Confidence level used rather than confidence interval.
• Presence of UI proxies	◕	UI can be ignored.	◔	UI proxies minimised.
• Level of task uncertainty	◕	UI beneficial at high uncertainty.	◔	NA
• UI adds obvious value	◕	Otherwise, suboptimal strategies.	◔	A task where UI is required in order to execute optimal strategy.
Person				
• Task training	◕	Increases UI understanding and use.	◑	Significant UI training provided to those receiving UI.
• Expertise/Experience	◔	May reduce UI utility.	◔	One-off decisions reduce likelihood of anyone having expertise.
Environment				
• Payoffs/Incentives	◑	Appear necessary for UI use.	◕	Monetary incentives provided to best performers.
• Feedback	◑	Possibly confusing with UI.	◑	One-off decision scenario with no feedback.
• Time pressure	◔	May reduce UI utility.	◔	No time pressure was applied to participants.
• Realistic task context/story	◔	Required for UI utilisation.	○	Use of an ecologically valid task.
• Loss function	◑	UI utility when loss function is asymmetric.	◔	Asymmetric, risk-averse, loss function applied.
Interactions				
• Intuitive expectations of risk	◕	UI may be useful when intuition and reality differ.	◑	NA
• Individual perceptions of risk	◑	UI beneficial when uncertainty perceived as high.	◑	NA
• Calibration of confidence	◔	Calibration may influence belief in UI.	◔	NA
• Probabilistic Ambiguity	◔	Point estimates preferred.	◔	UI provided in the form of a point estimate (confidence level).
Risky Behaviour				
• Losses weigh heavy	◕	Losses increase concentration.	◑	Risk-averse situation, with considerable theoretical losses.
• Avoid risks where target level jeopardised	◕	Target levels increase UI salience.	◑	Target levels included as a point of comparison for UI.
• Avoid ruinous losses	◔	Risk-averse behaviour with large losses.	◔	Integrated into the task to see if UI could help in such situations
• Risk seeking in threatening situations	◔	Threats may alter decision strategies.	◔	NA

Key:

Importance	The importance of that factor to understanding UI utility, combining normative and descriptive importance
Our Understanding	The extent to which this factor is understood: a combination of the number and strength of contributions
NA	Not addressed in Experiment 1
○	Very Low
◔	Low
◑	Medium
◕	High
●	Very High

Table 3-3: How each factor in the contingent model of UI utility was addressed in Experiment 1.

3.4 RESULTS

Key Messages:

- UI significantly improved overall decision performance (H_1 supported).
- This improvement in decision performance by those receiving UI was due to them being better able to adjust decision behaviour when the target level went from below to above the 75% confidence level.
- UI provided significantly greater relative decision performance as separation of the target level and 75% confidence level increased. (H_2 supported).

3.4.1 Analysis Methodology

How does one assess the goodness of decisions in a task where the objective is to be at least 75% sure? It would be inadequate just to investigate whether decisions were correct or not, relative to the actual. The actual may or may not agree with the UI, and therefore cannot be trusted to reward rational judgment in a one-off situation. The ideal way to evaluate decisions would be to simulate one hundred responses for each situation and determine whether each decision-maker is at least 75% correct. Since we don't have one hundred responses (we only have the one actual) we're going to construct a naïve policy to evaluate decisions. Our inferences about decision performance will be based on this naïve policy.

Two variables were used to measure results. First, whether the participant's response was 'yes' or 'no' defined decision behaviour. Decision performance was the second dependent variable, and was determined by comparing participants' responses with a naïve policy. The 'naïve' policy responds 'yes' when the 75% confidence level is greater than the target level, and 'no' otherwise. While this experiment used real time series to increase the external validity of results, the aim of this investigation is to understand the rationale of the judgments made by participants, rather than the correctness of their judgments or the accuracy of the information cues. Comparing participants' responses against a naïve policy, rather than against the actual forecast horizon values of only twenty time series, is equivalent to evaluating the participants against thousands of simulated actual results for these time series and therefore provides greater insight into the rationality of their judgments.[14,15] However, some limitations of this approach should be pointed out. First, and foremost, evaluating decision performance using a naïve policy is limited by the fact that real actuals are not used. Second, using the naïve policy to assess decision performance suffers from an historical bias, because the confidence level is constructed from the empirical distribution of historical forecast errors. The "real" decision performance may not be the same as that determined by the naïve policy, because the time series' behaviour may not be stable, and

---

[14] If decision performance were to be calculated using the forecast horizon's actual, rather than a naïve policy's responses, then a correct response would have been determined by where the actual appeared relative to the target level. If the actual ended up being equal to or above the target level, the decision would have been correct, and incorrect otherwise. Such an analysis could then be further dissected to look at only 'yes' decisions in order to confirm whether participants adhered to the task guidelines of being 'at least 75% sure' that their investments would be successful. This would be a valid extension to the current work.

[15] As mentioned earlier (and referred to in Appendices A and B), a more complex dependent variable was abandoned, because participants were unable to fully comprehend it.

therefore the confidence level may not be accurate. Both of these limitations reduce the ability to generalise results.

Due to the binary nature of the dependent variables, non-parametric analysis was used, with parametric analysis also used on some data transformations when appropriate data characteristics existed. Analysis of results is focussed on the overall impact of UI and the influence of the sign and magnitude of the CL-TL difference, but only to the extent that it influences the impact of UI.[16]

3.4.2 Analysis Results

To test H_1 the decision performance of the UI/F group was compared to that of the other treatment groups. Decision performance was determined by comparing responses with a naïve policy. UI had a significant impact upon overall performance. Figure 3-8 clearly shows the impact of the UI on overall decision performance, with a Kruskal Wallis test confirming the existence of a difference between the performances of the treatment groups ($\chi 2 = 34.605$, $p < .001$).

The UI/F group made a significantly higher percentage of correct investment decisions than the next best group (Mann-Whitney U test, Z = -

---

[16] It should be noted that two characteristics (sign and magnitude) of the within-subject factor, CL-TL difference, are being investigated. Sign is a categorical characteristic with two possible values: positive and negative. Magnitude is a scalar factor, which has four values in this experiment.

4.879, p < .001).[17] Thus, the provision of explicit, objective UI that had diagnostic or decision-making value improved decision performance and H_1 is supported.

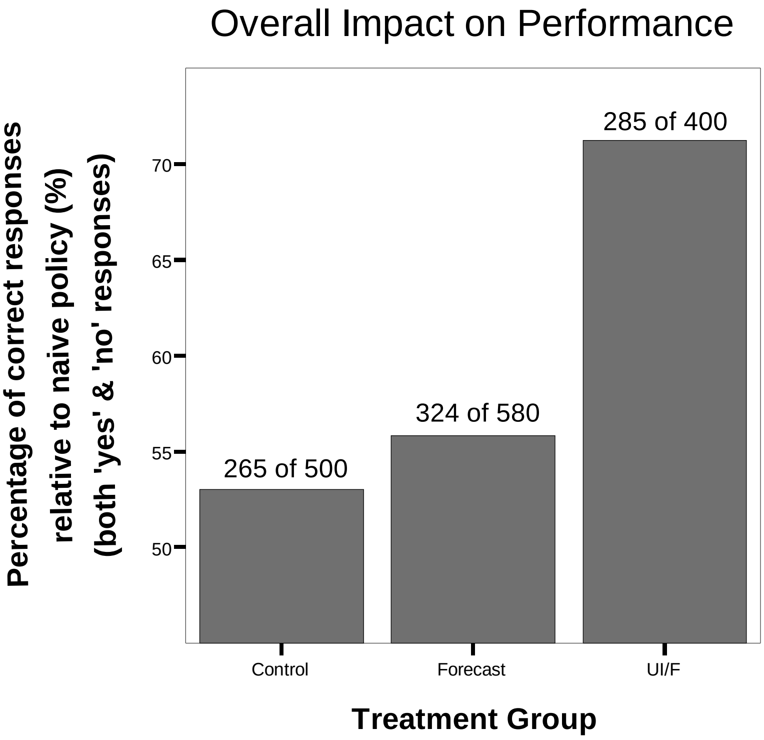


Figure 3-8: The percentage of correct investment responses, by treatment group (Kruskal Wallis test, $\chi 2 = 34.605, p < .001$).

This overall improvement in decision performance by those receiving UI was further investigated by considering the decision behaviour of the various treatment groups as the position of the target level changed relative to the 75% confidence level. CL-TL difference significantly impacted the extent to which UI influenced decision behaviour. While the behaviour of the Control and Forecast groups was relatively stable, the behaviour of the UI/F group was highly contingent upon CL-TL difference (see Figure 3-9).

---

[17] Due to the quasi-randomised design of this experiment it can be assumed that the better performance of the UI/F treatment group was due to the provision of UI and not due to any difference in the participants that made up the UI/F treatment group.

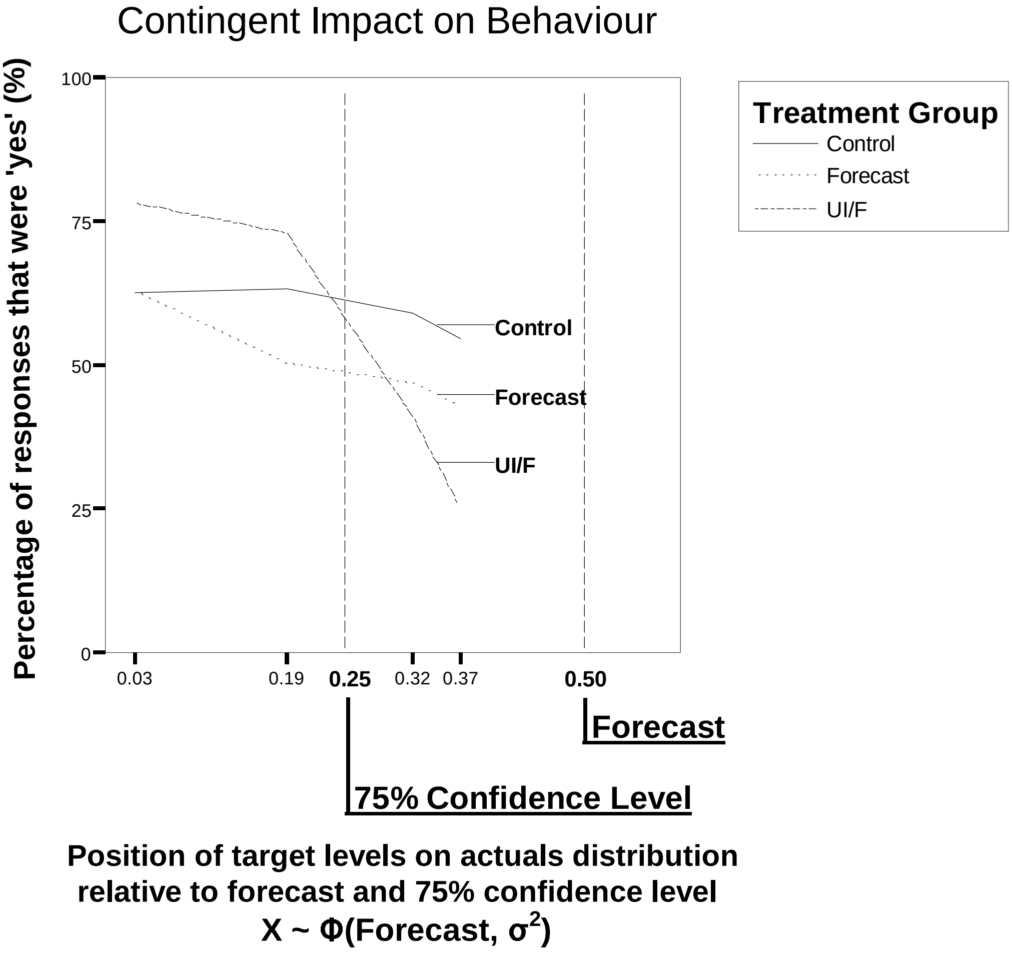


Figure 3-9: Decision behaviour and the interaction of CL-TL difference with treatment group.

Figure 3-9 shows the change in the percentage of ‘yes’ responses for each of the treatment groups as the position of the target level moves relative to the 75% confidence level and forecast.[18] As expected, the responses of the UI/F group were skewed towards ‘yes’ when the CL-TL difference was positive (the 75% confidence level was above the target level, and therefore there was a greater than 75% chance of success) and towards ‘no’ when the CL-TL

---

[18] It should be noted that the naïve policy switches from saying ‘yes’ always when the target level is below the confidence level to always saying ‘no’ when the target level is above the confidence level. If the naïve policy were to be included in Figure 3-9 it would be constantly at 100 to the left of the 75% confidence level (the 0.25 mark on the horizontal axis) and constantly zero to the right of the 75% confidence level.

difference was negative. The mean drop in the percentage of ‘yes’ responses as the CL-TL difference moved from positive (TL = 0.03 or 0.19) to negative (TL = 0.32 or 0.37) was analysed for each treatment group. The Control group had a mean drop in the percentage of ‘yes’ responses as the CL-TL difference moved from positive to negative of just 5.8%, the Forecast group 11.6% and the UI/F group of 42.5%. Analysis of variance confirmed the existence of a significant interaction between the sign of the CL-TL difference and treatment group regarding decision behaviour (F = 37.439, p < .001).[19]

Post hoc multiple comparisons, using the Games-Howell procedure (Games & Howell, 1976), revealed the treatment groups formed two distinct homogenous subsets: those with and without UI. The decision behaviour of the UI/F group was significantly more impacted by a change in the sign of the CL-TL difference than was the Forecast group (p < .001).[20] Those with UI were better able to adjust decision behaviour as the sign of the CL-TL difference changed, and this provided them with better overall decision performance.

---

[19] A data transformation was used to calculate the mean drop in ‘yes’ responses when moving from a positive CL-TL difference to a negative CL-TL difference. This involved subtracting the mean proportion of ‘yes’ responses for the negative CL-TL difference for a particular treatment group from the mean proportion of ‘yes’ responses for the positive CL-TL difference for the same treatment group. Consequently the distributions for the final variables differed. Non-parametric analysis assumes that data from each group is similarly distributed, and therefore non-parametric analysis could not be used. One-way analysis of variance was used instead. Analysis of variance is fairly robust to violations of normalcy when sample sizes are relatively large (Diekhoff, 1992) and approximately equal in size.

[20] Variances for the treatment groups were not homogenous, so a post hoc multiple comparison procedure that did not assume equal variances was required. The Games-Howell procedure was chosen. If variances are known to be unequal, then this procedure is more powerful than others such as Dunnett’s C test or Dunnett’s T3 test (Rafter, Abell & Braselton, 2002).

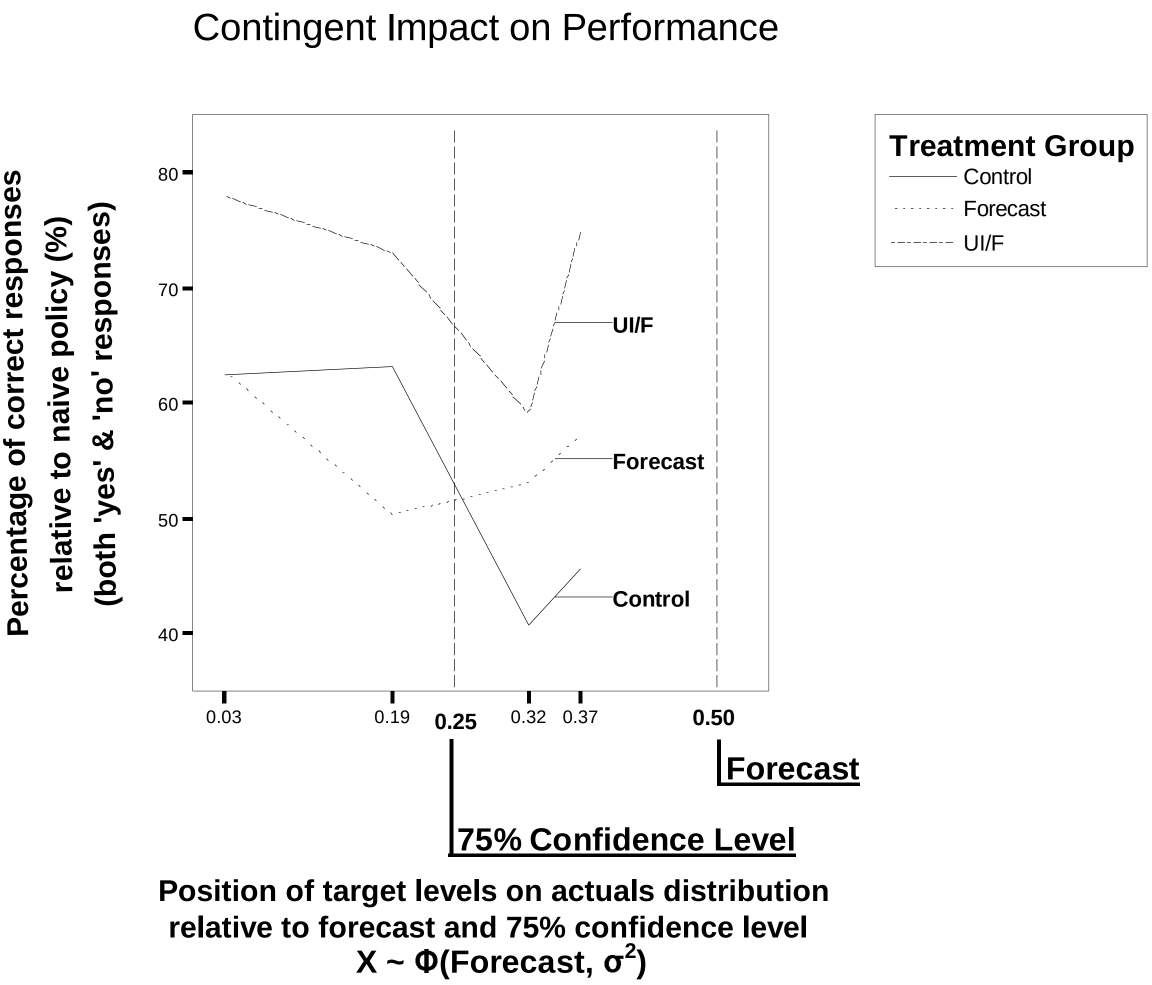


Figure 3-10: Decision performance and the interaction of CL-TL difference with treatment group.[21]

How the magnitude of the separation between the target level and the 75% confidence level influenced decision performance is shown in Figure 3-10. To test H_2 the decision performance of the UI/F group was compared to the next best group at each target level position. Again, decision performance was determined by comparing responses with a naïve policy. When the target level was close to the 75% confidence level position (TL = 0.19 or 0.32) there was no

---

[21] The astute reader will notice that the left half of Figure 3-10 is identical to the left half of Figure 3-9, apart from a change in scale. The left halves are identical because the naïve policy (used to determine the decision performance reported in Figure 3-10) grades all ‘yes’ responses as correct when the target level is less than the 75% confidence level.

significant difference between the decision performance of the UI/F group and the next best group. However, as reported in Table 3-4 the UI/F group had significantly higher decision performance when the target level was far below the 75% confidence level (TL = 0.03: Mann-Whitney U test, Z = -2.516, p < .05) and when the target level was far above the 75% confidence level (TL = 0.37: Mann-Whitney U test, Z = -2.849, p < .005). UI provided greater relative decision performance as separation of the target level and 75% confidence level increased, thus supporting H_2.

	Position of Target Level			
	0.03	**0.19**	**0.32**	**0.37**
Significance of difference between decision performance of **UI/F** group and best non-UI group	P < .05	Not significant	Not significant	P < .005

Table 3-4: Significance of difference between the decision performance of the UI/F group and the best non-UI group (results of Mann-Whitney U comparisons).

3.5 DISCUSSION

Key Messages:

- The overall impact on performance by UI is expected, given that this experiment was specifically designed using a contingent model of UI utility so that UI provided significant value to the decision-maker.
- The ability of those with UI to better adjust decision behaviour as the target level went from below to above the 75% confidence level is to be expected, since this must be a cue that directly influences the UI/F group's decisions.
- The inability of those with UI to outperform the other treatment groups when separation of the target level and 75% confidence level was minimal may have been due to: the range created by the forecast and confidence level indicating that the decision objective was being threatened; a perception that the confidence levels contained some degree of error; or, an indifference to the situation.
- These results raise further questions regarding UI utility, which can only be answered by extending this research to address issues such as providing the UI without a forecast.

UI significantly impacted both decision behaviour and decision performance. UI improved overall decision performance, because those with UI were better able to adjust their decision behaviour when the sign of the CL-TL difference changed. UI also improved relative decision performance as separation of the target level and confidence level increased. These results will be discussed, and then followed by an identification of research topics to be investigated in a second experiment.

The impact of UI on overall decision performance

In Figure 3-8 it can be seen that those individuals with UI significantly outperformed those in other treatment groups.[22] Only three (out of a total of eleven) other studies into UI utility have found an overall effect on either performance or behaviour due to UI (see Table 2-1 to Table 2-7 in the literature review). Yet, despite this sparsity of previous results, this result was expected since the experiment was specifically designed using a contingent model of UI utility so that UI provided significant value to the decision-maker.

Eckel (1983) found probabilistic information changed decision behaviour and improved performance in an experimental game with the objective of maximising the profits of a hypothetical firm. Hawkins (1974) found the provision of UI significantly altered decision behaviour in an investment allocation task, a task similar to that later advocated by Kennedy (1999, p. 132) in her constructive criticism of the study by Hirst, Koonce & Miller (1999), which did not find an overall effect due to UI. Herman, Ornstein & Bahrick (1964) carried out a visual search and attack game and showed discrete probabilistic displays (like confidence intervals) improved object location judgments.

Given that when making decisions under uncertainty people exhibit systematic biases and a limited information processing capacity (Hogarth,

---

[22] Again, note that performance was measured by comparing judgments against a naïve policy, which responds ‘yes’ when the target level is above the 75% confidence level (when there is at least a 75% chance of success) and ‘no’ otherwise.

1975), it should not be surprising that people benefit from information that reduces the processing load placed on a decision-maker and helps in assessing uncertainty. Providing UI to decision-makers has been suggested by a number of researchers for improving human probability assessment (Mohn & Reid, 1977; Remus & Simkin, 1987; Makridakis, 1988; Kydd, 1989).

The significant results of this study may be a confirmation of the contingent model of UI utility outlined in the literature review, since this experiment was holistically designed based on the model. However, to say that the overall result is due to the embryonic contingent model of UI utility is highly speculative and any such claim is premature. Further analysis is required to better understand the role of each part of the contingent model of UI utility, such as whether the removal of UI proxies in this experiment did improve UI utilisation and utility. Since the reasons for the overall result cannot be easily pinpointed, of much greater interest are the results contingent upon the sign and the magnitude of the CL-TL difference.

How the UI/F group were better able to adjust their decision behaviour as the sign of the CL-TL difference changed

Those participants in the UI/F group had better overall decision performance, because they were better able to adjust their decision behaviour as the sign of the CL-TL difference changed. While the target level was below the 75% confidence level the UI/F group responded ‘yes’ more than the other groups, and when the target level was above the 75% confidence level the UI/F group responded ‘no’ more than the other groups.

Participants were told to be “very sure” if they were going to answer ‘yes’ and increase the production capacity of a particular plant. Those with UI could be “very sure” by comparing the 75% confidence level with the target level. This decision strategy was only available to them. The proximity of the confidence level and target level to the UI/F treatment group was therefore crucial and directly influenced their decisions, but only because they were told to be risk-averse. Presumably, if they had not been told to be risk-averse, then the forecast would have been of greater comparison value than the 75% confidence level.

This change in behaviour due to the sign of the CL-TL difference was not as great for those treatment groups without UI. The Control group received only history and a target level to compare. The Forecast group received history, the target level and a forecast, but the target level was always below the forecast. In contrast, for the UI/F group, half the time the target level was above the confidence level and half the time it was below the confidence level (see Figure 3-5) so there was a need to think about what the change meant. The steep shape of the UI/F group’s decision behaviour as the position of the target level changes, depicted in Figure 3-9, clearly shows that whether the target level was above or below the confidence level strongly influenced decision behaviour.

Those participants in the UI/F group behaved similarly to the ‘naïve’ policy, responding ‘yes’ whenever the 75% confidence level was above the target level, but not exactly the same. When the confidence level was above the target level the UI/F group tended to respond ‘yes’ approximately 80% of the time. In this situation the ‘naïve’ policy always responds ‘yes’. Perhaps when both the confidence level and forecast were above the target level the UI/F group were encouraged to respond ‘yes’. When the confidence level was below the target level a more complex decision situation arose: the forecast was still above the target level. The confidence level and forecast were providing conflicting information as to whether the target level would be achieved. The UI/F group behaved somewhere between the Forecast group, by trusting the forecast rather than the confidence level, and the naïve policy. In this situation the ‘naïve’ policy responds ‘no’.[23]

It is not clear why the UI/F treatment group did not behave exactly like the ‘naïve’ policy, which would have improved their performance. Perhaps subjects did not comprehend the UI, the UI was not trusted or other information cues were favoured over the UI. Further investigation is required to explain the behaviour of the UI/F group. Whether the UI was not trusted will be discussed more in the next section.

While the UI/F group was not following the naïve policy exactly, there is some pattern to its decision behaviour. With the scenario provided to participants, it’s not unreasonable to assume that the UI/F group would have responded ‘yes’ approximately 50% of the time when the target level was near the 75% confidence level; probably wouldn’t have ever responded ‘yes’ when the target level was above the forecast (although this situation did not eventuate) and would have responded ‘yes’ approaching 100% of the time as the target level got further below the 75% confidence level.

---

[23] It should be noted that behaviour was not influenced by the individual time series used in this experiment. Behaviour by all treatment groups was consistent across different time series.

In a forecasting task, Bolger & Harvey (1995) confirmed that when asked to estimate the probability that the true value is below or above a reference point (a task very similar to the task used in this study, except for a lack of UI) people used the centre of the probability scale as an anchor. The researchers found that insufficient adjustment away from this anchor led to judgmental probability distributions that were too flat. Bolger & Harvey (1995) refer to this lack of adequate adjustment away from the centre of the probability scale as hypoprecision. Hypoprecision indicates underconfidence in one's knowledge of a task.

The task in this study, although only asking for a 'yes' or 'no' response is essentially a probability judgment like the one in Bolger & Harvey (1995), since in order to complete the task one needs to only invest (that is, say 'yes') when one is 75% sure of success, and therefore participants needed to determine whether there was a 75% chance of achieving the target level. Figure 3-9 shows that the Control and Forecast groups were unable to adjust away from the centre of the probability scale, and as a result they were hypoprecise. That is, their probability distributions were too flat. This result confirms the findings of Bolger & Harvey (1995), and contrasts with the results of Seaver et al. (1978), who did not find evidence of hypoprecision for probability judgments in a knowledge retrieval task. The Control and Forecast groups were hypoprecise relative to the naïve policy, which answered 'yes' always when the target level was less than the 75% confidence level, and 'no' otherwise. By anchoring on the centre of the probability scale, the Control and Forecast groups overestimated the likelihood of success when it was less than 50% (target level at the 0.32 and 0.37 positions) and underestimated the likelihood of success when it was greater than 50% (target level at 0.03 and 0.19 positions).

The behaviour of the UI/F group was also hypoprecise relative to the naïve policy, which is not surprising given the shape of the naïve policy’s probability distribution, but less so than the Control and Forecast groups. The provision of UI improved the calibration of the judgments. Additional target levels positioned closer to the forecast are required to confirm whether the UI/F group was hypoprecise at the extremities of the distribution of the likelihood of success.

Ho, Keller & Keltyka (2002) investigated decision behaviour using a similar decision task to that used in Experiment 1. They explored how decision framing influenced outcome ambiguity. Their work concerned the relative locations of a point forecast, a corresponding surrounding symmetric range forecast and a target level. When the point forecast was above the target level, and the range forecast straddled the target level (a gain condition since the target level is expected to be achieved) the forecast was preferred (participants were outcome ambiguity averse – they avoided the range). When the point forecast was below the target level (a loss condition since the target level is not expected to be achieved) with the range still straddling the target level, the range forecast was preferred (participants were outcome ambiguity seeking – they preferred the ambiguous range). The project managers participating in Ho, Keller & Keltyka (2002) believed whatever information gave them the best chance of achieving the target level.

The relevant message from Ho, Keller & Keltyka (2002) is that the form of the UI, and its location relative to a target level, can influence its attractiveness and utility to the decision maker. While this first experiment investigated the behaviour of a treatment group receiving UI/F, there was no investigation of a group receiving the UI without a forecast. The results of Ho, Keller & Keltyka (2002) tell us that such a change could have significant impact on the way the UI is utilised, and this may be an area for future investigation.[24]

Behaviour observed in this study can be better understood by reviewing earlier studies in decision-making. Several authors have observed that business executives define risk as the prospect of not being able to achieve a target level (for example, Mao, 1970; Conrath, 1973; Greer, 1974; Laughhunn, Payne & Crum, 1980; Crum, Laughhunn & Payne, 1981; Ho & Vera-Muñoz, 2001; March & Shapira, 1987). Some studies have also found people to weight the probability and magnitude of loss more than that of gain (Slovic, 1967; Alderfer & Bierman, 1970; MacCrimmon & Wehrung, 1984, 1986; March & Shapira, 1987). This experiment combined both situations, with participants willing to risk the potential losses associated with capital investment only when they thought there was at least a 75% chance the target level would be achieved. In this task risk and loss would be associated with investing in a plant and the target level not being achieved. It would therefore be assumed that the natural response to this experiment would be to not invest unless one was very sure. The natural risk-averse attitude towards target levels and losses was reflected in the risk-averse situation described to participants as part of the decision task.

---

[24] It would be interesting to repeat the study of Ho, Keller & Keltyka (2002), but replacing the range forecast with a lower limit or upper limit to see whether their results were due to outcome ambiguity or simply the placement of the range limits relative to the target level. An alternative modification would be to apply an asymmetric loss function, which favoured risk-averse behaviour. This would greatly increase the value of the range's lower limit.

When forecast and confidence level were both above the target, actual and perceived chances of achieving the target level were maximised. The weight of both information cues above the target level was enough to make participants agree to a capital investment despite the risk-averse scenario. When the forecast was above the target level and the confidence level below the target level the risk-averse scenario tilted belief towards the confidence level. While Ho, Keller & Keltyka (2002) were investigating the sign of the difference between the forecast and a target level, this experiment looked at the sign of the difference between the 75% confidence level and a target level. In this experiment UI provided the most value.

Why those with UI were unable to outperform the other treatment groups when separation of the target level and 75% confidence level was minimal

Figure 3-10 and Table 3-4 show that the UI/F group significantly outperformed the other groups when the target level was either far above (target level position 0.37) or far below (target level position 0.03) the confidence level. However, when the confidence and target levels were in close proximity (target level at position 0.19 or 0.32) the decision performance of the UI/F group was no better than the best non-UI treatment group.

It could be argued that the participants were troubled by the closeness of the confidence level to the target level and it is this that reduced their performance. As mentioned previously, it has been observed that business executives define risk as the prospect of not being able to achieve a target level (for example, Mao, 1970; Conrath, 1973; Greer, 1974; Laughhunn, Payne & Crum, 1980; Crum, Laughhunn & Payne, 1981; Ho & Vera-Muñoz, 2001; March

& Shapira, 1987). The UI/F group would have been able to see the location of the target level relative to the range created by the forecast and 75% confidence level, and therefore would have been able to tell when the target level was relatively close to the 75% confidence level. As a result, judgment may have suffered.

If participants had perceived that the confidence levels contained some degree of error, this would have added to the feeling that the decision objective of being 75% correct was being threatened when the target level and confidence level were close. When the target levels were just below the confidence level those with UI who were risk averse may not have been confident that they could be 75% sure that the target level would be achieved, and would only be sure as separation of the target level and confidence level increased. Equally, if the target level was just above the confidence level, those with UI who were risk seeking may have thought there could still be a 75% chance the target level might be achieved, due to any potential error in the calculation of the confidence level.

It may be that rather than the UI/F group having performance detrimentally affected when the target level was close to the 75% confidence level, perhaps the UI/F group was the only group able to significantly improve decision performance as the target level moved away from the 75% confidence level. This interpretation assumes that participants felt that as the target level approached the 75% confidence level it didn’t matter what they did: it was becoming a toss-up.

Participants were indifferent as the target level approached the 75% confidence level, because the yardstick for their performance was to be at least 75% sure of success. As a result, the difference between the outcome of a correct and incorrect answer may have been negligible. In contrast, as separation of the target level and 75% confidence level increased, improvement in relative decision performance by those with UI coincided with when it was most important that a correct decision be made: when there was a large difference between a correct and incorrect decision.

This indifferent behaviour is reflected in Figure 3-9 where the decision behaviour of all groups, including the group with UI, approached a level of 50% ‘yes’ responses as the target level approached the 75% confidence level. While the naïve policy immediately switched from saying ‘yes’ to ‘no’ as the target level went from below to above the 75% confidence level, it is wrong to expect people to change their decisions so sharply.

Results summary

Table 3-5 summarises the findings for Experiment 1. Both hypotheses are confirmed. UI did have an overall impact upon decision performance (Figure 3-8), and accordingly H_1 is confirmed. H_2 is verified since the relative decision performance of those with UI improved as separation of the target level and confidence level increased (Figure 3-10).

	Hypothesis	Result
H_1	UI will improve decision performance.	Confirmed (see Figure 3-8).
H_2	UI will provide greater relative decision performance as separation of a target level and UI (a 75% confidence level) increases.	Confirmed (see Figure 3-10).

Table 3-5: Summary of Experiment 1 results relating to each hypothesis.

Topics for investigation in a second experiment

Results from the first experiment are encouraging, indicating that UI can improve decision performance. The results also provide direction on what needs to be investigated further. There are four main ways in which the current experiment can be extended in order to further investigate UI utility.

First, further experimentation will need to include an additional treatment group. Despite those in the UI/F treatment group following a decision strategy like the ‘naïve’ policy more than any other strategy, the majority of people in the UI/F group did not appear to do this when a strategy analysis was performed. The fact that the ‘naïve’ policy was not followed more needs to be understood. One explanation may be that those with UI had their decisions confounded by the forecast that they were also provided with. There is a case for introducing a fourth treatment group that receives the UI without a forecast, and comparing the behaviour and performance of this fourth group with the UI/F group. This additional treatment group will also help to explain why the relative decision performance of those in the UI/F treatment group was compromised when separation between the target level and 75% confidence level was minimal. If the range created by the forecast and 75% confidence level provided them with a heightened sense of when the target level was relatively close to the confidence level, then this may have increased the UI/F group's perceptions of risk and their judgments may have suffered. The decision performance of a treatment group that receives UI without a forecast may not be as susceptible to changes in the separation of the target level and confidence level, because without a forecast they would be unaware of the location of the target level relative to the confidence level-forecast range.

The UI/F group in Experiment 1 had decision performance influenced by the graphically visual separation of the target level and 75% confidence level relative to the distribution of actuals. Since a group receiving UI without a forecast will not know where the target level is on the distribution of actuals, perhaps this group will have decision performance influenced by the absolute graphically visual separation of the target level and 75% confidence level. This would make sense, given the findings of previous studies that business executives define risk as the prospect of not being able to achieve a target level (Mao, 1970; Conrath, 1973; Greer, 1974; Laughhunn, Payne & Crum, 1980; Crum, Laughhunn & Payne, 1981; Ho & Vera-Muñoz, 2001; March & Shapira, 1987).

Although a treatment group receiving UI without a forecast might not be as susceptible to changes in the separation of the target level and 75% confidence level, such a treatment group should theoretically be impacted by changes in the underlying uncertainty of the data. The UI/F group would have been aware of changes in uncertainty, since the size of the range created by the forecast and 75% confidence level would have altered as the level of uncertainty changed. A treatment group receiving UI without a forecast would be unaware of changes in underlying uncertainty. Consider two situations with a target level just below the 75% confidence level, and with the separation of the target level and confidence level the same in each instance. Now imagine that in the first scenario the forecast is just above the confidence level such that the confidence level is equidistant from the target level and forecast (low uncertainty), while in the second scenario the forecast is far above the confidence level (high uncertainty). A treatment group receiving UI without a forecast, which will be introduced in Experiment 2, would not be able to distinguish these two situations. The second extension to this experiment is the introduction of a new factor that manipulates task uncertainty by altering the variance of the distribution from which the information cues in the forecast horizon come from. The introduction of the task uncertainty factor in Experiment 2 will provide an additional means of investigating how providing a forecast impacts the utility of UI.

Third, modifying the position of the target levels on the distribution of actuals so that one is closer to the forecast will help to better illustrate the hypoprecision (a flatter judgmental probability distribution, anchored on the centre of the distribution) of the treatment groups in their judgments about the likelihood of success (see Figure 3-9).

Finally, risk is known to be associated with whether or not a target level will be achieved, so if the achievement of a target level is threatened and that impacts the ability to accomplish a decision objective (such as being at least 75% sure of success), decision confidence may be affected. As a fourth change, decision confidence should be collected as a dependent perceptual variable in order to see whether providing UI (which indicates the likelihood of success) influences the calibration of decision confidence.

Complementing the current experiment with the additions listed above would progress the understanding of UI utilisation and utility. These extensions, and the questions this experiment has already started to address, shall be the subject of Experiment 2. Hypotheses, experimental design, results and discussion for Experiment 2 are presented in the next chapter.