

CHAPTER 2 : LITERATURE REVIEW

A contingent model for decision-making with uncertainty information

CHAPTER 2 : LITERATURE REVIEW.....	12
2.1 INTRODUCTION.....	13
2.1.1 <i>Theories of Decision-Making Under Risk and Uncertainty</i>	15
2.1.2 <i>Choice Behaviour</i>	23
2.1.2.1 Measures of Risk.....	24
2.1.2.2 Empirical Evidence of Risky Choice Behaviour	29
2.2 FACTORS INFLUENCING UI UTILITY – A MODEL OF CONTINGENT EFFECTS.....	40
2.2.1 <i>Person</i>	41
2.2.1.1 Experience/Ability.....	42
2.2.2 <i>Task</i>	46
2.2.2.1 Information Disclosure	48
2.2.2.1.1 Provision of Uncertainty Information.....	48
2.2.2.1.2 Cue Mix & Information Load	60
2.2.2.2 Task Uncertainty	72
2.2.3 <i>Environment</i>	77
2.2.3.1 Time Pressure.....	78
2.2.3.2 Incentives and Motivation.....	79
2.2.3.3 Task Setting.....	81
2.2.3.4 Feedback/Irreversibility	83
2.2.4 <i>Interactions of Person and Task</i>	87
2.2.4.1 Risk Profile and Perceived Risk.....	88
2.2.4.2 Confidence/Calibration.....	98
2.2.5 <i>A Model of Contingent Effects for Decision-Making under Uncertainty with UI</i>	108
2.3 A SUMMARY OF EMPIRICAL EVIDENCE REGARDING THE USE OF UNCERTAINTY INFORMATION – A REFINED MODEL OF CONTINGENT EFFECTS.....	115
2.3.1 <i>A Refined Model of Contingent Effects for Decision-Making under Uncertainty with UI</i>	149
2.4 RESEARCH QUESTION FOR EXPERIMENT 1	152

2.1 INTRODUCTION

Key Messages:

- Normatively and rationally Uncertainty Information (UI) should improve decisions.
- Empirical results have not conclusively shown the benefits of UI.
- This chapter will produce a model of contingent effects to guide future investigation into UI utility.

Normatively and rationally the provision of Uncertainty Information (UI) should improve the quality of decisions. People's decisions are detrimentally affected by their inability to judge uncertainty effectively (for example, see Hogarth, 1975) and so any indication of the underlying uncertainty inherent in a decision setting should enhance decision performance. In response, the explicit representation of UI, such as in the form of a confidence interval, has been suggested by a number of researchers as a decision aid appropriate for improving decision-making by providing a better indication of uncertainty (Mohn & Reid, 1977; Remus & Simkin, 1987; Makridakis, 1988; Kydd, 1989).

Despite the existence of strong arguments for the value of UI in decision-making, empirical results have failed to conclusively show that UI improves decision performance (for example, Foong, 1992, 2003; Johnson, 1982; Keys, 1978; Oliver, 1972). Results have been inconclusive because the utility of UI appears highly contingent upon task related factors, and there have only been eleven studies in forty years directly investigating this area. A better

understanding of how the provision of UI can benefit decision-makers is required.

The potential utility of objective UI has not attracted the same research effort as the subjective estimation of UI. An enormous amount of effort has been put into understanding the accuracy and validity of people's judgments of confidence. As a result this 'calibration' research has moved towards the development of a contingency model describing what influences confidence in judgment (for example, Ayton & McClelland, 1997; Erev, Wallsten & Budescu, 1994; O'Connor, 1989). Future research aimed at understanding the downstream role of UI, as an input into decisions under uncertainty, requires a similarly rigorous framework.

This chapter attempts to produce such a framework by defining a model of contingent effects for the effective utilisation of UI in decision-making under uncertainty. The model provides guidance on what factors influence whether or not UI will be of value to decision-makers. It will be developed by:

- Investigating all normative factors that influence UI utility relating to person, task and environment (the model prescribed by Beach & Mitchell, 1978) and producing a preliminary model of contingent effects for the effective utilisation of UI in decision-making under uncertainty (section 2.2), and
- Reviewing the inconclusive empirical results from studies investigating the impact of providing UI to decision-makers, comparing these results to the preliminary model and extrapolating to a refined model (section 2.3).

Concluding the chapter are research questions, derived from the model of contingent effects, which will be investigated in Experiment 1. The chapter begins by first reviewing the theoretical foundations of decision-making under risk and uncertainty.

2.1.1 Theories of Decision-Making Under Risk and Uncertainty

Key Messages:

- Decisions are generally made without definite knowledge of outcomes: they are made under risk (with full knowledge of the probabilities of potential outcomes), ignorance (with no knowledge of the likelihood of outcomes) or somewhere in between (uncertainty).
- The majority of decisions are made under uncertainty, and these are the decisions of interest in this study.
- When UI is provided to people making decisions under uncertainty the distinction between decision-making under risk and decision-making under uncertainty is blurred: UI moves the decision towards the domain of risk.
- Decision-making under uncertainty appears to exhibit similar behaviour to decision-making under risk, except people are less sensitive to uncertainty and therefore more prone to departures from expected utility theory.
- Decision-making under uncertainty with UI needs to consider theories of decision-making under both risk and uncertainty.

Decisions are usually made without definite knowledge of outcomes. Some decisions are made with complete ignorance of the relevant outcome

possibilities. Other decisions, such as games of chance, are made with exact knowledge of the distribution of outcome probabilities: these are known as decisions under risk. The majority of decisions lie somewhere between the domains of ignorance and risk: most decisions are made under uncertainty.

Knight (1921) and Keynes (1921) were the first to distinguish between risk and uncertainty. What distinguishes risk from uncertainty is the extent to which the distribution of outcome probabilities is known and measurable. Decisions are made under uncertainty when people do not know the exact probabilities associated with the outcomes of their decisions, but have some inclination of the likelihood of outcomes. Examples might include determining the prospect of war, or guessing the reactions of a competitor.

Decisions made under uncertainty are the domain of this research. The role of UI is to increase knowledge about the likelihood of potential outcomes, thereby facilitating better decisions. Without UI decision-makers can only have some (presumably inaccurate) inclination about the likelihood of outcomes, which makes decision-making very difficult. For example, it is hard for a senior manager to make good capital expenditure decisions that hinge upon the achievement of some target-level (such as sales), without some indication of uncertainty.

However, when UI is provided the increase in knowledge about the likelihood of possible outcomes reduces the complexity of the task. If the UI is trusted as an accurate assessment of the likelihood of outcomes and is utilised when decision-makers are effectively operating towards the domain of risk. That is, those decision-makers that trust and utilise the UI effectively may perceive

themselves to be operating under risk. Accordingly they should behave like decision-makers operating under risk.

How UI is perceived will depend upon factors such as its source. For example, UI from experts may be perceived differently to statistically based UI. As a result, providing UI may not necessarily help those making decisions under uncertainty. Decision-makers may neither trust, understand nor utilise UI. When the UI is not utilised there is no decision-making benefit and decision-makers clearly operate under uncertainty.

Decision-making under uncertainty with UI lies under the domain of either uncertainty or risk, depending upon the decision-maker's perception of the UI. For those who trust, understand and utilise UI, the provision of UI moves a decision-maker along the domain of uncertainty towards the domain of risk, where knowledge of the likelihood of outcomes is well known (see Figure 2-1).

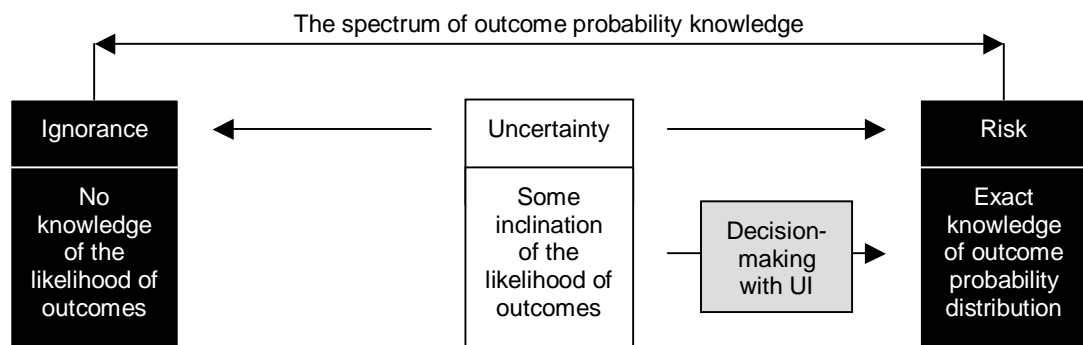


Figure 2-1: The domains of decision-making, and how UI moves decision-making under uncertainty towards the domain of risk.

Providing UI blurs the distinction between uncertainty and risk, and therefore any research investigating the utility of UI needs to consider prior research regarding both decision-making under uncertainty and decision-making under risk. After Knight (1921) and Keynes (1921) clearly distinguished

between risk and uncertainty, research on decision-making within these two domains diverged. Aspects of each domain of decision-making will now be briefly discussed.

Decision-making under risk

Traditionally research into decision-making has focussed on decisions under risk. Theories of how people make decisions under risk provide us with some indication of how UI may be utilised by a decision-maker who trusts and utilises the UI. Several theories have attempted to explain the choices made by decision-makers operating under risk, however none have been generally accepted as adequate predictors of behaviour (Schoemaker, 1990). Normative theories of decision-making under risk, including Expected Value¹ (EV) theory and Expected Utility² (EU) theory (Bernoulli, 1738; Bentham, 1789; Jevons, 1871; von Neuman & Morgenstern, 1947)³, are regularly violated by human risky choice behaviour. For example, EU theory does not account for either the

¹ Expected Value theory states that individuals base decisions on the objective combination of potential gain multiplied by probability.

² Individuals make decisions based on expected utility, which is defined as a combination of expected value and the unique situation of the decision-maker.

³ For a layman's review of the development of Expected Utility Theory see Bernstein (1996).

certainty effect⁴ (Allais, 1953; Kahneman & Tversky, 1979) or the decision framing effect⁵ (Tversky & Kahneman, 1981).

Prospect Theory⁶ (Kahneman & Tversky, 1979), a more recent, descriptive, theory of risky choice behaviour, has increased understanding of decision-making under risk by recognising people's apparent overweighting of low probabilities and the certainty effect (Allais, 1953; Kahneman & Tversky, 1979). However, Prospect Theory also has shortcomings. Prospect Theory cannot accommodate preference reversals (Slovic & Lichtenstein, 1983) between P bets (high probability of winning a modest sum) and \$ bets (low probability of winning a large sum), nor can it explain the inconsistencies between the probability equivalence⁷ and certainty equivalence⁸ methods in utility measurement (Hershey & Schoemaker, 1985). More recently, Cumulative

⁴ People behave in a risk-averse manner and prefer a certain \$1 more than they prefer to gamble with a 50% chance of \$2 and a 50% of nothing, even though the mathematical expectation of each situation is the same.

⁵ Modifying the way a problem is posed can influence the decision-maker's choice. This inconsistency in choice is known as the failure of invariance and is caused by modifying the decision-maker's reference point. For an example, see Tversky & Kahneman (1981).

⁶ Prospect Theory states that there is an asymmetry in the way people deal with gains compared to losses. People are risk-averse with gains and risk seeking with losses. The driving factor behind this is loss aversion. Another Prospect Theory concept states that decisions are made based on some reference point, and it is the deviations in utility away from the reference point, rather than Daniel Bernoulli's concept of resulting wealth, that influences decisions.

⁷ Probability equivalence occurs when a probability P of winning a simple gamble is equal in utility to a sure thing. For example, for risk-averse individuals a sure \$100 may be the same as a gamble where there is a 75% chance of winning \$200 and a 25% of winning \$0.

⁸ Certainty equivalence occurs when a sure amount S is found to be equivalent in utility to a simple gamble. For example, for risk-averse individuals \$70 may be equivalent to a 50-50 chance at \$200 or \$0 (Hershey & Schoemaker, 1985). The amount S is the certainty equivalence for the gamble.

Prospect Theory confirms “a distinctive fourfold pattern of risk attitudes: risk aversion for gains and risk seeking for losses of high probability; risk seeking for gains and risk aversion for losses of low probability” (Tversky & Kahneman, 1992).

While each of these theories falls short of broadly describing risky choice behaviour, until better theories are developed they are still an adequate basis for decision-making comprehension.

Decision-making under uncertainty

Theories of decision-making under uncertainty developed in parallel with those theories describing decision-making under risk. Following Knight's distinction of risk and uncertainty, Arrow (1953) and Debreu (1959) developed a state-preference approach to describing decision-making that did not require the assignment of probabilities (mathematical or subjective), thus lending these approaches towards describing decision-making under uncertainty.

Somewhere between uncertainty and risk a stream of decision-making theory emerged by incorporating subjective probabilities. Ramsey (1926) and de Finetti (1931, 1937) independently suggested a way of deriving a consistent theory of choice under uncertainty that could isolate beliefs from preferences, still maintaining subjective probabilities. These views were taken and turned into a full theory by Savage (1954) called subjective expected utility theory.

Subjective expected utility theory is not clearly under the domain of risk or uncertainty. The assignment of numerical probabilities, regardless of whether

they are subjective, implies decision-making under risk. However, these probabilities are subjective and therefore contain uncertainty.

The intersection of risk and uncertainty

More recently, theories describing decision-making under risk and uncertainty have begun to merge. Tversky & Kahneman (1992) extended prospect theory in several ways, including its application to uncertain as well as risky prospects. Known as cumulative prospect theory, the generalisation to risk and uncertainty was achieved by combining the cumulative choice model for uncertain assets by Smeidler (1989) with the cumulative choice models for probabilistic prospects from Quiggin (1982) and Yaari (1987).

Tversky & Kahneman (1992) demonstrated a four-fold pattern of risk attitudes: risk aversion for gains and risk seeking for losses of high probability; risk seeking for gains and risk aversion for losses of low probability. Tversky & Wakker (1995) and Tversky & Fox (1995) used the principle of bounded subadditivity⁹ and confirmed this relationship for uncertain prospects.

Subsequently, researchers have attempted to incorporate subjective probabilities with theories of risk in order to explain decision-making under uncertainty. In their study of subadditivity in both risk and uncertainty, Tversky & Fox (1995) investigated the relationship between subjective probability and decision weights. Their results were consistent with a two-stage model of choice in which the decision maker first assesses the probability of an uncertain event,

⁹ An event has greater impact when it turns impossibility into possibility, or possibility into certainty, than when it merely makes a possibility more or less likely.

and then transforms this value using a risky weighting function. Fox & Tversky (1998) and Wu & Gonzalez (1999) specified such models and found empirical support for their predictions. These models of choice use support theory¹⁰ (Tversky & Koehler, 1994) as the basis for how the decision-maker assesses probability, and prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) for how the subjective probabilities are converted using a risky weighting function.

Knight (1921) and Keynes (1921) rightly distinguished between risk and uncertainty, and others such as Cohen, Jaffray & Said (1987) have shown that people behaved differently under risk and uncertainty. However, the recent convergence of theory treats risk more like a special case of uncertainty (Tversky & Kahneman, 1992; Fox & Tversky, 1998; Wu & Gonzalez, 1999). As mentioned, Tversky & Wakker (1995) and Tversky & Fox (1995) showed that choice under uncertainty exhibits some of the main characteristics of choice under risk. Others also found support for the similarity between risk and uncertainty. For example, Di Mauro & Maffioletti (2001) found uncertainty about probabilities (that is, moving from risk to uncertainty) to have a weak impact on consumers' valuation of insurance.

Perhaps the most pervasive finding regarding risk and uncertainty is that although choice under uncertainty does exhibit many of the main characteristics of choice under risk, people are less sensitive to uncertainty than to risk, and therefore decision-making under uncertainty should be more prone to

¹⁰ A descriptive theory of how people make probability judgments. According to support theory different descriptions of the same event can give rise to different judgments (nonextensionality).

departures from expected utility theory (Tversky & Wakker, 1995; Tversky & Fox, 1995).

This dissertation investigates decision-making under uncertainty with UI. As mentioned, UI helps move decisions from the domain of uncertainty towards the domain of risk by making information about the probability distribution of outcomes more salient to the decision-maker. Because providing UI blurs the distinction between risk and uncertainty, and because risk and uncertainty are clearly closely related, in order to understand the impact of providing UI to decision-makers it is necessary to investigate prior research on decision-making under both risk and uncertainty, as well as the production of subjective probabilities. The remainder of this chapter considers prior research on decision-making under both risk and uncertainty. Next is a closer look at how people actually make choices and under what situations normative models of decision-making under risk or uncertainty are violated.

2.1.2 Choice Behaviour

Key Messages:

- Understanding choice behaviour under risk and uncertainty requires understanding measures of risk.

Much empirical research into decision-making under risk and uncertainty has concentrated on measures of risk. The measurement of risk, and subsequent conveying of that information to a decision-maker, is central to

understanding behaviour under risk or uncertainty¹¹. Decision-makers can only operate optimally if they fully comprehend the situation that surrounds them. This can only occur if risk information is measured, conveyed and comprehended by the decision-maker. In short, understanding choice behaviour to a significant extent requires understanding how risk is best measured. It shall be seen in section 2.1.2.1 that risk is a nebulous concept, as noted by Rapoport & Wallsten (1972), making it hard to measure and making its impacts upon the decision-maker difficult to understand.

Empirical evidence of risky choice behaviour will then be summarised in section 2.1.2.2. Behavioural observations frequently contrast with theory, however some consistent themes have emerged.

2.1.2.1 Measures of Risk

Key Messages:

- The traditional measure of risk is variance.
- There is growing agreement that risk is a multidimensional phenomenon.
- Measures are starting to incorporate individual measures of risk preference and risk perception.

The traditional measure of risk is variance (Markowitz, 1959). That is, the risk of a particular decision or choice is the variance of the returns relating to that choice. Some studies, such as Cooley (1977), have at least partially

¹¹ The focus is on measures of risk rather than measures of uncertainty, because measures of uncertainty are generally estimates of risk measures and therefore are based on risk measures.

confirmed that variance plays a significant role in communicating risk. However, studies across a broad range of situations have found the variance of returns to be of secondary concern to decision-makers at best. Studies of gambling behaviour (for example, Aschenbrenner, 1978; Edwards, 1953, 1954; Slovic & Lichtenstein, 1968a, 1968b), managerial decisions (for example, Alderfer & Bierman, 1970; Conrath, 1973; MacCrimmon & Wehrung, 1986; Mao, 1970; March & Shapira, 1987) and stock-market investing (Farrelly & Reichenstein, 1984; Gooding, 1975; McDonald & Stehle, 1975) have all questioned the importance of variance as an all-encompassing measure of risk, pointing towards the need for an alternative, and possibly multi-dimensional, approach to risk measurement. As MacCrimmon & Wehrung (1990, p. 425) noted:

“We have used multiple measures of willingness to take risks because our research and that of many others have shown that a single measure of risk taking is inadequate – risk propensity is too complex to be captured by a single measure.”

Cooley, Roenfeldt & Modani (1977) investigated the multi-dimensional nature of risk by analysing the level of redundancy between eleven commonly used measures of risk. The study found that six of the measures, range, semi-interquartile deviation, mean absolute deviation, standard deviation, semivariance, and lower confidence limit, formed a homogeneous group with a high level of substitutability. The five remaining measures, coefficient of variation, coefficient of quartile variation, skewness, kurtosis and beta, captured other dimensions of risk.

MacCrimmon & Wehrung (1990) also looked at the consolidation of risk measures. Thirteen measures were reduced to seven factors, made up of: personal utility (attitude towards personal investments), risk return, debt

gambling (attitudes towards personal debt and gambling) risky assets (percentage of gross personal assets held in the most risky asset categories), business utility (attitude towards business investments), in-basket (reactions to “in-basket” situations, choosing between risky and sure alternatives as described in MacCrimmon & Wehrung, 1984) and a self rating (willingness to take business risks).

More recent studies looking at measures of risk have focussed on the incorporation of individual risk preferences and risk perceptions, the integration of risk measures with risk models, and the unification of different risky decision activities such as risk judgment and modelling choice.

Sarin & Weber (1993) proposed a risk-value model for evaluating decisions under risk, where decision preference is determined by both the risk and value of a particular choice. This approach helps to unify research on measuring risk judgments (for example, Coombs & Lehner, 1981, 1984; Fishburn, 1982, 1984; Keller, Sarin & Weber, 1986; Luce, 1980; Weber, 1990) and on modelling choice (for example, von Neumann & Morgenstern, 1947).

Jia & Dyer (1996) also developed a measure of risk directly linked to a preference model in their standard risk-value model. Unlike the risk-return structure developed by Bell (1995), Jia & Dyer’s model is compatible with measures of expected utility.

Weber & Milliman (1997) reported empirical evidence supporting the risk-value models of Sarin & Weber (1993), Bell (1995) and an earlier version of Jia & Dyer’s model (Jia & Dyer, 1994). The integration of research on perceived

risk and models of decisions under risk, as prescribed by Brachinger & Weber (1997) in their review of risk perception measures, is well advanced.

While the risk-value models unified hitherto separate areas of decision research, the nature, and importance, of risk judgments has been further advanced by numerous researchers.

Brachinger & Weber (1997) and Sitkin & Weingart (1995) saw a greater role for risk perception as a determinant of risky decision-making behaviour. Sitkin & Weingart (1995) demonstrated empirically that perceptions of risk influenced decision-making under risk. Jia, Dyer & Butler (1999) extended earlier work on standard measures of risk to develop a standard measure of perceived risk. This measure has the potential to unify a significant amount of empirical evidence relating to decision-making under risk. In particular, synthesising risk judgments, risk perceptions and preference modelling within the one model. However, this work is yet to be supported empirically.

Risk-value models are not the only contender for accurately measuring risk. While research into risk measures in decision theory has gravitated towards risk-value models, work within finance theory literature has recently focussed on lower partial moments (LPMs)¹². Unser (2000) demonstrates how this risk-value model can be enhanced through the use of naturally asymmetric LPMs.

¹² Lower Partial Moments (LPMs) are measures of downside or shortfall risk where only negative deviations from a target outcome are considered. Positive deviations above the target are considered desirable and do not impact an alternative's risk (Bawa, 1975; Jean, 1975; Unser, 2000).

In summary, there is currently no dominant generally accepted measure of risk, and it appears to be a multidimensional phenomenon. However, research is helping to better understand risk's nebulous nature. We have seen:

- Variance does play a significant role in communicating risk¹³ and is perhaps the most widely used measure of all, but variance alone does not adequately represent risk,
- Models integrating risk and value appear to be more generalizable,
- Measures directly compatible with decision models are most practical,
- There is growing support for the inclusion of risk perception in any form of risk measure, and
- There is increasing evidence suggesting that attitude towards risk is a personal trait (primitive) and not just a function of a decision task or the environment under which the decision is made.

So despite significant effort in developing alternate measures of risk, the simple measure of variance (Markowitz, 1959) and its extensions are still highly relevant. Risk is difficult to measure. Perhaps Keynes (1921) and Knight (1921) were right arguing that uncertainty (immeasurable risk) rather than mathematical probability is the ruling paradigm of decision-making.

Measuring risk is one problem, communicating that risk information to the decision-maker is a second problem. We may not know how to precisely measure risk, but we know that people react to risk in both rational and

¹³ Beta values from the Capital Asset Pricing Model (Sharpe, 1964) are an example of a measure of risk based on variance. Beta values measure the variance of an individual stock's price movements in comparison to an overall stock market index.

unpredictable ways. Looking at these reactions to risk will help to better understand how, and under what circumstances, measures of risk can be best utilised by the decision-maker. The next section investigates evidence of risky behaviour.

2.1.2.2 Empirical Evidence of Risky Choice Behaviour

Key Messages:

- Four commonly observed behaviours: losses weigh heavier than gains; ruinous losses are avoided; situations where target levels are jeopardised are avoided; and, risks are more likely taken in threatening situations.
- Framing conditions are a major effect on risk preference in choice tasks.
- Observed behaviour has threatened the validity of risk and uncertainty theories.
- Cumulative Prospect Theory (Tversky & Kahneman, 1992) is perhaps currently the best predictor of risky and uncertain behaviour.

Measures and models of risk attempt to understand and help predict risky choice behaviour. Certain types of behaviour are unexpected, difficult to explain and violate established theories, measures and models of risk. This section discusses some of these observed risky behaviours, discusses how they have violated established theories of risk, and identifies the most current theories that describe decisions under risk (see Figure 2-2 for a conceptual overview of this section).

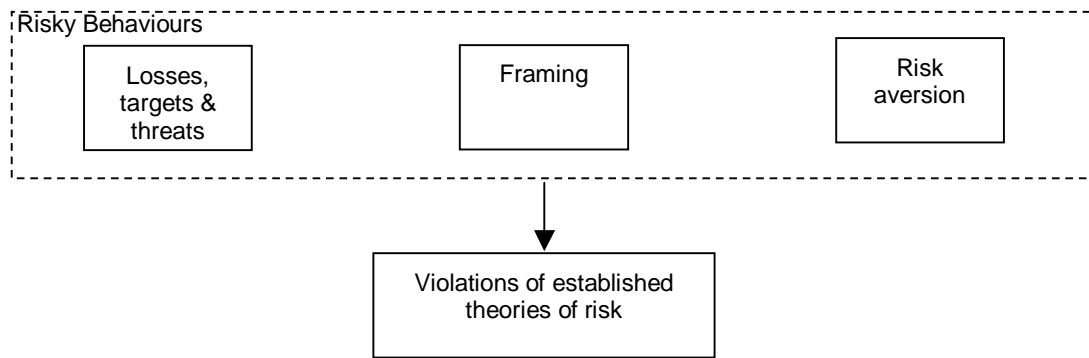


Figure 2-2: Risky behaviours that have violated established theories of risk.

The impact of losses, targets and threats

There are four particular behaviours that have been repeatedly observed and empirically confirmed under more than one decision-making context. First, it is clear that people weight the probability and magnitude of losses more so than they do that of gains. Second, empirical observations indicate that decision-makers under risk particularly aim to avoid ruinous losses. Third, decision-makers aim to avoid situations where a target level cannot be achieved. Fourth and finally, risks are more likely to be taken in threatening situations.¹⁴

Numerous studies looking at gambling behaviour have found that the probability and magnitude of outcomes can impact risky choice decisions (Aschenbrenner, 1978; Edwards, 1953, 1954; Payne, 1975; Payne & Braunstein, 1971; Slovic & Lichtenstein, 1968a). More specifically, some studies have found people to weight the probability and magnitude of loss more than

¹⁴ Clearly there are more than four patterns of behaviour that are commonly observed in decision-making under risk (for example, ambiguity aversion in Ellsberg, 1961, which will be discussed later), however these are the four most important behavioural patterns relating to this particular piece of research.

that of gain. Slovic (1967) found subject's risk ratings correlated most highly with probability of loss. Probability of loss was also found to be a key determinant of investment choice by Alderfer & Bierman (1970). They also found the amount of maximum possible loss to be another key factor. MacCrimmon & Wehrung (1984, 1986) found the probability and magnitude of loss, as well as exposure to loss, defined risk for a large group of American and Canadian business executives. American and Israeli executives were interviewed by March & Shapira (1987). Participants in March & Shapira's study tended to associate risk with negative outcomes, however of particular importance to them was the magnitude of the outcome rather than the probability.

Business executives interviewed by Mao (1970) associated risk with ruinous losses. In particular, they associated risk with investments that involved enough company resources to threaten insolvency upon failure. Laughhunn, Payne & Crum (1980), Crum, Laughhunn & Payne (1981), MacCrimmon & Wehrung (1984, 1986) and Kunreuther et al. (1995) also found people to be risk-averse to potentially ruinous losses.

Apart from predominantly associating risk with losses, people also appear to associate risk with target levels. Several studies have found business managers to be wary of taking risks where there is a high likelihood that their decisions will jeopardise reaching a target level. Mao (1970), Conrath (1973) and Greer (1974) observed that business executives defined risk as the prospect of not being able to meet a target level. It appeared the executives were under significant pressure to avoid any chance that an investment

decision did not achieve its target level. Similar behaviour was observed in Laughhunn, Payne & Crum (1980) and Crum, Laughhunn & Payne (1981).

March & Shapira (1987, p. 1413) summarised this phenomenon:

“In general, if one is above a performance target, the primary focus is on avoiding actions that might place one below it This leads to relative risk aversion on the part of successful managers, particularly those who are barely above the target

For decision makers who are, or expect to be, below the performance target, the desire to reach the target focuses attention in a way that leads generally to risk taking. “

Both MacCrimmon & Wehrung (1986) and March & Shapira (1987) observed that the executives participating in their studies were more likely to take risks in threatening situations. These observations are consistent with the value function presented in Prospect Theory (Kahneman & Tversky, 1979), where the steepest part of the hypothetical value function is in the convex loss segment. So there is clearly an interaction between the desire to take risks in threatening situations, and the need to protect the achievement of a target level.

Over the past decade considerable debate has continued to circulate regarding the completeness and validity of existing risky behaviour theories in light of empirical behavioural observations. Some of these discussion themes include:

- The importance of framing effects
- The questioning of behaviours previously thought of as inherent human risky behaviour, and

- Criticism of various theories of risky behaviour

Each of these themes has been investigated in empirical studies, and some of their results will now be discussed.

Framing effects

Tversky & Kahneman (1981) were among the first to observe framing effects in risk taking. They found that the framing of a problem could cause systematic preference reversals. In particular, results showed that choices framed as gains could cause risk-averse behaviour, while the same choices framed as losses gave rise to risk seeking behaviour. These observations sparked considerable debate, and caused many to rethink past empirical observations.

Framing effects themselves have also been questioned. Kühberger (1995) criticised the wording of the disease problem¹⁵ used by Tversky & Kahneman (1981) and showed that framing effects, especially by positively framed choices, were significantly reduced if no information was missing. Similarly, Mandel (2001), reports findings that indicate the results of Tversky & Kahneman (1981) were due to not just missing information, but to asymmetries in the ambiguity of the sure and risky prospects in their disease problem.

¹⁵ Two groups of participants were asked to choose a program to combat an unusual Asian disease expected to kill 600 people. The first group was asked to select between Program A, which would save 200 people, and Program B, which had a 1/3 chance of saving 600 people and a 2/3 chance of saving none. Not surprisingly, the risk-averse sure-bet of saving 200 people was preferred (72%). The second group was asked to select between Program C, which would see 400 people die, and Program D, which had a 1/3 chance of nobody dying and a 2/3 chance of 600 people dying. The reframing of this identical problem resulted in 78% of respondents preferring the risk seeking option (Tversky & Kahneman, 1981).

However, there is considerable evidence that framing effects do occur (Kühberger, 1998; Kühberger, Schulte-Mecklenbeck & Perner, 1999). In a comprehensive meta-analysis of research investigating framing effects Kühberger (1998) found that framing effects do occur, but the effect is “small to medium”. The analysis showed that framing effects, which are operationalised through a form of risk manipulation, were most commonly associated with manipulations of the reference point and were less apparent with problems differing from the Tversky & Kahneman (1981) style disease problem.

Kühberger, Schulte-Mecklenbeck & Perner (1999) expanded the work of Kühberger (1998) by looking at all studies similar to the Tversky & Kahneman (1981) disease problem to determine the major effects on risk preference in choice tasks. Their meta-analysis found that framing conditions were the most dominant effect, exceeding the impact of payoff magnitude, probability levels and the good at stake.

Risk aversion

The impact of framing effects is not the only reason that the previous results of studies investigating risky behaviour have been questioned. Recently the previously assumed risk-averse behaviour of decision-makers has been questioned, from varying directions. March (1996) hypothesized that greater risk aversion towards gains rather than losses may be due to accumulated learning, and not a human trait. Levy & Levy (2001) conducted an experiment to

investigate risk aversion using a stochastic dominance¹⁶ approach. By isolating effects caused by subjective probability distortion, the certainty effect (Allais, 1953; Kahneman & Tversky, 1979), and framing effects, Levy & Levy (2001) found that most individuals are not risk-averse.

However, findings questioning risk-averse behaviour currently appear to be in the minority. The risk preference meta-analysis performed by Kühberger, Schulte-Mecklenbeck & Perner (1999) showed that risk aversion could be observed not only due to framing, but also because of payoff quantity and level of probability. This meta-analysis showed support for risk aversion occurring with high payoffs, and also high probabilities leading to risk aversion for gains.

Modern theories of risky behaviour – Prospect Theory

Empirical investigation of risky behaviour puts formal theories of decision-making under risk into the spotlight. Expected Utility theory has for a long time been challenged by inconsistent empirical evidence (Camerer, 1995; Schoemaker, 1982; Starmer, 2001). These observations led to the development of alternative theories. Hastie (2001, p. 675) noted:

“There have been a few dramatic intellectual events in the recent history of the field of judgment and decision making. One is the sudden acceptance of non-expected utility theories with labile reference (inflection) points,

¹⁶ Consider two cumulative probability distributions F and G , representing two alternative decision options. First degree stochastic dominance refers to the dominance of one choice's cumulative probability distribution over another choice's cumulative probability distribution. That is, F dominates G if $F(x) \leq G(x)$ for all x , and F is preferred to G by all rational decision-makers who think more is better. See Goodwin & Wright (1999) for a thorough explanation.

If we assume risk-aversion, second degree stochastic dominance by a cumulative probability distribution F occurs when the cumulative area under F is less than the cumulative area under G . This allows stochastic dominance to be determined even when F and G intersect. The use of a stochastic dominance methodology allows the researcher to ask decision-makers to decide on more realistic choices that have both positive and negative outcomes, not just one or the other.

separate value functions for gains and losses, and nonadditive probability weighting functions.”

Cumulative Prospect Theory by Tversky & Kahneman (1992) is probably the best known of these non-expected utility theories, which advanced their earlier work on Prospect Theory (Kahneman & Tversky, 1979). However, in recent times, these theories have also been brought into question.

As previously mentioned, the standard framing effect demonstrates that choices framed as gains can cause risk-averse behaviour, while the same choices framed as losses can lead to risk seeking behaviour. Kühberger (1995) demonstrated a reversal of the standard framing effect when choice descriptions were incomplete. Prospect Theory was unable to predict this behaviour.

Prospect Theory has also been criticised for its prediction of risk taking for losses and risk aversion for gains. Several researchers (for example, MacCrimmon & Wehrung, 1986) have indicated that Prospect Theory conflicts with accepted observations that potential opportunity induces risk taking while threats cause risk aversion. In response to this theoretical conflict, Highhouse & Yüce (1996) attempted to separate threat and opportunity perceptions from loss and gain perspectives in an empirical experiment. They showed that threat and opportunity perceptions are distinguishable from loss and gain perspectives. Their results showed support for Prospect Theory, with participants of their experiment being risk seeking for losses and risk-averse for gains. However, Highhouse & Yüce (1996) also found that opportunity-framed risks led to risk taking while threat-framed risks caused risk aversion. This contrasts with Prospect Theory’s prediction of risk taking for losses and risk aversion for gains.

The meta-analysis by Kühberger, Schulte-Mecklenbeck & Perner (1999) supported Prospect Theory's general prediction of risk aversion for gains and risk seeking for losses. The same study found no support for Cumulative Prospect Theory's prediction of preference reversal for low probabilities. However, low probabilities did lead to a weakening of support for the behaviour predicted by Prospect Theory.

It can be seen that there is a wealth of empirical observations indicating that human responses to decision-making do not follow the behavioural patterns predicted by normative models of judgment. This gap between theory and reality has sparked significant recent debate. Alternative explanations to normative predictions include: humans are irrational; the norm has been inappropriately applied (Gigerenzer, 1991); the problem has been misconstrued by the decision-maker and as a result they are responding to another problem (the differential construal argument, for example see Stanovich, 1999); or the participant has failed to process or understand the task (the understanding/acceptance principle, Slovic & Tversky, 1974; Stanovich & West, 1999). Stanovich & West (2000) found that this normative/descriptive gap was predominantly due to computational limits in line with the differential construal argument and the understanding/acceptance principle. In contrast Shafir & LeBoeuf (2002), in a comprehensive review of normative violations, conclude that such "violations cannot be dismissed as either random or trivial, nor can they be attributed to experimenters' misinterpretation of answers that are actually appropriate to alternative, valid interpretations of the problems. The systematic and well-documented findings cannot be attributed to simple computational limitations, nor does it appear that inappropriate types of

questions are being asked or inappropriate norms applied (p. 509)". Oaksford & Chater (2001) believe that apparent irrational behaviour is in fact due to people using everyday reasoning strategies inappropriately in the laboratory.¹⁷ They are therefore being incorrectly evaluated for using a probabilistic approach¹⁸ to laboratory problems, when actually they are making a "rational attempt to make sense of the tasks they are set, by applying strategies adapted for coping with the uncertainty of the everyday world (p. 356)". The debate continues.

The empirical evidence presented in this section illustrates that our knowledge of how people make decisions under risk has increased dramatically through observation, however we are still far from developing a comprehensive theory of risky behaviour. While there is no one dominant theory of risky behaviour, there is growing agreement on what aspects various theories require to have high validity. For example, Mellers, Schwartz & Cooke (1998, p. 467) commented:

"... there is widespread agreement that risky and uncertain choices are rank dependent. But specific issues about the utility function and the weighting function – including shape, form (cumulative vs noncumulative), and factors that influence it – are still points of controversy."

¹⁷ This is based on the observation that people process information in two very different ways (Evans & Over, 1996; Oaksford & Chater, 2001). 'System-1' processes are automatic, unconscious and based on knowledge acquired through everyday life. 'System-2' processes are controlled, analytical and based on explicitly acquired formal rules. Problems occur when System-1 processes are applied when System-2 processes should be. This appears to occur frequently.

¹⁸ A probabilistic approach to logical tasks sees people apply everyday reasoning strategies (which are probabilistic) in a laboratory setting, often inappropriately. This approach was first observed by Wason (1968). For a more recent discussion see Oaksford & Chater (2001).

In a comprehensive review of Non-Expected Utility Theory, Starmer (2000) attempted to look for a replacement for Expected Utility Theory. Following on from Mellers et al. (1998), Starmer (2000, p. 376) stated:

“From a practical point of view the rank-dependent models (including the sign-dependent variants) provide a convenient way of modelling these well established influences on choice, and there seems good reason to push forward the task of examining what implications such models have in general economic contexts.”

So while there is still no comprehensive theory of risky choice behaviour, it appears that some of the best theories at present include models like Cumulative Prospect Theory from Tversky & Kahneman (1992):

“Perhaps there is a case for thinking that the position we should now aim for is one in which models like cumulative prospect theory become the default in applied economics with EUT used as a convenient special case, but only when we can be confident that loss aversion and probability weighting are insignificant (Starmer, 2000, p. 370).”

So far we have reviewed measures of risk (inputs to a decision-making process), theories of decision-making under risk and uncertainty and empirical observations of behaviour under risk (outcomes from a decision-making process). We now turn our attention to the decision-making process itself, and the interaction of decision-maker, task and environment. Analysis of the decision-making process will indicate the normative value of providing UI in uncertain situations.

2.2 FACTORS INFLUENCING UI UTILITY – A MODEL OF CONTINGENT EFFECTS

Key Messages:

- This section defines a model of contingent effects based on a person, task, and environment framework.

What does the literature tell us about how decision-makers should benefit from UI? Decision-making under uncertainty is highly contingent upon relatively small changes in task and environment (Einhorn & Hogarth, 1981; Payne, 1982, 1985). Accordingly, apparently minor changes in task and environment (and possibly the person) should influence the utility of UI. In this section a model of contingent effects describing factors that should influence the utility of UI is developed. This model will then be used to evaluate the inconclusive empirical results of prior research investigating the utility of UI (section 2.3).

Beach & Mitchell (1978) proposed a contingent model for the selection of decision strategies based on personal characteristics of the decision-maker, the type of decision task, and the surrounding environment. This is a comprehensive model for understanding decision-making strategies, and has been thoroughly tested (Christensen-Szalanski, 1978; McAllister, Mitchell & Beach, 1979; Waller & Mitchell, 1984). The Beach & Mitchell (1978) contingent model of behaviour has been found to explain the calibration of confidence in human judgment (O'Connor, 1989a), so it would seem reasonable that the same model could be extended to understanding the utility of UI in decision-making under uncertainty: from UI production to UI utilisation.

This section will investigate theoretical and empirical research in order to understand what factors (task, person or environment) influence the utility of UI in decision-making under uncertainty. The first three sections investigate the influence of personal (2.2.1) and task (2.2.2) characteristics along with the impact of environmental factors (2.2.3). The fourth section (2.2.4) examines interactions between personal and task factors that may influence decision performance. Finally, a model of contingent effects describing UI utility in decision-making under uncertainty is presented (2.2.5).

2.2.1 Person

Key Messages:

- Experience/Ability is the only personal characteristic to be discussed in this section.

Beach & Mitchell (1978) specified three personal characteristics in their decision strategy model: knowledge, ability and motivation. The potential influence of knowledge and ability will be investigated as part of section 2.2.1.1 (Experience/Ability). For the purposes of this study, motivation is seen as a function of incentive and perceived meaningfulness of task, and as such its investigation is included in assessing the impact of environmental factors (2.2.3.3 Incentives and Motivation). Interactions of personal characteristics with task variables (for example, individual attitudes towards risk and uncertainty) are discussed separately in section 2.2.4 (Interactions of Person and Task) since they influence fundamentally new behaviours, such as perceptions of risk and assumptions about underlying uncertainty.

2.2.1.1 Experience/Ability

Key Messages:

- Despite differences in task knowledge, students (particularly post-graduate) are adequate surrogates for managers in decision-making experiments.
- Even when there are differences in behaviour due to experience, differences are with the magnitude of effect rather than the direction of effect.
- Competence in a decision-aid (such as UI) is greatly improved through training.

Beach & Mitchell (1978) observed that the knowledge and ability of the decision-maker had the potential to influence the selection of a decision strategy. Knowledge can be further categorised into task specific knowledge and decision-aid competency. In practice, the level of task specific knowledge is often operationalised in experimental procedure in the form of experience through the use of experts as opposed to novices, or students. Decision-aid competency is operationalised through decision-aid specific training.

Much empirical research into decision-making is conducted using students as surrogates for managers. This is understandable given the generous supply of students most researchers have access to. However, concern over the external validity of results has led several researchers to investigate whether students are indeed adequate surrogates for managers in decision-making experiments. Some found that students were often not appropriate surrogates

for managers (for a summary see Gordon, Slade & Schmitt, 1986). Abdolmohammadi & Wright (1987) found experience to be of significant importance in complex audit decisions, and therefore questioned the appropriateness of using students as surrogates for CPAs in audit decision settings. In another audit related experiment completed by practising auditors, Simnett (1996) showed that expertise improved selecting appropriate information for making an optimal decision, although information load (or overload) had equivalent impacts regardless of expertise.

There is a significant body of work focussed on identifying and understanding expertise (for example, Shanteau, Weiss, Thomas & Pounds, 2002). Shanteau et al. would argue that experts are fundamentally different to novices, and have developed techniques for identifying and measuring expertise.

In contrast, Remus (1996) found no significant difference between the decision-making of graduate business students and managers. However, in the same study undergraduate business students were found to behave differently, and in particular their decisions were more erratic. Similar results have also been found by Whitecotton (1996b) and by Whyte & Sebenius (1997). Others have found that in some task settings, such as the probabilistic forecasts of stock prices, there was an “inverse-expertise” effect (for example, Staël von Holstein, 1972; Yates, McDaniel & Brown, 1991).

It should be noted that task specific experience has been shown to interact with task characteristics, such as task format, and modify decision behaviour (Stone & Dilla, 1994; Önköl & Muradođlu, 1996; Trotman & Wright, 1996;

Wilkie-Thomson et al., 1997). Experience was also found to modify the decision-maker's attitude towards risk (Snowball, 1980; March & Shapira, 1992). However, more recently Kamalanabhan & Sunder (1999) found no significant difference between the risk-taking propensity of managers, non-managerial employees and MBA students (potential managers).

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 an effect rather than the direction of the effect (Zelditch & Evan, 1962; Ashton & Kramer, 1980; Remus, 1986). Accordingly, students, and particularly graduate students, should be seen as a valuable source of experimental subjects, having adequate knowledge to evaluate and select appropriate unaided-analytic decision strategies. This is particularly pertinent when considering that even realistic decision-making experiments have been criticised for being just as ecologically invalid as laboratory settings (Bolger & Wright, 1994), without the same level of internal validity offered by laboratory experiments.

The other part of knowledge specified by Beach & Mitchell (1978) relates to decision-aid competency. While task related knowledge comes through experience, decision-aid competence generally comes about through training, such as being trained to utilise UI. Whitecotton (1996a) found decision-aid competence (or experience) did increase reliance on a decision aid. Presumably training in a decision aid or on available information cues (such as UI) should then increase the likelihood of the use and therefore utility of such aids and information cues.

Beach & Mitchell's (1978) concept of ability comes down to the intelligence of the decision-maker combined with task experience and possibly decision-aid competence. Since intelligence is generally not controlled for in an experimental setting, ability comes down to task knowledge and decision-aid competence, both which have already been discussed.

In summary, there is strong evidence suggesting experience does not significantly modify decision-making. If there is any difference caused by experience, it is in the magnitude of the effect, rather than the direction. As a result, students may be used as surrogates for managers in empirical investigations of decision-making behaviour. However, it is important to note four criteria relating to personal characteristics that should aid investigation into the utility of UI:

- If students are to be used as experimental subjects, then the use of graduate business students will maximise the chance results will generalise to a managerial audience.
- The use of a consistent group (for example, all graduate business students rather than a mix of graduates, undergraduates or managers) of experimental subjects will remove the chance of interaction effects taking place between experience and other task characteristics. Using a consistent group will hopefully also help to standardise attitudes towards risk.
- Selection of a task that is unlikely to have been completed previously by participants will reduce the likelihood of task knowledge, and provide a "level playing field" where differences in decision performance will be due

to nascent experience based on available decision-aids and information cues.

- Ensure training in decision-aids and all information cues takes place to increase the probability that these decision-aids and information cues are understood and utilised.

While the experience of a decision-maker may impact their ability to utilise UI, task characteristics are likely to have a far greater impact upon the utility of UI and decision performance. These impacts are discussed next.

2.2.2 Task

Key Messages:

- Judgment cannot be understood without understanding the task.
- Tasks can be defined in terms of 'information disclosure' and 'task uncertainty'.

The relationship between judge and task is of central importance to the study of judgment. Brunswik's Probabilistic Functionalism (1952, 1956)¹⁹ argues that judgment cannot be comprehended without understanding the properties of the task. It may be that the appropriateness of judgment compared to models for a particular task (the accuracy of expert judgment) also cannot be determined without understanding the properties of a task (Stewart, Roebber & Bosart, 1997).

¹⁹ Probabilistic Functionalism (Brunswik, 1952, 1956) has been extended through Social Judgment Theory (Hammond, Stewart, Brehmer & Steinmann, 1975; Doherty & Kurz, 1996) and Cognitive Continuum Theory (Hammond, 1980, 1981).

Beach & Mitchell (1978) provided a comprehensive break-down of what type of characteristics make up a task or decision problem: 'unfamiliarity' with the task by the decision-maker; the 'ambiguity' of the task, such as unclear goals; 'complexity' of the task measured by its number of components; and, the 'instability' of the task's goals and constraints.

The categories provided by Beach & Mitchell (1978) are too vague and arbitrary to precisely evaluate what task characteristics might influence the utility of UI. More precision is required. In this section alternate categories will be used to investigate what task characteristics may influence the utility of UI. These categories are 'information disclosure' (section 2.2.2.1) and 'task uncertainty' (section 2.2.2.2).

Information disclosure refers to the number and type of information cues, the format of the instrument used to present the information cues and whether there is any interaction between the information cues. This is a subset of the complexity category used in Beach & Mitchell (1978). Task uncertainty refers to the level of uncertainty associated with the task's information cues. This might be operationalised through manipulating the variance of the data used to produce the information cues. This is a subset of the ambiguity category used in Beach & Mitchell (1978). Each of these categories will now be thoroughly investigated.

2.2.2.1 Information Disclosure

Key Messages:

- Information disclosure refers to the number and type of information cues, the format of the instrument used to present the information cues and whether there is any interaction between the information cues.
- For UI, 'information disclosure' is determined by the way in which the UI is provided ('provision of UI') as well as 'cue mix & information load'.

The type, quantity, quality and mix of information cues given to decision-makers should influence the utility of UI. The information disclosed to the decision-maker changes according to the 'provision of UI' (section 2.2.2.1.1) and 'cue mix & information load' (section 2.2.2.1.2)²⁰. It is these characteristics of information disclosure that are of interest in determining the best way to improve UI utility and ultimately for designing a task adequate for testing the utility of UI. First, and most importantly, the way in which UI is provided will be investigated.

2.2.2.1.1 Provision of Uncertainty Information

Key Messages:

- The format used to convey UI can significantly affect the information's utility.

²⁰ Task complexity covers more than just cue mix and information load, however that is the limit to which this topic shall be discussed in this literature review. See Wood (1986) for a comprehensive, structured definition of task complexity, and Paquette & Kida (1988) for an example of how task complexity can influence decision performance.

- People prefer, anchor on and are more confident with point estimates than they are with ranges, probability distributions, and confidence intervals.
- Informativeness is favoured over accuracy, explaining the preference for tighter ranges and confidence intervals.

There are strong theoretical arguments for assisting people's assessment of uncertainty (see Chapter 1: Introduction). People have limited information processing capacity (Hogarth, 1975) and therefore need help in assessing uncertainty. However, equally driving the need for UI is the need to properly communicate forecasts and predictions (Fischhoff, 1994). Decisions based on forecasts should be made more effectively when the recipient of the forecast understands how accurate the prediction is, and what the chance of variation is.

There are a variety of ways in which UI can be provided to a decision-maker. This discussion will be limited to the provision of UI relating to a continuous variable. UI can differ in terms of presentation format (for example, numerical, verbal, graphical, tabular), source (expert, historical data) and level of detail or complexity (level, interval, probability distribution). Each way has benefits (such as reducing the computational workload of the decision-maker or increasing the use of rational, analytical decision strategies) and disadvantages, the utility ultimately determined by how the UI is interpreted by the recipient. Interpretation is influenced by their ability to abstract useful information from the UI, and their willingness to trust the given information (see Figure 2-3).

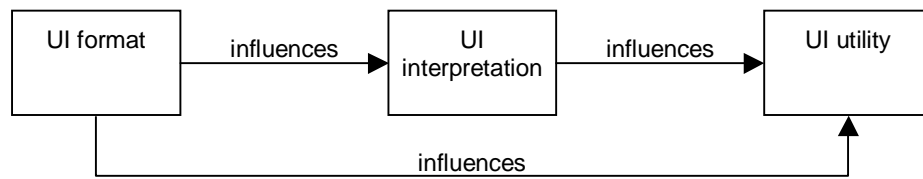


Figure 2-3: The relationship between UI format and UI utilisation, and the mediating effects of interpretation.

The different ways in which UI can be formatted

So how can the information that a forecast conveys be supplemented? Normally a forecast is provided as a point-estimate, a format that communicates precision and confidence. Unfortunately, that confidence is often only perceived rather than real. A point-estimate without additional contextual information can lead to unexpected and unintended consequences through decisions by the end-users of this information. For example, in an analysis of methods used to disclose contingent environmental liabilities, Kennedy, Mitchell & Sefcik (1998) found that financial statement users clearly anchored on the point-estimates provided. In this study management reported either a minimum, maximum, best estimate or range for contingent environmental liabilities²¹. The users of the financial statements exhibited anchoring behaviour when either minimum or maximum values were provided.

The simplest way to augment a point-estimate forecast is to provide an interval-estimate, an upper and lower bound relating to the forecast. The size of

²¹ Contingent environmental liabilities are recorded for environmental investigatory and non-capital remediation costs at identified sites when litigation has commenced or a claim or assessment has been asserted or is imminent and the likelihood of an unfavourable outcome is probable.

the interval then gives the recipient of the information some indication of the confidence they can place in the forecast. As a result interval forecasts can be more useful since they explicitly address uncertainty in forecasts (Pindyck & Rubinfeld, 1991), identifying alternatives and igniting discussion and debate. However, the interval does not contain any explicit indication of mathematical uncertainty.

Providing decision-makers with a probability distribution describing the likely value of a variable being forecast should greatly improve the individual's ability to assess uncertainty. Nakano (1976) argued for an accounting system based on probabilistic accounting data, due to the potential advantages to decision-making. However, it appears that the information provided by a probability distribution is often too complex for decision-makers to easily incorporate into their decision process. Conrath (1973) found in a clinical experiment that decision-makers were not comfortable with probability distributions, and preferred point-estimates. This study also showed people were better with less detailed probability distributions. Participants preferred cumulative probability distributions made up of 8 discrete steps rather than more detailed distributions made up of 20 steps. In line with this finding, Preckel & DeVuyst (1992) proposed an alternate method of presenting numeric/tabular probability distributions, condensing them using Gaussian quadrature approximate distributions.

Conrath speculated that the reason people might not prefer probability distributions may be because they know probability distributions are often based on dubious data. Given the choice between a complex decision-aid based on

unreliable data, and a simple point-estimate the simple information cue is preferred. However, with training this type of information has been shown to dramatically improve forecasting performance. For example, Abt, Borja, Menke & Pezier (1979) found the use of probabilistic forecasts for indicating the future sales of new pharmaceutical products to be very effective in improving decision-making previously based on point-estimate forecasts. Even if effective, it is often difficult and time-consuming to produce forecasts accompanied by probability distribution information.

A confidence interval is an explicit representation of UI. Compared to probability distributions, confidence intervals are “a less ambitious expression of uncertainty ... bounding the set of possible values with some degree of certainty (Fischhoff, 1994, p. 392)”. Their use by decision-makers has been advocated by a number of researchers (Mohn & Reid, 1977; Remus & Simkin, 1987; Makridakis, 1988; Kydd, 1989). The benefit of the confidence interval is that it provides the power of a probability distribution in a simple, easy to produce, and easy to process, format. Most forecasts assume a Gaussian confidence interval²² for forecast errors, although recent work by Gordon & Kammen (1996) investigating the confidence intervals of forecasts for the Standard & Poor’s 500 Stock Index shows that often this may be an inappropriate assumption.

Herman, Ornstein & Bahrack (1964) compared the utility of probability distributions with confidence intervals in a two dimensional object location experiment. Participants played an aircraft search-attack mission, in which the probabilistic information referred to the location and likelihood of hitting an

²² A Gaussian confidence interval is based on a Gaussian, or normal, distribution.

enemy aircraft. Location information altered as the participant's aircraft drew nearer. Participants received either a point-estimate of the enemy's location, a complete probability density function centred on the point-estimate or two contours centred on the point-estimate representing the 50% and 90% points of the distribution function of the point-estimate. Those receiving the two contours (the graphical confidence intervals) significantly outperformed the other groups. Those who received the complete probability density function had more information about the possible location of the enemy aircraft than those receiving just the two contours, however they were unable to fully utilise that information.

Confidence intervals can be thought of as discrete probability distributions, with only two values. Discrete probability distributions can have more than just two points, with their purpose being to communicate the same information as a continuous probability distribution, but with greater simplicity. Zaino & D'Errico (1989) discussed several of these approximations, and demonstrated their robustness and ease of use.

The final way to communicate UI is verbally, through terms such as 'doubtful', 'probable' or 'likely'. Wallsten et al. (1986) showed that these terms could be converted to numerical equivalents within the [0,1] probability interval (see also Wallsten et al., 1993). Some studies have found that the provision of numerical probability information can cause judgments to be less than optimal (for example, Erev, Bornstein & Wallsten, 1990) and people have been shown to be poor at processing numbers (for example, Hammerton, 1973), so there may be some benefit in communicating UI verbally. Results in Dilla & Stone

(1997) supported this prediction, finding that verbal risk cues increased judgment consistency. Contrastingly, they found that numeric risk responses were more consistent. The same study found that while verbal cues and numeric response formats encouraged better judgment, they also led to greater cognitive effort, with judgments taking longer. Other studies have supported these results from the risk response side. Verbal judgments of uncertainty in Rapoport et al. (1990) were less conservative (better calibrated), but were more variable and as a result less accurate. Judgments were more consistent using numerical representations of risk in Stone & Dilla (1994), but only by experienced participants (auditors).

How UI may be interpreted

Having reviewed the different ways in which UI can be conveyed, how do humans interpret this information? When not receiving any explicit information regarding the uncertainty of a decision environment, people may abstract probability information from the data sample. For example, people may abstract cues from time series information (Eggleton, 1982; Pitz, 1980; Pitz et al., 1976).

People are good at estimating central tendency (Beach & Swensson, 1966; Pitz et al., 1976), but have great difficulty with interpreting uncertainty. Difficulties with uncertainty stem from limited information processing capacity (Hogarth, 1975; Pitz et al., 1976), an inability to compare the relative variability of small and large numbers (Beach & Scopp, 1968), and by being too heavily influenced by the sequence of numbers or objects (Lathrop, 1967).

Despite theoretical support for the value of probability information in decision-making, there is strong evidence that people do not consider

probability in judgment. March & Shapira (1987) found managers to be insensitive to estimates of probability for possible outcomes. Of far greater importance to the managers was the achievement of set performance targets. Conrath (1973) found a similar combination of behaviours, with a fixation on target levels and a dislike for probability information. People do not think of probability in terms of a frequentist²³ model either. Teigen, Brun & Frydenlund (1999) found participants in several experiments appeared to be following a dispositionist²⁴ model of probability and risk.

How does this inability to consider probability impact the way people behave when provided with UI? In an investigation of investor judgment Hirst, Koonce & Miller (1999) showed that investors see ranges as less precise than point estimates. As a result investor's confidence is reduced when provided with range forecasts. However, ranges do not elicit different responses because investors use the midpoint of the range as their estimate.

Yaniv & Foster (1997) looked at the communication of judgmental probability information, and the existence of an accuracy-informativeness trade-off (Yaniv & Foster, 1995). They showed that interval estimates vary between accurate (large, with a high hit-rate of predicting a correct estimate) and informative (small, but with a lower proportion of correct estimates). They found

²³ A frequentist is someone who believes that the probability of an event should be defined as the limit of its relative frequency in a large number of trials.

²⁴ What the authors mean by dispositionist, is a person who tends to base their understanding of probability and risk on "if-then" scenarios. That is, something will happen if a certain condition is satisfied or realised. In other words, frequencies and statistics were not important in understanding probability and risk.

that the recipients of this information did look for a balance between accuracy and informativeness, however they did exhibit a greater need for informativeness. While this does not point towards the preference for point estimates (Chen & Summers, 1977, used information theory to show point estimates are less informative than probabilistic information) it does show a dislike for large interval estimates. This behaviour parallels a phenomenon known as ambiguity aversion (for example, see Einhorn & Hogarth, 1985), which will be discussed in greater detail in section 2.2.4.1.

How UI may be utilised

The way people utilise statistical forecasts when making judgmental forecasts should give us some clues as to how people will utilise UI. Goodwin & Fildes (1999) ran a judgmental forecasting task to see how statistical forecasts would be used as an input to the forecasting process. Results showed that the forecast did sometimes improve decisions, especially when accompanied with an explanation of the statistical forecast, however it was not used optimally and was modified and ignored at inappropriate times.

Within this abyss of apparently irrational behaviour it's difficult to know how people can be made to utilise UI. However, there is some light at the end of the tunnel. Snizek & Buckley (1995) showed that decision-makers do react to indications of confidence, specifically in judge-advisor situations. Moreover, judges employed a 'confidence utilization' strategy, selecting advice from the most confident advisor when there was conflict in advice.

There appear to be significant hurdles to overcome before humans will be willing to utilise probability information. Davis & Kottemann (1995) reviewed the

literature and found that there is a reluctance to use all forms of decision aids and decision rules, not just those relating to probability information. So what is the best way for decision-makers to receive UI? Yates et al. (1996) evaluated probabilistic forecasts from the users' perspective, and discovered three distinctive subjective principles of evaluation (p. 55):

“First, there is the emphasis on categorical ‘correctness’. Although a good case can be made for the value of probabilistic rather than deterministic forecasts, it is not apparent that the typical consumer appreciates the advantages. Second ... some consumers interpret (forecast) extremeness as an indicator of forecaster competence, others the opposite ... And then there is the seemingly very strong desire by consumers that forecasters be able to give good explanations for their predictions.”

It can be seen that the way in which users of probabilistic forecasts evaluate the information they receive may not be in their best interests, such as in the desire for categorical correctness rather than more informative probabilistic information cues.

In line with the third finding in Yates et al. (1996), that forecasts should have good underlying explanations, Dawes (1999) also found humans require significant contextual understanding in order to effectively utilise probabilistic information. Specifically, Dawes (1999) found humans to be “‘the primates whose cognitive capacity shuts down in the absence of a story’ when attempting to incorporate probabilistic information to make a coherent probabilistic inference (p. 29)”.

Fischhoff (1994) also argues that the background to a forecast needs to be known. One way to do that is to use a concept known as ‘pedigree’, from Funtowicz & Ravetz (1990), which describes numerical estimates according to

the quality of the model used to generate the forecast, the quality of the data and the degree of social acceptance for the procedure.

Similar to Yates et al. (1996), Fischhoff (1994) specified a framework for improving the communication of forecasts. To successfully communicate a forecast, the forecast must avoid the following: ambiguity, irrelevance, immodesty and impoverishment²⁵. UI should help to avoid ambiguity and immodesty by clarifying the uncertainty of the situation and by clearly articulating the limits of the forecast.

Lessons from a study of decision aid reliance by Kaplan, Reneau & Whitecotton (2001) provide two more clues as to how probability information should be communicated. The study also found decision aid reliance to be higher if validity was not disclosed, which appears to correlate other findings that accuracy and correctness are generally expected from information cues such as forecasts. In a similar study Lawrence, Goodwin & Fildes (2002) observed similar behaviour. Persons involved in the development of a forecasting decision support system were significantly more satisfied with resulting forecasts when compared to those individuals who had no involvement in the development of the system. However, despite increased forecast satisfaction, decision accuracy for those involved in the development of the forecasting system was slightly inferior. These studies showed that decision aid reliance was improved if the decision-maker was involved in the development of the decision aid. Unfortunately this is often impractical or inappropriate.

²⁵ If a forecast is impoverished, it fails to address the broader context within which the forecast has been made (Fischhoff, 1994).

In summary, there is clear evidence that humans should greatly benefit from the availability of UI when decision-making under uncertainty. However, the utility of the UI appears to be dependent on the form in which it's provided. People clearly prefer the preciseness and simplicity of point estimates, but such information cues are deficient in terms of UI. Probability distributions and confidence intervals have far more UI, although the former generally seem to be too complex for decision-makers to competently utilise. The challenge is to determine how decision-makers will relinquish their strong relationship with point estimates for UI in the form of something like confidence intervals.

Given that people appear to prefer point estimates (Kennedy, Mitchell & Sefcik, 1998; Hirst, Koonce & Miller, 1999), focus their attention on target levels and are known to ignore confidence intervals by utilising the mid-point, perhaps a promising format for UI, when appropriate for the task, is to use a confidence level²⁶ (a one sided confidence interval) rather than a confidence interval. A forecast generally indicates a point at which there is a 50% chance of a variable ending up either above or below the forecast (assuming the distribution is symmetric). A confidence level would indicate a point at which there is an X% chance of a variable ending up above the forecast and a (100 – X)% chance of a variable ending up below the forecast, where X is between 0 and 50. As long as this confidence level is provided along with a good story (for context), and with appropriate indications of quality (to improve trust), the utility of UI should be improved.

²⁶ Confidence levels are also known as forecast limits (Wei, 1990) or prediction bounds (Brockwell & Davis, 1991).

2.2.2.1.2 Cue Mix & Information Load

Key Messages:

- People may seek to extract UI from history information rather than utilising explicit UI when both cues are provided.
- Preference for point estimates could lead to forecasts dominating UI when both cues are provided.
- Cue mix and information load can detrimentally impact UI utilisation because limited information processing capacity leads humans to seek simplifying strategies.

Having looked at the various ways in which UI can be presented (and utilised) we turn our attention to other information cues commonly available in a judgmental forecasting task²⁷. Apart from UI, what other information cues are available to the judgmental forecaster/decision-maker? How does the mix or load of information cues impact the potential utility of UI?

Generally, judgmental forecasters have at their disposal a relatively small number of information types. This information can be simplistically categorised into four groupings: historical data (for example, a time series), forecasts (either statistical or expert based), UI generally derived from the historical data and relating to the forecast, and causal/contextual information. Various combinations of these information types are used to develop explicit or implicit judgmental forecasts and/or make judgmental decisions. The impact that

²⁷ The primary users of UI are judgmental forecasters and decision-makers, and accordingly this review focuses on those roles.

historical data and forecasts have on the decision-maker will now be discussed individually²⁸. This will then be followed by a review of problems arising when various cues are presented simultaneously. Finally, some ways to alleviate cue problems will be discussed. Figure 2-4 conceptually illustrates this section's structure, showing how UI utility is influenced by information load and cue mix.

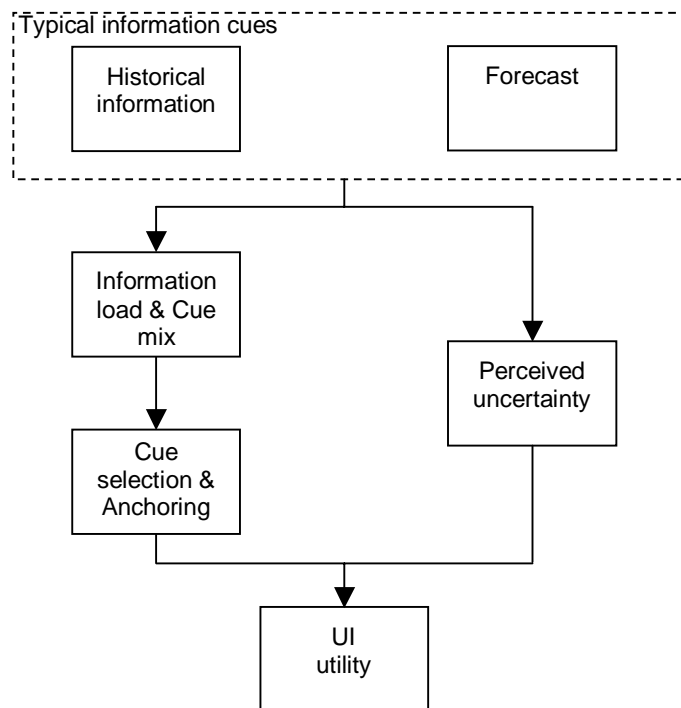


Figure 2-4: The influence of certain information load, cue mix, cue selection, anchoring and perceived uncertainty on UI utility.

It should be noted that due to the lack of research into the utility of UI (see Section 2.3) it is necessary to look at other areas of research that may provide clues as to how cue mix and information load may influence UI utility. The 'calibration' research, looking at the production and calibration of judgmental

²⁸ The impact of providing UI has already been discussed in section 2.2.2.1.1, so will not be discussed here. The provision of causal information will not be discussed, as it is not directly related to this work. Reviews of literature relating to the provision of causal information in judgmental forecasting can be found in Webby & O'Connor (1996) and also Goodwin & Fildes (1999).

confidence intervals, is one such area of investigation. Yaniv & Foster (1995) demonstrated the relevance of calibration to UI utility by showing that the upstream production of judgmental confidence intervals is closely related to the downstream utilisation of UI since the producers of confidence intervals think about how their information will be used, and lean towards being informative rather than technically accurate.

How perceived uncertainty and the presence of historical information influences UI utilisation

Historical data is the most elementary information cue that can be provided to a judgmental forecaster. It has already been mentioned that people are quite good at assessing certain aspects of historical information, such as central tendency (Beach & Swensson, 1966; Pitz et al., 1976), but very bad at evaluating other aspects like variability (Beach & Scopp, 1968; Hogarth, 1975; Lathrop, 1967; Pitz et al., 1976). However, people's ability to assess historical data is probably limited to situations where there are enough data points to adequately infer underlying patterns of behaviour in the data. For example, Pitz et al. (1976) obtained their results by providing subjects with nine historical data points. Fildes (1989), in an editorial highlighting future research directions in forecasting, emphasised the need for more research into situations where either little or no historical data is available to the forecaster (such as new product forecasting).

How does the provision of historical information influence the perception of uncertainty, and therefore the utility of UI? The decision-maker perceives the level of uncertainty in a situation based on the information or signal observed in

the historical data. This formation of perceived uncertainty might even dominate the presence of any objective UI. The concept of perceived uncertainty is discussed further in section 2.2.4.1 (see Sitkin & Pablo, 1992, for the role of perceived risk in decision making).

Lawrence & Makridakis (1989) looked at the production of judgmental confidence intervals and found that they were heavily influenced by the trend of historical data. In line with observations of our inability to deal with data variability, the same study found the judgmental confidence intervals to be insufficiently influenced by randomness. Also, O'Connor & Lawrence (1989) showed that confidence intervals were influenced by time series seasonality.

The behaviour observed in the judgmental production of UI should influence the way in which UI is utilised. First, willingness to believe in UI may be influenced by trend in historical data. Second, the extent to which UI is believed may be dependent upon the observed randomness of the historical data. In particular, data with extreme randomness (very high or very low) may reduce trust in UI (such as a confidence interval) since judgmental confidence intervals are insufficiently influenced by randomness.

Providing historical information together with other cues can alter the utility of the other information cues. For example, in an investigation of interactive forecasting support systems, Lim & O'Connor (1996b) found that participants favoured their own judgment rather than relying on the support system. The authors' speculated that one reason why this may have happened was because participants felt they could get all the information they required from the historical time series information. This was despite the fact that more useful

information cues were available. So the provision of historical information may provide decision-makers with a favoured information cue that is not necessarily as accurate or as useful as other information cues.

How forecasts may affect UI utilisation

To aid the job of decision-makers and judgmental forecasters, historical data is often replaced by, or supplemented with, a statistical forecast. A variety of statistical forecasting methods exist, and their advantage over historical information is that they objectively summarise the historical information into a single figure and simultaneously indicate what a future value may be. When statistical forecasts are provided to judgmental forecasters and decision-makers this information is frequently ignored and inappropriately modified.

Even when provided with a statistical forecast, people inevitably produce their own judgmental forecast either as a point of comparison or out of lack of trust for the statistical forecast. Judgmental forecasting then becomes an integration task. Lim & O'Connor (1995) studied this integration process and found "an extremely strong tendency to place too much faith in one's own forecast rather than that of the statistical forecast (p. 162)".

As mentioned previously, Goodwin & Fildes (1999) found forecasts to be inappropriately used, modified and ignored by individuals involved in judgmental forecasting. Earlier, Lawrence, Edmundson & O'Connor (1986) showed the ineffectiveness of judgmentally combining forecasts compared with simple averaging. Webby & O'Connor (1996) reviewed the literature on judgmental and statistical forecasting and concluded that probably the most effective method is not direct adjustment of a forecast but rather mechanically combining statistical

and judgmental forecasts. In this way the judgmental adjustments are dampened, but still allow the integration of additional, contextual information.

So despite the obvious benefits of an objective, statistical forecast, people do not optimally use them, instead desiring to mould them into a product of their own judgment. Given this behaviour, how might the provision of a forecast influence the utility of UI? The preference of people for point estimates over ranges or probability distributions (for example, see Conrath, 1973) indicates there is some chance forecast information might dominate and limit the impact of UI when provided in a form such as a confidence interval. However, even if the UI does grab the attention of the forecaster it may be judgmentally adjusted just as we saw forecasters inappropriately adjust statistical forecasts.

Cue selection and anchoring with UI

What happens when several information cues are provided to a decision-maker simultaneously? People are then faced with a situation where they need to select what cues they wish to use, or perhaps combine and process multiple cues. Additional problems are then encountered, including anchoring on one cue rather than another, being overloaded with too much information, or determining which information is relevant.

When provided with multiple information cues people tend to have difficulty aggregating all the pieces of information. For example, Lim & O'Connor (1996b) found this when they analysed people's efficiency at selecting and using different pieces of information. As a result people tend to favour one cue, or a summarised representation of all the cues such as a mean, over the others. Subsequent decisions are then 'adjusted' away from this chosen 'anchor'. For

stationary series Lawrence & O'Connor (1992) found the anchor to be the long-term mean of the series. Forecasts were then adjusted based on the difference between the long-term mean and the last observation. The most recent segment (difference between the most recent observation and its predecessor) slope was also influential. Harvey, Bolger & McClelland (1994) found that for series with a trend the anchor was the last observation, while the adjustment was relative to the last segment. In an investigation into the accuracy of judgmental extrapolation of time series Welch, Bretschneider & Rohrbaugh (1998) found a similar 'recency' bias. They concluded that judgmental forecasters would perform best if asked to make more immediate forecasts of just one or two steps ahead since forecasting over a longer horizon, such as three steps or more, increases the negative effects of this recency bias. In the same study, results also suggested that the provision of statistical information (explicitly highlighting characteristics of underlying data such as long-term levels and trends) in addition to historical information was beneficial because "the forecaster is able to use these statistics as cues for a more appropriate anchoring and adjustment process (p. 108)".

In practice forecasters often have significant contextual information available to them. Despite this, a common anchor for judgmental forecasting practitioners is still their last forecast (Fildes, 1991; Bromiley, 1987).

One way of assessing how good people are at understanding uncertainty is to look at how they assess subjective probability. Like judgmental forecasts, judgmental confidence intervals are also influenced by anchoring and adjustment behaviour. The production of judgmental confidence intervals

theoretically should be symmetrically anchored on a forecast. However O'Connor, Remus & Griggs (2001) found judgmental confidence intervals to be asymmetric, with them being anchored to the difference between the last actual value and the forecast.

It is not known exactly how people select what information to anchor on when there are multiple anchors to choose from. There is no theory of anchor selection, with the theory of anchoring still based on the work of Tversky & Kahneman (1974). Whyte & Sebenius (1997) investigated the effect of multiple anchors on anchoring in a judgmental task in order to improve our understanding of the anchor selection process. Their research demonstrated that the selection of an anchor does not appear to depend on the anchor's relevance to the task. Anchor selection was impacted significantly more by the order in which anchors were encountered, regardless of whether the anchors were relevant to the task or were in fact useless information. Those pieces of information encountered immediately before a decision was to be made were anchored on.

Apart from acting as 'red herring' anchors, research into the impact of seemingly useless information cues has produced other interesting results. Davis, Lohse & Kottmann (1994) conducted an experiment investigating the impact that additional information cues (financial news information) had on a stock earnings forecasting task where the baseline information provided were time series. They found that "the additional news information, whether redundant or nonredundant, had the effect of degrading performance while increasing confidence (p. 253)". In an empirical study examining the

effectiveness of interactive forecasting support systems Lim & O'Connor (1996b) found people often chose less-reliable information when confronted with multiple information cues. Simnett (1996) also observed inadequacies with information selection, although this was worst with those participants having limited expertise.

The results of Davis, Lohse & Kottemann (1994), Lim & O'Connor (1996b) and Whyte & Sebenius (1997) point to people being poor judges of the utility of newly obtained information. However, in a related piece of work, Remus, O'Connor & Griggs (1998) observed encouraging behaviour by individuals completing a judgmental forecasting task and provided with additional contextual information of unknown and varying correctness. Participants produced more accurate forecasts with correct information. More importantly there was no difference between the forecast accuracy of those provided with incorrect information and those who were not provided with any contextual information. This implies that participants were able to discern what information was useful.

The impact of information load on UI utilisation

People's ability to determine what information is useful to them in a decision-making or judgmental forecasting task is ultimately limited by the amount and quality of information available to them. More information is useful, but only up to a point, after which humans suffer from limited processing

capacity. Unfortunately, people have been shown to seek too much information²⁹ (for example, Tversky & Edwards, 1966), which exacerbates the limited processing problem, and leads to deviations from optimal strategy. The acquisition of too much information can be exacerbated in positive-payoff situations. Connolly & Gilani (1982) showed individuals were willing to purchase more information in such situations. The purchasing of additional information can be seen as a risk-reducing strategy, and as such those individuals were behaving in accordance with Prospect Theory (Kahneman & Tversky, 1979).

What happens when people receive too much information? It has already been mentioned that participants in Davis, Lohse & Kottmann (1994) were unable to utilise additional information in forecasting stock earnings, whether the information was redundant or not. Several authors have discovered that too much information in a decision-making task can overload the decision-maker (for example, Conrath, 1973; Parry, 1967) and detrimentally affect decision performance. A number of studies have also indicated that as information load increases there is an associated reduction in the reliability of judgment (Einhorn, 1971; Hogarth, 1987; Stewart et al., 1992; Lee & Yates, 1992)³⁰. Information overload has other consequences including impacts upon confidence in judgment and time to execute a decision.

²⁹ It should be noted that in some situations people have been shown to behave in a reverse manner. For example, Lim & O'Connor (1996b) found participants in their forecasting support system experiment under-acquired information.

³⁰ A number of studies investigating the impact of large numbers of information cues were reviewed in Faust (1986).

Casey (1980) investigated the impact of varying the load of accounting information. The author found that varying information load had no impact on the accuracy of decisions. However, increased information load did significantly lengthen time taken to make decisions. This also occurred in Onken, Hastie & Revelle (1985) although for very high information loads decision time started to decrease, presumably because the participant's processing capacity had been reached and simplification strategies were put in place. In a similar study Simnett (1996) had contrasting results. Information load did limit information processing capability. Interestingly this was regardless of experience (participants were practising auditors covering all experience levels). Simnett (1996) found information load did not influence confidence in judgment, differing from the findings of an earlier study by Snowball (1980). Snowball (1980) manipulated information load by varying processing time rather than the number of information cues in another accounting task. He found that increased expertise and lower information load interacted to make participants record wider confidence intervals. In other words, the experts knew when they didn't have enough information to complete the task effectively.

Another way to manipulate information load is through summarisation of data. Summarising and replacing several information cues with one can dramatically reduce information load and possibly illuminate additional insight. Chervany & Dickson (1974) manipulated information load in this way and saw summarised information significantly increase the quality of decisions. Simultaneously confidence in those decisions was reduced, and decisions took longer to make. Conrath (1973) observed choice behaviour was impacted by the format of probabilistic information, with those receiving more summarised

information making more optimal decisions and better able to comprehend task uncertainty.

Summarising data is one way to combat the problems of information selection and information overload. Alternatives include emphasising, and improved labelling for, certain information cues. Welch, Bretschneider & Rohrbaugh (1998) found that emphasising specific cues, like the last data point or summary information such as the average of the last three historical data points, significantly decreased forecast error. These were cues already provided in raw, time series format, but were explicitly highlighted when asking for the forecast. In contrast with the information overload studies this work saw benefit in additional information, but presumably because it was emphasising and summarising existing (available) information cues.

Providing better labels for information cues should also help selection and overload issues. Congruent labels have been found to improve information retrieval, reduce the number of strategies used and therefore increase consistency and the quality of decisions for probability related information cues (Sniezek, 1986).

Cue mix and information load can dramatically impact decision-making performance, predominantly because the limited processing capacity of humans leads us to attempt simplifying strategies that introduce bias and behaviour that is less than optimal. Such behaviour has the potential to negatively impact the utility of UI. Characteristics of historical data, such as trend and the position of most recent data points, may influence the credibility of UI. There is evidence that historical information is preferred by decision-makers, and could lead to UI

being ignored. UI should be provided in a simplified format, and should not be diluted by a large number of information cues that may risk the occurrence of information overload. If UI is to be provided at the same time as a significant amount of historical information, it may be best to emphasise the UI immediately before a decision needs to be made so that the decision-maker is less likely to favour the historical data and attempt to independently (and most likely inadequately) evaluate uncertainty. Ultimately, to understand and study the utility of UI, other information cues that can act as a (possibly less accurate) proxy for UI, such as historical data, must be carefully controlled.

2.2.2.2 Task Uncertainty

Key Messages:

- Task uncertainty refers to the level of uncertainty associated with a task's information cues, potentially operationalised through data variability.
- People perceive implicit task uncertainty as signal rather than randomness and attempt to simulate and incorporate this uncertainty into their decisions, resulting in less predictable and worse decisions.
- Judgmental confidence intervals are insufficiently influenced by randomness and inappropriately manipulated by scale: there appears to be regression towards an expected medium level of randomness.
- Unexpected task uncertainty reduces the ability to assess variability and highlights the need for providing UI in decision-making.

The purpose of explicit, objective UI is to communicate the underlying uncertainty in a situation. The UI represents the actual uncertainty (see Figure 2-6, p. 90). Can the underlying, implicit uncertainty perceived through other information cues impact the way in which objective UI is interpreted? For example, does an environment with low predictability alter the willingness of a judge to rely on information cues, including those communicating objective UI?

This section covers how the decision maker perceives task uncertainty and how this might influence UI utility. This issue can be investigated indirectly, by looking at how task uncertainty impacts the accuracy of forecasting, and directly by exploring the impact that task uncertainty has on the production of judgmental confidence intervals.

In forecasting, one has to constantly deal with perceptions of uncertainty for a given task and therefore it provides a good setting for this discussion. Whereas complexity is met with simplifying heuristics in the forecasting task, uncertainty or variability is coped with by attempting to incorporate and simulate randomness. As a result forecasting performance is degraded. People are incapable of distinguishing systematic behaviour in data from noise³¹. Several studies have shown people to interpret random series as systematic (Bolger & Harvey, 1993; Eggleton, 1982; Goodwin & Fildes, 1999; Harvey, 1995; O'Connor & Lawrence, 1992) and to then add noise to the output of their forecasting heuristics. For example, Harvey (1995) found participants in an

³¹ Noise is the random, irregular, or unexplained component in a measurement process. Noise can be found in cross-sectional data as well as in time-series data (Armstrong, 2001). Noise is also referred to as variability, predictability, randomness and randomness in some studies. In a study on judgmental weather forecasting, Stewart et al. (1997) defined task predictability as “the strength of the relation between the observed event and the cues (p. 208)”.

experiment consistently attempted to simulate noise in their forecasts. This behaviour was then confirmed in a second experiment, which asked participants to simulate the time series. O'Connor, Remus & Griggs (1993) observed comparable behaviour in a judgmental forecasting experiment using time series with discontinuities. In this study forecasts were changed due to random fluctuations by identifying nonexistent signals especially under high variability.

The effect of this behaviour is less accurate forecasting. Goodwin & Fildes (1999), O'Connor, Remus & Griggs (1993) and Stewart, Roebber & Bosart (1997) all found higher levels of data variability or randomness (or the presence of discontinuities) reduced judgmental forecasting performance. For example, data discontinuities clearly caused people to perform worse than statistical forecasts (O'Connor, Remus & Griggs, 1993)³². Harvey (1995) speculated that the reason people incorporate noise into their forecasts is either (i) the removal of noise from the forecast seems inconsistent with the randomness in the data, or (ii) they are simply trying to represent the randomness in their forecasts. Less predictable situations also render certain forecasting strategies inadequate. Anchoring on similar situations and adjusting away (a pattern matching strategy) does not work in less predictable environments (Hoch & Schkade, 1996).

People's desire to find a signal in data, whether it exists or not, is beneficial in some situations. Smithson (1997) found that judgmental forecasts were more accurate with chaotic time series than with series having

³² Sanders & Ritzman (1992) slightly disagreed with these results, finding judgmental forecasts to be worse than statistical forecasts for low and low-medium variable data, but for there to be no difference in forecast accuracy for high variability.

randomness. This is not surprising considering the existence of occasionally repeatable patterns of behaviour in chaotic data.

Data variability can be imitated by manipulating the presentation of data. For example, the scale used to present graphical time series data can be altered to manipulate perceived data variability without actually modifying the data. Lawrence & O'Connor (1992) conducted a judgmental forecasting experiment with results showing that scale did not influence forecast accuracy. Presumably participants were able to discern scale manipulation from data variability.

Increased data variability also causes people to be less consistent in their judgments. As Stewart et al. (1997, p. 206) stated:

“Judges respond to unpredictable tasks by behaving less predictably themselves.”

Several studies have observed this behaviour, including Brehmer (1976, 1978), Camerer (1981) and Harvey (1995). Harvey (1995) also observed that people were unable to judgmentally assess data variability because they anchored on the range created by the upper and lower endpoints of time series data and used them as a crude assessment of uncertainty. Range can be highly influenced by seasonality and therefore may not be a valid measure of uncertainty in the near future.

In separate studies Lawrence & Makridakis (1989) and O'Connor & Lawrence (1989) showed that judgmental confidence intervals were insufficiently influenced by data variability. However, the judgmental confidence intervals were significantly influenced by trend (Lawrence & Makridakis, 1989)

and time series difficulty³³ (O'Connor & Lawrence, 1989). Lawrence & O'Connor (1993) obtained more conclusive results regarding presentation scale and data variability. Small scale and low variability led to excessive confidence intervals (under-confidence) whereas large scale and high variability produced narrow confidence intervals (over-confidence). Whether data variability causes people to insufficiently alter their judgmental confidence intervals or leads them to regress towards some expected mid-range variability, these results are a good demonstration of our inability to assess variance, and a strong argument for the provision of UI.

Data variability detrimentally impacts forecast accuracy and estimates for forecast uncertainty, but there are some simple ways to reduce the impact of less predictable situations on judgment. Goodwin & Wright (1994) explain one way to better deal with very large period-to-period changes is to provide fewer data points more widely spaced apart, thus dampening the variability. Wagenaar & Timmers (1978) effectively demonstrated this proposal. This is counter to the salience hypothesis from Andreassen & Kraus (1990), which recommends time series characteristics be best discerned through the provision of more data points.

We have seen that people interpret noise as signal and attempt to simulate this noise in their judgments, resulting in reduced forecasting accuracy. Variability also causes inconsistency in judgment and renders people worse at assessing variance. Judgmental confidence intervals are insufficiently

³³ Time series difficulty or complexity is generally proportional to the presence of trend and/or seasonality.

influenced both by randomness or scale (which can be thought of as perceived randomness), regardless of whether either manipulation is high or low, and there appears to be regression towards an expected medium level of randomness. All these demonstrations of behaviour act as strong arguments for the provision of UI. These results indicate when UI is provided with significant amounts of historical data there may be a reluctance to believe the UI. Better understanding of underlying uncertainty may occur when UI is provided with limited historical data and when provided with a statistical forecast.

2.2.3 Environment

Key Messages:

- Environmental factors should be of less significance in determining the utility of UI.

The first two parts of Beach & Mitchell's model (1978), person and task, have been reviewed for potential decision characteristics that may influence the utility of UI. The third part of the Beach & Mitchell model, environmental factors, will be reviewed in this section for any possible impact on UI utility. It is not expected that environmental factors will be as influential on UI utility as some of the task factors, however major environmental factors which will now be reviewed include time pressure (2.2.3.1), incentives (2.2.3.2), task setting (2.2.3.3) and feedback (2.2.3.4).³⁴

³⁴ The payoff or loss function can also influence the utility of UI, however this will not be discussed in detail. It will be mentioned in section 2.3.1, which defines a model of contingent effects for decision-making under uncertainty with UI.

2.2.3.1 Time Pressure

Key Messages:

- Time pressure introduces erratic behaviour that diminishes decision performance.
- Time pressure may lead to less risky behaviour, which could influence UI utility.

A key element of any decision-making environment is the time pressure placed on the decision-maker. Several studies have shown increased time pressure detrimentally effects decision performance (for example, Ahituv, Igbaria & Sella, 1998; Kerstholt, 1994; Rothstein, 1986; Wright, 1974). What commonly happens is people manage to stick to their original decision strategy, but erratic behaviour is introduced which diminishes performance (Kerstholt, 1994; Rothstein, 1986). There is no reason to assume that this main effect would be any different regarding the provision of UI. However, time pressure does elicit secondary effects that may influence UI utility.

Increased time pressure leads to the equivalent results of less risky behaviour (Ben Zur & Breznitz, 1982) and wider judgmental confidence intervals (Snowball, 1980). The production of a wider judgmental confidence interval is a symptom of less risky behaviour. Any catalyst that influences perceptions of uncertainty (such as judgmental confidence intervals) has the potential to influence the utility of UI. Increased time pressure could therefore potentially limit the extent to which people are willing to trust and use UI.

This result, coupled with the fact that time pressure also leads to the overweighting of negative evidence (Wright, 1974) leads to the conclusion that in any study investigating a main effect due to UI, time pressure should not be manipulated. This is something for a follow-up experiment.

2.2.3.2 Incentives and Motivation

Key Messages:

- Incentives have been found to help in the majority of studies investigating probability judgment.

The provision of financial incentives in empirical studies of human behaviour is a highly contentious issue. Researchers are polarised into those believing financial incentives are necessary to motivate and maximise the performance of experimental participants, and those who think financial incentives not only detrimentally affect judgmental performance, but also cause deviations from optimal decision-making strategies³⁵. Moreover, there is a clear delineation between the experimental designs used by economists and behavioural psychologists. As part of a comprehensive comparison between the experimental practices of economists and psychologists Hertwig & Ortmann (2001) reviewed the use of financial incentives. In a ten-year sample (186 studies) of behavioural decision-making experiments, 26% used financial incentives. Camerer & Hogarth (1999) completed a similar analysis for

³⁵ There is also some evidence that incentives have no effect. For example, Wiseman & Levin (1996) found no difference between the impacts of real and hypothetical consequences.

experimental economics over three decades: all studies used financial incentives.

The main focus of Camerer & Hogarth (1999) was to complete a meta-analysis reviewing the impact of financial incentives on psychological experiments, and they had mixed results. Incentives were useful in some areas of research (for example, judgment and decision-making) and not in others. The authors found that incentives reduced variance in performance, but there was no main effect on performance: that was dependent upon what type of task was performed. Also, incentives were never able to remove violations of rationality alone. Based on their capital-labor-production framework, Camerer & Hogarth (1999) concluded that incentives were required for difficult tasks, where more intellectual capital was required.

Hertwig & Ortmann (2001) concluded by urging psychologists to use financial incentives as part of their experimental design. They were convinced by the results of Camerer & Hogarth (1999), and because of the reduced variability resulting from incentives.

Although some studies have found incentives to reduce calibration in probability judgments and increase overconfidence (Henry & Sniezek, 1993; Wallsten, Budescu & Zwick, 1993), the findings of Camerer & Hogarth (1999) point 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 example, in an investigation into the effectiveness of confidence interval information Foong (1992, 2003) found confidence intervals to be useful in a judgmental forecasting

task only when financial incentives were present. When no financial incentives were offered, the control group (those without the confidence interval information) actually performed better.

2.2.3.3 Task Setting

Key Messages:

- Any task used to investigate UI utilisation must be realistic enough so that UI can be applied within a believable context, while balancing the considerable advantages of cost, control and generalizability of results obtained from laboratory experiments.

The vast majority of empirical psychological decision-making research is conducted using a laboratory setting. Laboratory based research provides several benefits, but is particularly attractive because of cost and control. The cost of running a laboratory study, in both time and money, is much less than what would be incurred for related field research. Laboratory studies also provide considerably more control over the task and participants, and accordingly significant results are easier to find. However, observations of human behaviour in a laboratory setting only become valuable when they can be generally applied. This is often not possible, and is seen as a fundamental weakness in laboratory based research.

Winkler & Murphy (1973) and Beach (1975) raised doubts about the generalizability of results obtained from laboratory based psychology studies, both using the observation of conservatism as an illustration, while others have questioned the artificiality of the laboratory setting (for example, Dobbins, Lane

& Steiner, 1988; Griffin & Kacmar, 1991). One aspect of this artificiality is the reduced impact of visceral influences³⁶ (Loewenstein, 1996).

In light of these concerns, and in response to Hertwig & Ortmann's (2001) comparison of economic and psychological experimental procedures, Hogarth (2001) concludes that psychologists need to better test the generality of their theories. The way to do this is to pursue more realistic psychological experiments, something that has been suggested in this field at least since Winkler & Murphy (1973), and studies such as Fischhoff (1996) have examined ways of conducting experiments in a more realistic manner. Fischhoff (1996) outlines ways of conducting less structured, more open-ended, experiments in order to increase external validity of results where participants have much more freedom in the way they make decisions.

Despite this call for more realistic psychological and decision-making experiments, caution must be taken since realistic experiments have their own pitfalls. Real-world tasks can be just as ecologically invalid as laboratory studies because of the similar methods used (Bolger & Wright, 1994).

Earlier the conclusions of Dawes (1999) were discussed and it was postulated that people are unable to make decisions utilising or inferring probabilistic information without significant context (an accompanying story). The implications of that finding is that it may be even more important for tasks

³⁶ Visceral influences (or factors) include states such as hunger, thirst, pain, moods and emotions, and cravings or addictions. Visceral influences are characterised by a direct pleasure or painful (generally painful) impact, and an effect on the relative desirability of some good or action.

involving probabilistic information (such as UI) to be realistic, assuming that realistic experiments provide the best and most believable stories.

2.2.3.4 Feedback/Irreversibility

Key Messages:

- Outcome feedback can have both positive and negative effects upon decision performance.
- Feedback may not be effective in probabilistic inference tasks, such as when UI is provided, because the function linking information cues to performance may be too complex.

When individuals must make a number of similar decisions within one system, performance, or outcome, feedback has the potential to influence judgment by allowing the correction and refinement of decision strategies. Many researchers believe feedback is of significant benefit to judgment. Bolger & Wright (1994) went so far as to state “that good performance will be manifest when both ecological validity and learnability³⁷ are high, but that performance will be poor when one of these is low (p. 1)”. In a summary of literature on feedback, Hertwig & Ortmann (2001) also concluded that feedback could improve judgment, although they stressed that this can only occur with a significant number of repeated trials. Davis & Kottmann (1995) found other feedback benefits relating to decision rule use. Feedback increased perceived

³⁷ Learnability means the opportunity to learn, and therefore inevitably refers to the availability of feedback information.

usefulness, level of usage, and decision performance with production planning decision rules.

Feedback doesn't always improve judgment. For example, feedback has no effect on the occurrence of anchoring-and-adjustment behaviour (Remus & Kotteman, 1995). Feedback can also detrimentally affect judgmental forecasting performance³⁸. When the decision-making environment contains noise (which in reality is frequently), feedback may only compound the chance of noise being interpreted as signal since the feedback will contain noise also (Hammond, Summers & Deane, 1973; Kessler & Ashton, 1981; Siegel-Jacobs & Yates, 1996). The result is judgments that also contain more noise, and less accuracy.

Even if feedback is high quality and noise free, people don't always utilise feedback effectively. In judgmental forecasting the provision of a statistical forecast can be thought of as a form of task information feedback (Goodwin & Fildes, 1999). Lim & O'Connor (1995, 1996a) demonstrated that this form of task information feedback can improve judgments, but people under-utilised the feedback, relying too heavily on their initial judgments. Providing graphical, or more salient, feedback had little effect.

Feedback has other negative effects. Feedback can cause people to change their behaviour in suboptimal ways, for example by trying to avoid post-decisional regret. Zeelenberg (1999) found that anticipated regret promoted both risk-averse and risk seeking behaviour, depending on which option was

³⁸ For a summary of the effects of feedback on forecasting see Goodwin & Wright (1994) and Goodwin & Fildes (1999).

regret minimising. Knowledge of prior gains or losses can influence decision-making behaviour. People may be more willing to accept risk after experiencing a gain (Sullivan & Kida, 1995). Frequent feedback can also exacerbate myopic loss aversion, where an aversion to losses and a tendency to frequently evaluate outcomes leads to severe loss and risk aversion (Thaler et al., 1997).

Perhaps feedback is not well used by people because we have difficulty understanding the whole system, and how feedback can and should be utilised in a complex situation (Diehl & Sterman, 1995). The alternative to simple outcome feedback is to provide feedback regarding the system that the decision-maker is operating in, or feedback relating to the decision-making procedure. This task information feedback (commonly referred to as cognitive feedback) was superior to outcome feedback in numerous studies, including general decision-making and judgmental forecasting (Balzar et al., 1992; Balzar et al., 1994; Benson & Önköl, 1992; Hammond, Summers & Deane, 1973; Kessler & Ashton, 1981; Remus, O'Connor & Griggs, 1996; Sanders, 1997; Siegel-Jacobs, 1996). However, in a related study Davis & Kottemann (1995) found rule description not as effective as outcome feedback, but explained it would be useful in situations when it's not feasible to provide feedback.

In summary, in decision-making and judgmental forecasting feedback can have both positive and negative effects upon decision performance, and it appears that cognitive feedback is often more effective than outcome feedback. The effects of feedback on probabilistic inference and confidence in judgment are more specific.

There is some belief that no feedback is effective in probabilistic inference tasks because too often the function linking information cues to performance evaluation is too complex, and accordingly people do not understand how to utilise the feedback to change their behaviour and improve performance (for example, Brehmer, 1974, 1976, 1980; Brehmer & Kuylenstierna, 1978). Hammerton (1973) found subjects performed better without feedback in a probabilistic inference task. However, feedback has been shown to be of use in the calibration of confidence in judgment (O'Connor & Lawrence, 1989; Bolger & Önkcal-Atay, 2004). Confidence in judgment increases with both practice and effort (Paese & Snizek, 1991). In O'Connor & Lawrence (1989) outcome feedback moderated this over-confidence and lead to more calibrated judgment. Benson & Önkcal (1992) and Bolger & Önkcal-Atay (2004) also improved calibration, but through specific calibration feedback, not through outcome feedback.

In the real world, tasks vary in learnability (Bolger & Wright, 1994): the ability to make use of feedback. Individuals, like weather forecasters or race handicappers, have immediate access to feedback that is clear in meaning and allows for refinement of decision strategies. Others, such as clinical psychologists, physicians, insurance underwriters of low probability events (Kunreuther et al., 1995), or audit managers, have greater difficulty learning from their judgments due to delays in feedback, and unclear causal linkages between decision and outcome (Bolger & Wright, 1994).

Given that feedback is of doubtful use in probabilistic inference tasks, and there appear to be a number of situations in reality where feedback is

unavailable or of little use, there is a strong ecological case for investigating UI utility without providing feedback.

2.2.4 Interactions of Person and Task

Key Messages:

- Interactions of person and task have the potential to influence the interpretation and utility of information cues like UI.

So far we have reviewed a model of decision-making based on person, task and environment in order to develop a customised model for understanding the utility of UI. To complete this model, personality traits apparent during decisions under uncertainty that are functions of person and task must be considered.

Beach and Mitchell (1978, p 445) purposely did not include any reference to personality related variables in their model of decision strategies:

“Note that we have omitted most of the “personality” variables that might be expected to be included (e.g. risk-taking and risk aversion). This is because few solid data exist that reliably relate personality characteristics to specific decision strategies.”

Kozielecki (1981) agreed with this assessment, explaining that attitudes towards risk were not a personality trait, but rather were in fact a function of environmental factors and other personality traits such as anxiety and aggression. O'Connor (1989) interpreted previous experimental findings to show that one such personality trait, ability to calibrate judgments of confidence, can to some extent be explained by differences in person, task and environment.

Whether these concepts are personality traits or are interactions of person and task, they have the potential to influence the interpretation and utility of information cues even though aspects of person, task or environment may themselves influence the traits. Two such traits involve perceptions of probability, namely, the risk propensity of the individual, and the calibration of their decision confidence. In the next two sections an individual's risk profile and perception of risk (section 2.2.4.1) and the calibration of their confidence (section 2.2.4.2) will be analysed for their potential impacts on UI utility.

2.2.4.1 Risk Profile and Perceived Risk

Key Messages:

- Because risk propensity and perceptions of risk can alter decisions, these traits have the potential to influence the utility of UI.
- UI may be most useful in risk-averse situations by making underlying uncertainty salient, and reducing the tendency to treat decisions as unique and in isolation.
- Providing multiple reference points (such as target levels) relating to current and future performance may optimise UI utility.
- The format of UI should be carefully selected in order to minimise ambiguity avoidance.
- UI utility appears dependent upon the interaction of underlying uncertainty and the decision-maker's intuitive expectation of risk.

People’s perceptions of risk can influence the way in which they make judgments (Sitkin & Weingart, 1995). This is apparent from the integration of research on perceived risk and models of decisions under risk, as reviewed earlier (for example, Brachinger & Weber, 1997; Jia, Dyer & Butler, 1999; Sitkin & Weingart, 1995). Because perceptions of risk and risk propensity can alter decisions, these traits have the potential to influence the utility of information cues, including UI. This section investigates ways in which risk perceptions and propensities influence UI utility (see Figure 2-5).

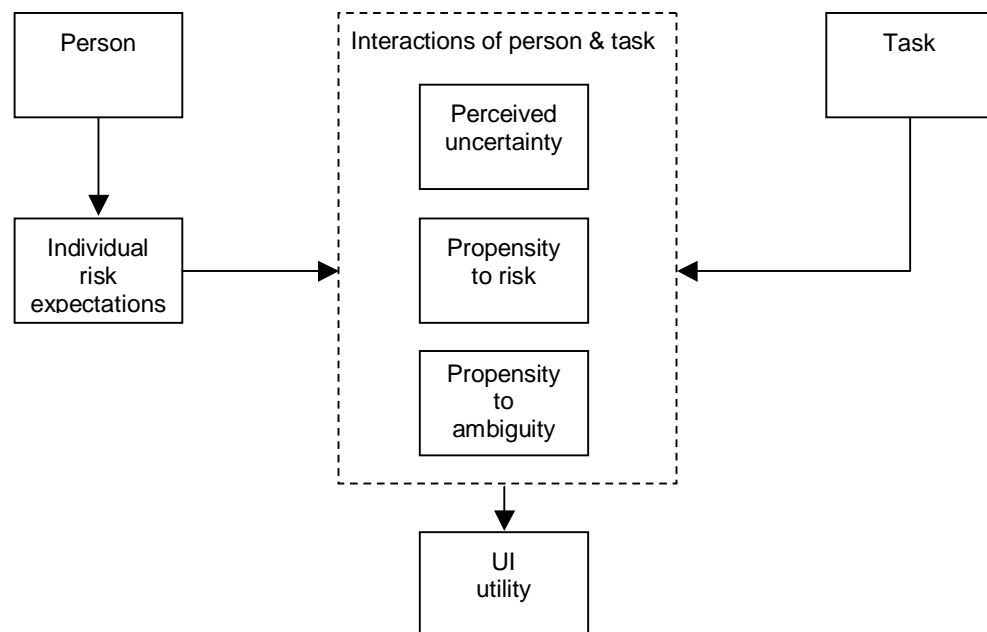


Figure 2-5: The mediating effects of perceptions and propensities on UI utility.

Perceived uncertainty

People have a perception of uncertainty that is formulated through an individual’s propensity to risk and the situation or task. The decision-maker perceives the level of uncertainty in a situation based on the information or signal observed in the available data.

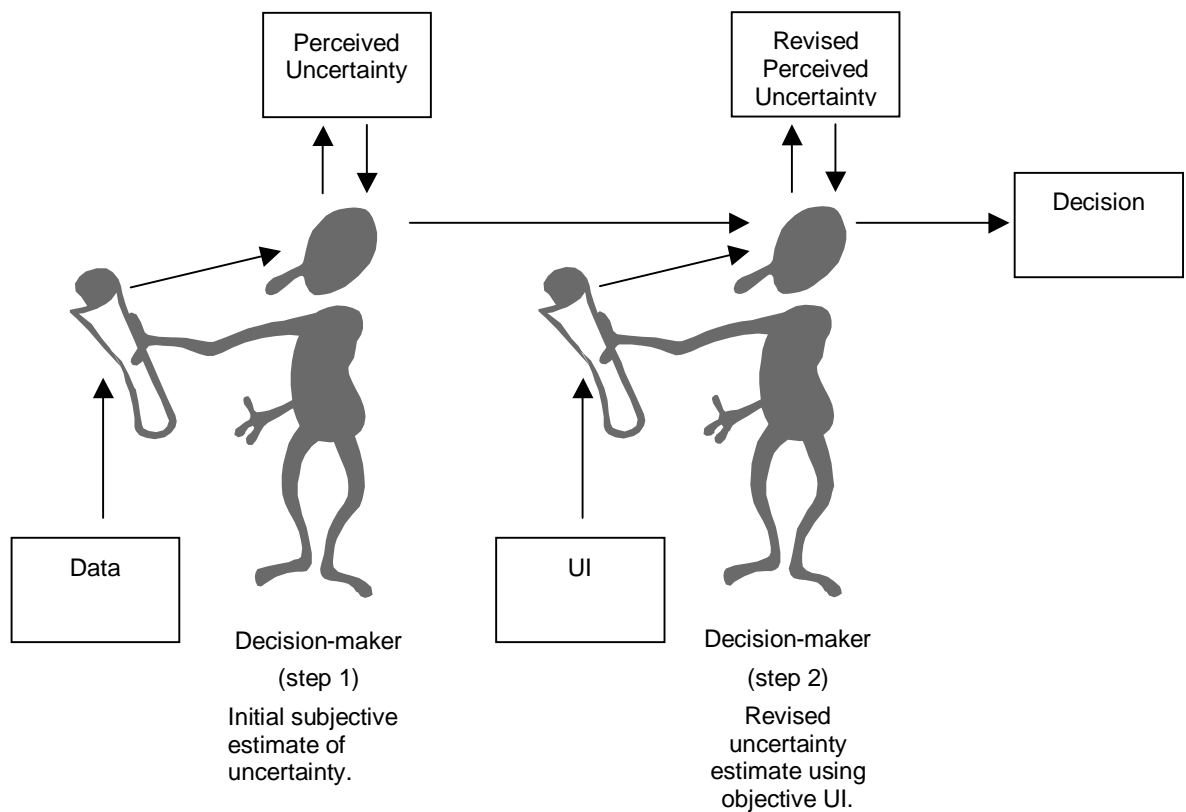


Figure 2-6: The relationship between objective UI and perceived uncertainty.

It is important to distinguish between the concepts of objective UI and perceived uncertainty (PU). Figure 2-6 helps to explain the relationship between objective UI and PU. Figure 2-6 describes a typical situation where a decision-maker is placed in a situation where a decision needs to be made using a particular set of data. Accompanying the data (and most likely based on the same data) is some objective UI. The decision-maker first perceives the underlying uncertainty of the task based on information or signal observed in the data, such as historical information (see section 2.2.2.1.2). These thoughts may well be unconscious. The revision of PU through the incorporation of explicitly communicated UI occurs afterwards. While the perceptions about

uncertainty should be most accurate when based on objective UI, clearly it is possible that perceived uncertainty might be based upon other, less accurate data. Of course, perceptions of uncertainty exist even when objective UI is not provided.

Others have investigated the role of PU. For example, risk perception has been of interest to researching in marketing (Bauer, 1967; Bettman, 1973; Baird & Thomas, 1985). Bettman (1973) and Baird & Thomas (1985) introduced the concept of perceived risk into models of consumer behaviour. Sitkin & Pablo (1992) went beyond consumer behaviour and proposed a model of risky decision-making that placed risk propensity and risk perception in a central role. The authors speculated that the effects of many variables on risk taking were mediated by risk propensity and risk perception. Sitkin & Weingart (1995) found empirical support for the mediating effects of risk propensity and risk perception from two decision-making experiments. Brachinger & Weber (1997) highlighted the importance of perceived risk in modelling risk judgments.

The perception of uncertainty created by a decision-maker varies in accuracy depending upon the situation and the information used as the basis for the PU. The provision of objective UI should improve PU. In particular, UI should improve PU when information cues that might contribute to inaccurate PU are limited.

Propensity to risk and its influence on UI utility

Propensity to risk can influence judgment through behaviours such as risk seeking and risk aversion, loss aversion, ambiguity aversion, pre-conceived (intuitive) expectations of risk levels, or can manifest itself as differences

between contrasting professions or cultures (for example, the difference between entrepreneurs and bankers: Sarasvathy, Simon & Lave, 1998). Some of these behaviours are individual differences (related to personality) while others are more related to the environment or the overall population (human nature), but regardless of their source they all have the potential to influence UI utility.

People's risk-aversion to potentially ruinous losses (Crum, Laughhunn & Payne, 1981; Kunreuther et al., 1995; Laughhunn, Payne & Crum, 1980; MacCrimmon & Wehrung, 1984, 1986; Mao, 1970) and Prospect Theory's general prediction of risk aversion for gains and risk seeking for losses (Kahneman & Tversky, 1979) have already been discussed, and now some more specific observations will be highlighted.

Kahneman & Lovallo (1993) explained that much of the observed risk aversion might be due to people treating decisions as unique and in isolation, neglecting the statistics describing the past, and therefore being unable to mitigate risk in current plans. The provision of UI should make the statistics describing the past more salient, and therefore help to reduce risk-averse behaviour when evaluating alternatives and at the same time should dampen overly enthusiastic forecasts. Contrastingly, the same research by Kahneman & Lovallo (1993) found people to be 'bold' when making forecasts, for many of the same reasons. In follow-up research Lovallo & Kahneman (2003) showed that cognitive bias and organisational pressures, rather than any rational assessment of risk, lead managers to make overly optimistic forecasts. The authors' solution to this situation was to use a different forecasting technique,

reference-class forecasting, which involves identifying similar projects, creating a distribution of outcomes for this reference class and positioning the current project on that distribution. This technique forces managers to associate a measure of uncertainty with each potential project.

Sullivan & Kida (1995) found corporate managers participating in a corporate investment decision task considered both current and target performance levels as significant determinants of propensity to risk taking or risk avoiding. Managers had multiple reference points when evaluating investment alternatives, and the relative positions of alternatives with respect to the reference points may have been considered when determining risk taking or risk avoiding behaviour. Accordingly, UI utility may also be dependent upon multiple reference points (or target levels) relating to both current and future performance in business decisions.

Ambiguity aversion and UI utility

In addition to displaying aversions towards risks and losses, people also demonstrate aversions to ambiguity. While uncertainty results from not knowing what the outcome of a decision will be conditional on some model (for example, the toss of a fair coin), ambiguity results from having limited understanding of the procedure that generates outcomes (for example pulling a red or green ball from an urn with unknown numbers of red and green balls). Ambiguity is second-order uncertainty: uncertainty about uncertainty³⁹. Ellsberg (1961) first

³⁹ Some authors (for example, see Tversky & Wakker, 1995) use an alternate terminology, with 'risk' referring to probabilities that are assumed known, and 'uncertainty' referring to probabilities associated with outcomes not given in advance.

showed that people prefer lotteries with precise probabilities rather than ones with ambiguous probabilities. This behaviour has been observed in the field, for example by underwriters pricing insurance (Kunreuther et al. 1995).

Ambiguous probabilities are generally operationalised by either providing no indication of the probability of a particular outcome, or by providing a range of probabilities (for example, there is a 20% - 40% chance). Einhorn & Hogarth (1985) proposed a descriptive model of how people make judgments under ambiguity, explaining that judgments under ambiguity are based first on an initial estimate (anchor) and then adjustments are made away from this estimate based on the amount of ambiguity present, and the individual's attitude towards ambiguity.

Highhouse (1994) found some support for this anchoring and adjustment behaviour, where decision-makers underweight imagined probabilities above the initial estimate anchor. Ambiguity avoiders were more likely to consider pessimistic probabilities while ambiguity seekers were more likely to consider optimistic probabilities. However, contrary to what has already been mentioned Highhouse (1994) found interval probabilities led to ambiguity seeking and precise, point probabilities led to ambiguity avoidance. The avoidance of ambiguous point probabilities (for example, "estimated at a 50% chance") compared to situations of known uncertainty (for example, "known to be a 50% chance") is not surprising. However, the preference for an ambiguous range of probabilities (for example, "estimated to be successful between 25% and 75%") over known uncertainty is a more contentious result. It may be that these results were in fact merely reflecting the general prediction of Prospect Theory

(Kahneman & Tversky, 1979), whereby individuals are risk seeking for losses, because the task involved selecting between various treatments for a patient. The ambiguous probability interval may therefore have just coincided with a riskier alternative in a loss-related setting. Ho, Keller & Keltyka (2002) confirmed this behaviour, noting that participants in an experiment making decisions under probabilistic ambiguity were ambiguity prone in the loss condition and ambiguity averse in the gain condition.

Although this discussion first reviewed people's attitudes towards risk (or uncertainty) and then separately reviewed people's attitudes towards ambiguity (second-order uncertainty) there is some evidence that people behave similarly towards both. Tversky & Wakker (1995) found that the characteristics of a weighting function for describing observed risk attitudes tended to hold for attitudes towards ambiguity also. In addition, it was found that people were less sensitive to changes in ambiguity than to changes in risk. However, the same study also weakened general claims of ambiguity aversion. Tversky & Wakker (1995) found people often preferred to bet on unknown probabilities. People preferred precise probabilities when made to feel ignorant or incompetent. Otherwise, people often preferred betting on unknown sources (such as sport or weather). The reasons for this behaviour are yet to be understood, but they clearly indicate a need for improved understanding of behaviour under ambiguity.

Behaviour towards ambiguity modifies choices when there are more certain alternatives available. People become ambiguity seeking or ambiguity avoiding, depending upon the conditions. Consider the situation when a

decision-maker is confronted with no such alternatives, but rather is confronted with an investment decision based on various information cues including forecasts (precise) and UI (either in an ambiguous format such as a range or a precise format such as a confidence level). Observed attitudes towards ambiguity (for example, Ho, Keller & Keltyka, 2002) lead to the conclusion that in such a situation the form of the UI may influence decision behaviour. In a loss situation a confidence interval may draw the attention of decision-makers away from a forecast (an ambiguity reducing information cue), while in a gain situation, a confidence interval may lead to ambiguity avoidance and decision-makers may be hesitant to trust the UI. If these behaviours are correct, then in gain situations (for example, investment decisions) UI should be provided as a precise, point estimate (for example, a confidence level).

Intuitive expectations of risk and the impact on UI utility

So far attitudes towards risk and ambiguity have been discussed, but people also appear to have preconceived ideas about risk, which can influence their decisions. Preconceived ideas about risk are based on intuition rather than mathematical probability. Hacking (1975) described this alternative view of probability as epistemic, while Cohen (1977) simply called it 'intuitive probability'. Behaviours have already been observed that indicate the presence of preconceived ideas about risk.

Risky and uncertain decisions are made in relation to some reference point (for example, Kahneman & Tversky, 1979; Tversky & Kahneman, 1981, 1992). Einhorn & Hogarth's (1985) ambiguity model describes decision-makers faced with an ambiguous problem anchoring on an initial probability assessment

away from which adjustments are made. People apply knowledge acquired through everyday life to all types of decision tasks (Wason, 1968; Oaksford & Chater, 2001). These are all examples of people bringing and using their intuition in risky and uncertain decision settings.

Kuhn & Budescu (1996) investigated intuition about probabilities and outcomes in hazard risk decisions by manipulating the vagueness of the probability (ambiguity) and outcome (uncertainty) dimensions of the task. The task was operationalised using three levels of probability (low, medium and high) and three levels of outcomes (low, medium and high). A fourth treatment level of probability and outcome was also included, which was unknown (or vague). Participants were asked to provide risk ratings under the various levels of probability or outcome. There were two results of note. Firstly, the vague probabilities were rated the same as the medium levels of probability. Secondly, the vague outcomes were rated the same as the medium levels of outcomes. Participants had a preconceived idea about the underlying risk, which coincided with the medium levels of risk.

If people possess an intuitive sense of underlying uncertainty, then UI may only be useful when underlying uncertainty differs from intuitive uncertainty, such as in the low and high conditions in Kuhn & Budescu's (1996) experiment. Oliver (1972) found that the provision of confidence intervals in a loan evaluation task had no impact upon decision performance and speculated that the reason for this was because the underlying risk was at an expected level and therefore the confidence intervals provided no additional information.

In summary, preconceived ideas about, and attitudes towards, risk or uncertainty, and ambiguity, have the potential to influence UI utility in a number of ways. UI may be most useful in risk-averse situations by making statistics of the past more salient. Providing multiple reference points relating to current and future performance may optimise UI utility. The format of UI should be carefully selected in order to minimise ambiguity avoidance. Finally, UI utility appears dependent upon the interaction of underlying uncertainty and the decision-maker's intuitive expectation of risk.

2.2.4.2 Confidence/Calibration

Key Messages:

- People appear to exhibit consistent overconfidence.
- This result has been challenged over the past decade, with much research arguing that the overconfidence results were due to methodological and analytical effects: the debate continues.
- More recently, the calibration research has focussed on contingent explanations for the calibration of judgmental probability assessments.
- Judges may appear overconfident because they are trying to be more informative, by providing tighter interval judgments. Accordingly, it would seem equally important to investigate the utility of probability judgments rather than the production of those judgments.
- While calibration may impact UI utility, it appears more likely that the presence of UI will influence calibration.

In Figure 2-6 (p. 90) it was hypothesised that decision-makers perceive uncertainty from available data before incorporating objective UI in order to get a revised estimate of uncertainty. How good are these initial perceptions of uncertainty? This can be ascertained by considering the calibration of decision-makers.

Calibration refers to the accuracy of people's subjective probability assessments. People are well calibrated for a given set of judgments when there is agreement between their degree of confidence (generally measured as a percentage) and the actual proportion of correct answers given. Subjective probabilities play a key role in many forms of judgments, and accordingly the calibration of these judgments is of great importance (see Beach, 1975, for a discussion on the potential benefits of subjective probabilities). As a result, a large amount of research has been conducted on calibration over the past two decades. The main result of this research has been the conclusion that people exhibit consistent overconfidence. However, this result has been challenged over the past decade, with much research arguing that the overconfidence results were due to methodological and analytical effects. More recently, the calibration research has focussed on more contingent explanations for the calibration of judgmental probability assessments.

Overconfidence

The most prominent finding to come from research into the calibration of judgmental probability assessments has been the discovery of systematic overconfidence. Lichtenstein, Fischhoff & Phillips (1982), Yates (1990) and Keren (1991) produced some of the earliest and most convincing work on

overconfidence. Their results did not seem surprising given the conclusions of Pitz (1974), that due to limited information processing capacity people are poor at assessing the probable error of their judgments, even though the judgments themselves may be relatively accurate.⁴⁰

A large number of studies have confirmed this result, in various settings by numerous authors. For example, overconfidence has been observed in predicting the likelihood of desirable, personal events (Wright & Ayton, 1989), using graphical information systems (Sen & Boe, 1991), in currency forecasting (Pollock & Wilkie, 1992), forecasting of stock prices (Önkal & Muradoğlu, 1996), audit judgments (Simnett, 1996) which are normally found to be underconfident, and in the estimation of effort required for software writing tasks (Connolly & Dean, 1997). The overconfidence phenomenon is not restricted to judgment. Gordon & Kammen (1996) even found evidence of overconfidence in statistical confidence intervals produced from time series forecasts of the Standard & Poor's 500 Stock Index.

Arguments against overconfidence

Despite the prevalence of research supporting overconfidence, this finding has been strongly and persistently challenged. Calibration research has been criticised in a number of ways, with four problems most apparent (Keren, 1997, provides an overview of overconfidence criticisms).

⁴⁰ The other major empirical result from the calibration literature is the identification of the hard-easy effect (Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff & Phillips, 1977), which describes the systematic movement from over to under-confidence as task difficulty decreases. However, it has been concluded that this effect is largely an artefact of experimental method (Gigerenzer et al, 1991; Juslin, 1993, 1994) and recent results have confirmed this (for example, see Klayman et al., 1999). As a result, this effect will not be discussed here.

First, calibration studies generally provide information in a probabilistic format, and assume elicited probabilities are subjective (a single judgment relating only to the current task), yet evaluate confidence frequentistically. It is argued that probabilities do not apply to the often-used unique, one-off questions asked in calibration studies (Gigerenzer, 1991; Gigerenzer et al., 1991). Gigerenzer (1994) found that calibration concerns disappeared when information was presented frequentistically, and argued that this was the natural way in which people process probabilities, rather than by single event probabilities.

The possible existence of a trade-off between calibration and resolution⁴¹ may explain some empirical results indicating overconfidence, and is the second major criticism of the overconfidence claim. Yates (1982) showed that calibration (absolute accuracy of confidence in judgment) and resolution (relative ability to distinguish situations which modify confidence in judgment) are not independent, Yates et al. (1996) found that recipients of forecasts prefer discrimination⁴² (effectively the same as resolution) to calibration, while Keren (1991) further claimed that the maximisation of calibration and resolution could not happen simultaneously. Similar conclusions have been reached in research investigating interval forecasts (Yaniv & Foster, 1995; Yaniv & Foster, 1997).

⁴¹ Resolution is a measure of a probability assessor's ability to assign events into subcategories for which the proportion of events correct is different from the overall proportion of events correct. A higher resolution score reflects the ability of an assessor to discriminate between differing degrees of uncertainty in their predictions using the probability scale. While calibration is an absolute measure, resolution is a relative measure.

⁴² Good discrimination by forecasters equates to an ability to produce forecasts for a target event that is ultimately going to happen that are significantly different to forecasts for an event that does not occur.

Yaniv & Foster (1995, 1997) consider the accuracy-informativeness trade-off apparent in the communication of interval forecasts, which may explain the reason for any calibration-resolution trade-off. Judges may appear overconfident because they are trying to be more informative, by providing tighter interval judgments. O'Connor, Remus & Griggs (2001) found that judgmental confidence intervals in time series forecasting were consistently asymmetric (the forecast was not the mid-point of the confidence interval) and, like Yaniv & Foster (1995, 1997), they believed this finding was due to a desire to be more informative.

Third, the analysis and interpretation of calibration studies may have confused apparent overconfidence with an experimental artefact caused by regression toward the mean by random error and truncation to fit a probability scale. For example, Erev, Wallsten & Budescu (1994) and Pfeifer (1994) showed that overconfidence may result from judgment errors creating a regression, and therefore a change in analysis technique can make overconfidence become underconfidence. Others have found supporting results (for example, Dawes & Mulford, 1996). Similarly, Ayton & McClelland (1997) reviewed the studies of Budescu, Erev & Wallsten (1997) and Juslin, Olsson & Björkman (1997) and concluded: "with the assumption of perfect calibration and with truncation at the end points of the probability scale, virtually any empirical calibration curve (half-range or full-range) ... can be mimicked ... (p. 281)". Overconfidence may be a methodological illusion.

The fourth and final criticism of calibration research relates to the typical general knowledge questions used in calibration studies, with overconfidence

being a phenomenon caused by such artificial tasks. Several studies have shown people to make well calibrated judgments when operating in a familiar environment. For example, well calibrated judgments have been recorded by weather forecasters (Murphy & Winkler, 1984), experienced tournament bridge players (Keren, 1987) and bookmakers (Phillips, 1987). These are all circumstances where judgments are precise and numerical, with prompt, unambiguous feedback (Ayton & Pascoe, 1995). In another example, Melone et al. (1995) found no evidence of overconfidence in a realistic production and workforce decision-making experiment where participants used a computer-based decision support system. In contrast, the artificiality of the experimental tasks used in calibration research combined with inappropriate sampling techniques contributes to the inappropriate claims of overconfidence (Gigerenzer et al., 1991; Juslin, 1994; McClelland & Bolger, 1994).

The outcome of such repeated criticism is that there are a large number of researchers who feel that it is inappropriate to accept overconfidence as an established fact of psychology (for example, see Dawes & Mulford, 1996). In fact some studies have clearly shown underconfidence to occur under some circumstances.

As previously mentioned, Erev, Wallsten & Budescu (1994) and Pfeifer (1994) showed that overconfidence may result from judgment errors creating a regression, and therefore a change in analysis technique can make overconfidence become underconfidence. Bolger & Harvey (1995) found people to be underconfident (hypoprecise) when judging the probability that the next point in a time-series was below a given value. The result was more significant

with trended series, and in a second experiment that involved judging the probability that the next point in a time-series was above a given value.

In the face of such criticism, researchers claiming fundamental, human overconfidence have launched counter-claims in an attempt to justify their findings: “there is overwhelming evidence to suggest that genuine overconfidence ... is a common phenomenon (Keren, 1997, p. 273)”. Brenner et al. (1996) compared probability judgments for single events with judgments of relative frequency and found both groups to exhibit significant overconfidence. Budescu, Erev & Wallsten (1997) allowed for the effects of random error, and used a representative sample of items in order to eliminate false overconfidence, but still found overconfidence in 87% of participants in their study. Klayman et al. (1999) also separated out systematic effects from statistical artefacts and found an overall bias towards overconfidence. See Keren (1997) for further details on why overconfidence is real and not an artefact.

Factors influencing calibration

Calibration is clearly a contingent measure. Whether overconfidence is a prevalent human condition or not, its manifestation or absence appears to be dependent upon numerous factors (See O'Connor, 1989, for a description of calibration as a contingent model of behaviour and Klayman et al, 1999, for a more recent contribution). For example, overconfidence is higher among people with greater numerical skills and when using graphical displays (Sen & Boe, 1991), when desirability for an event to occur is high (Wright & Ayton, 1989), among those living in poorer socio-economic conditions (Whitcomb et al.,

1995), when giving verbal rather than numerical probability judgments (Wallsten, Budescu & Zwick, 1993), during initial self-efficacy judgments (Stone, 1994), with the provision of positive payoffs (Wallsten, Budescu & Zwick, 1993; Henry & Sniezek, 1993), when there is greater perceived controllability (Wright & Ayton, 1989; Henry & Sniezek, 1993), as the amount of available information increases (Peterson & Pitz, 1988), with practice and effort and in the absence of feedback (Paese & Sniezek, 1991), when there is unfamiliarity with the task or topic of interest (O'Connor, 1989), when producing judgmental confidence intervals around time series forecasts based on series with no trend (Lawrence & Makridakis, 1989), when producing prediction intervals using large scales or when data is more variable (Lawrence & O'Connor, 1993), by asking questions that require confidence range judgments (Klayman et al., 1999) and by the desire to be more informative than accurate with interval predictions (Yaniv & Foster, 1997).

By contrast, overconfidence is reduced (and calibration improved) when feedback is provided (O'Connor, 1989; O'Connor & Lawrence, 1989; Ayton & Pascoe, 1995; Harvey, Koehler & Ayton, 1997; Bolger & Önköl-Atay, 2004), by asking two-choice questions (Klayman et al., 1999), or when information is presented in terms of relative frequency (Gigerenzer, 1994).

Other factors to influence calibration include environmental inducements, such as the threat of legal action against auditors (O'Connor, 1989). Inducements may cause miscalibration by making people more concerned with being informative rather than being accurate (Yaniv & Foster, 1997) due to an underlying asymmetric loss function.

The influence of some factors on calibration causes disagreement. For example, Trafimow & Sniezek (1994) found overconfidence to be greater amongst individuals with higher perceived expertise, whereas Wright et al. (1994) found that those who rated themselves as more expert were in fact better calibrated. Understanding the contingency of calibration is in its adolescence. In addition to the contingent factors summarised above, it may be that some individuals are simply better calibrated than others (Wright & Ayton, 1989; Klayman et al., 1999).

Given that calibration (or the precision of an interval estimate) is not necessarily a good indicator of accuracy in judgment (Schul & Yaniv, 1997), and calibration is often sacrificed for the sake of informativeness (Yaniv & Foster, 1995, 1997; O'Connor, Remus & Griggs, 2001), perhaps it is wrong to expect perfect calibration in decision-making (Pfeifer, 1994):

“To expect probability forecasters to be well calibrated is neither reasonable nor important. The only situations in which calibration can be expected are those in which the subjects have considerable experience (with feedback) in forecasting a stable situation. In such situations, the forecaster can be expected to have a clear understanding of the prior distribution – the process that generates the forecasting tasks (p. 212)”.

The potential relationship between confidence calibration and UI utility

The main result of the calibration literature, an inability to accurately predict confidence in judgment, forms a strong argument for the provision of UI. It's reasonable to expect people to be better calibrated if they are aware of the level of uncertainty in a task. As mentioned, one reason for this poor calibration may be because people seem more intent on being informative with their probability judgments rather than calibrated. As a result it is important to

investigate the utilisation of probability judgments rather than the production of those judgments. As shall be seen in section 2.3, research investigating the utilisation of UI is incomplete and inconclusive.

Our lack of calibration might influence the potential utility of UI in a number of different ways. Poor calibration may mean that UI based on statistical analysis may be trusted more, and be more useful, than UI based on expert judgment. Of course, in some situations it may only be possible to have UI based on expert judgment. General findings of overconfidence may make people less likely to use UI, but also makes it far more valuable for them to do so. Individual differences in calibration may lead to differences in the ability to utilise UI. It remains to be seen whether a trade-off similar to the accuracy – information trade-off for the production of confidence judgments exists for UI utilisation.

Although the reasons given above mean the calibration of a decision-maker could influence UI utility, calibration is an output of a decision-task, and it is more likely that the presence of UI will influence (presumably improve) judgmental calibration (see Figure 2-7). The presence of UI should provide decision-makers with knowledge of underlying uncertainty, which in turn should make them better able to understand how good their judgements are, thus improving calibration. Figure 2-7 illustrates this by showing that it is the combination of a decision-maker's innate sense of calibration, and their ability to obtain utility from UI, which should influence task specific calibration.

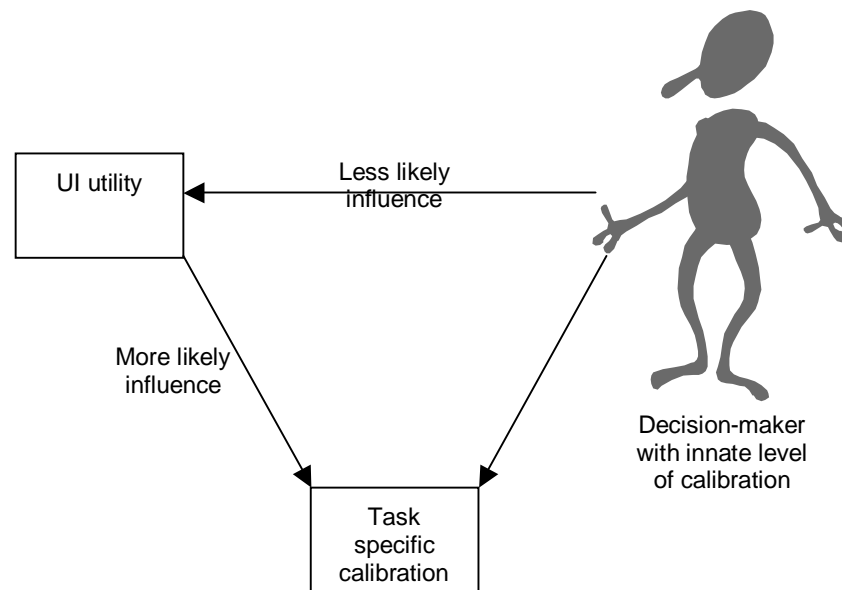


Figure 2-7: The relationship between UI and calibration.

2.2.5 A Model of Contingent Effects for Decision-Making under Uncertainty with UI

Key Messages:

- This section outlines a contingency model for the utilisation of UI, based on factors relating to person, task, and environment, interactions of person and task, and fundamental risky behaviours.
- The model highlights the importance of perceived uncertainty and the extent to which decision-makers believe in objective UI.
- The most influential and controllable areas of the model appear to relate to task, interactions between task and person, and fundamental risky behaviours.

Having reviewed the literature for factors with the potential to influence the utility of UI, it is now possible to specify a model of contingent effects describing the contingent utility of UI. This section outlines such a contingency model, continuing the use of a person, task, and environment framework while incorporating fundamental risky behaviour.

High-level factors most influential in determining UI utility are summarised in Figure 2-8. Figure 2-8 depicts a model showing 'UI Utility & Utilisation' to be a function of an individual's perceived uncertainty (PU) in a situation, objective UI and an individual's belief in that objective UI. The model further divides factors affecting 'UI Utility & Utilisation' between those influencing PU and those influencing 'Belief in UI'. Environmental factors and fundamental risky behaviours are portrayed impacting 'Belief in UI', and also UI utility. Task and personal factors, as well as interactions of task and person, are shown directly influencing PU. UI, while also a task related factor, is highlighted separately due to its significance in this model. Of central importance to the model is the 'UI Utility & Utilisation'. This is a qualitative measure of how useful and how used UI is at the point of making a decision. This part of the model could also be replaced by another concept, revised PU, since it is at this point in the decision that individuals would attempt to incorporate UI with their own PU. However, this research is not concerned with how PU is revised, but under what conditions objective UI has utility and is utilised.

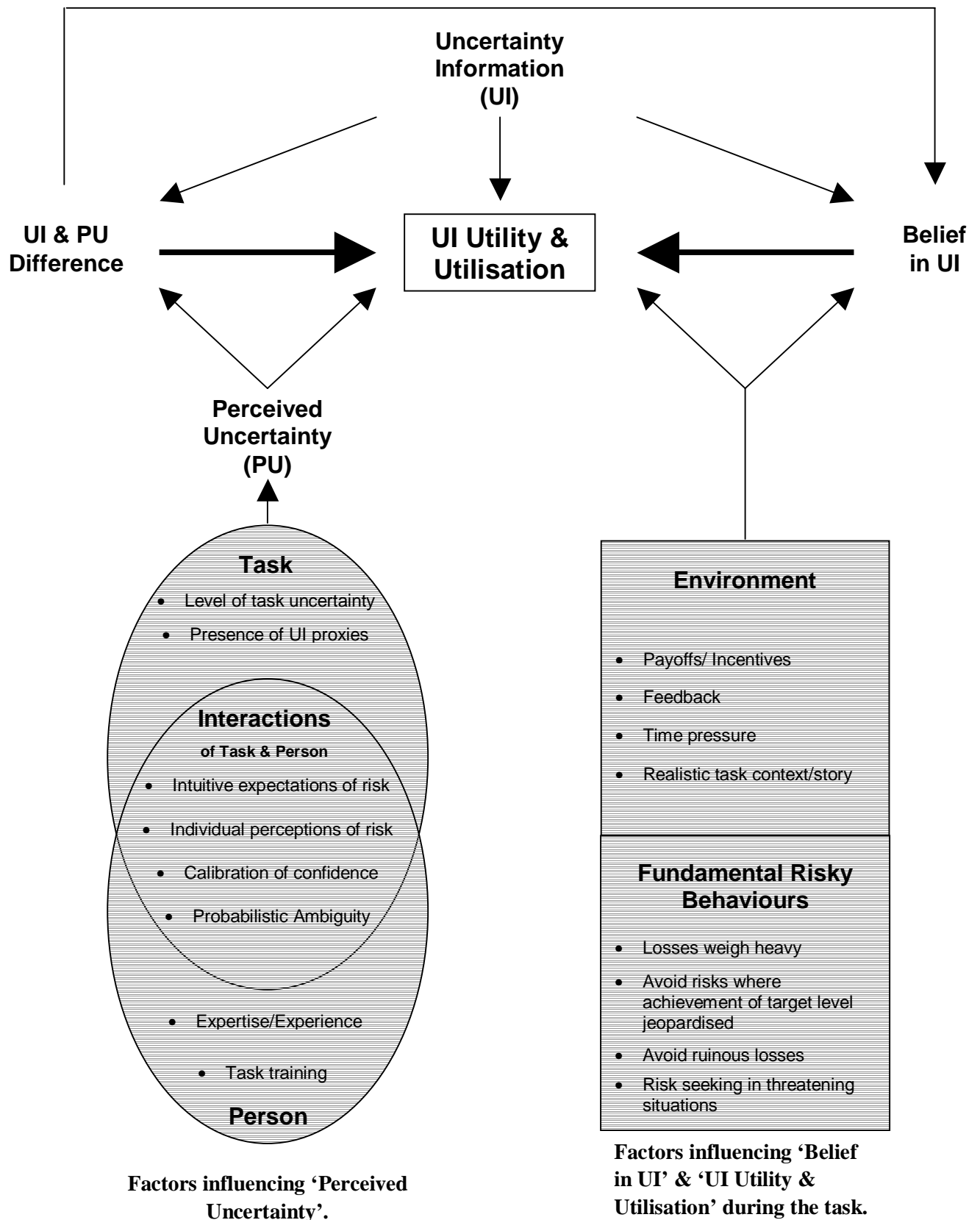


Figure 2-8: A model of contingent effects describing the contingent utility of UI

The details of this model will now be briefly described, and it will become apparent that the most influential areas relate to the task, interactions between task and person, and fundamental risky behaviours. The other areas of influence, while not as significant as the task or interactions, are still important.

Factors relating to task, person and interactions of task and person impact the decision-maker's perceptions of uncertainty. This in turn influences the decision-maker's belief in UI, particularly when there are differences between PU and UI, and ultimately has an impact on the utilisation of UI.

Task characteristics probably have the greatest ability to influence UI utility through modifying PU, but fortunately are also the easiest factors to control. UI utilisation can be improved by modifying the format of UI: keeping it as simple as possible and perhaps by presenting the information in point form (a level) rather than as a range. This prevents users from using the midpoint of a range as a forecast rather than using the information that a range conveys. The presence of UI proxies may lead to accurate UI cues (like a confidence interval) being ignored. For example, it may be easier or more intuitive for some individuals to use history information (especially when presented graphically) in a judgmental forecasting task to provide an estimate of uncertainty rather than to rely upon a specific UI cue generated by unknown statistical means or by an expert. Finally, task uncertainty may also impact UI utilisation, although this may be due to a mismatch between perceived and actual task uncertainty.

Factors relating to the decision-maker (person) are much less significant in determining UI utilisation. However there are some noteworthy considerations

that are appropriate for the successful analysis of any decision-making task. Ensuring that adequate training in UI and all information cues takes place should increase the probability that UI is understood and utilised. Selection of a task that is unlikely to have been completed previously will reduce the likelihood of task knowledge, and force participants onto a 'level playing field' where differences in decision performance will be due to nascent experience based on available decision-aids and information cues, such as UI.

The task and person components of any decision interact with each other and cause certain behaviours by the decision-maker. These behaviours have the potential to modify decisions and decision performance. They also have great potential to influence the utilisation of UI. Individuals have certain expectations about the level of risk, or uncertainty, in any given situation due to past experience, biases or both. When expectations of risk do not match underlying uncertainty, decision performance suffers. However, these may be the exact situations where UI is most valuable. Perceptions of risk are influenced by the interactions of task and the individual (for example, by manipulating the scale of presentation used to graph historical data). Perceptions of risk determine propensity towards certain risky behaviour, and have the potential to influence UI utility just like risk expectations. The literature on the calibration of confidence in judgment has found that there are individual differences in the way people are calibrated: some are well calibrated while some are not. In addition, calibration itself is contingent upon other task and personal factors. The extent to which a situation evokes calibration, or the extent to which an individual decision-maker is well calibrated, may have an effect on whether or not UI is utilised effectively. The last interaction factor

relates to probabilistic ambiguity. The commonly observed behaviour of ambiguity avoidance predicts that any form of probabilistic ambiguity within UI will lead to the UI being avoided.⁴³

Environmental factors and fundamental risky behaviours limit or increase the extent to which objective UI is believed. These factors also directly influence the utility and utilisation of UI during the task.

Similar to those factors relating to person, environmental factors that may influence UI utilisation are also common to any contingent model of decision-making, but again are not as significant as task, interaction or risky behaviour factors. The use of incentives in psychological and economic experiments is controversial, but it appears that for probability judgment tasks incentives are beneficial, and it can therefore be assumed that they can improve the likelihood that UI will be utilised effectively. Although feedback is beneficial in calibration there appears to be evidence that feedback is of little use in probability learning tasks because of the complexity of the feedback loop. As a result feedback is of doubtful use, and may distract individuals away from UI by increasing the complexity of the task. It should also be noted that in reality feedback is often unavailable anyway. Too much time pressure may reduce utilisation of UI, particularly when there are alternative, more familiar information cues available for determining uncertainty. The final environmental factor that has the potential to influence UI utilisation relates to task setting, and in particular to the context

⁴³ Of course, UI is generally ambiguous information, but the ambiguity comes from outcomes rather than from probabilities (See Ho et al., 2002, for an explanation of the different types of ambiguity).

of the task. Tasks that are as realistic and meaningful as possible are required for individuals to be able to make sense of probability information.

The foundations of this model of contingent effects are based upon four fundamental risky behaviours, each having the potential to influence or perhaps override the utilisation of UI. First, people weigh the probability and magnitude of losses more than gains. Second, decision-makers avoid ruinous losses. Third, decision-makers avoid situations where a target cannot be achieved. Fourth and finally, risks are more likely taken in threatening situations. If individuals are placed in situations that produce any of these behaviours, there is potential for UI utilisation to be altered. For example, UI may help decision-makers determine when a target can or cannot be achieved.

It is important to note the distinction that this model makes between PU and objective UI, a distinction first introduced in Figure 2-6 (p. 90). PU is the subjective determination of uncertainty by decision-makers using available information, while being influenced by their own biases, beliefs and cognitive limitations. The model acknowledges that the information conveyed by objective UI, might be drowned out by other, less accurate, factors. It is proposed that any difference between PU and UI has both a direct and indirect impact upon UI utility. When there is a difference UI provides utility, however the difference can also detrimentally reduce trust, or belief, in the UI and reduce UI utilisation. When PU and UI are in agreement UI utility is negligible, since it no longer provides additional information, although ironically this probably increases belief in the UI. Ideally, belief in UI should come from characteristics of the UI, such as how it was produced, rather than whether or not there is a difference between

UI and PU. That way the utility and utilisation of UI can be maximised when there is most potential benefit: when UI and PU differ.

The model of contingent effects describing the contingent utility of UI (Figure 2-8) succinctly identifies those factors most likely to influence the utilisation of UI. This model will now be used to evaluate previous studies that have investigated the utility of UI with mixed conclusions. By evaluating these previous UI studies it will be possible to determine the most appropriate way to explore UI utility, and provide a means of putting these mixed results into context.

2.3 A SUMMARY OF EMPIRICAL EVIDENCE REGARDING THE USE OF UNCERTAINTY INFORMATION – A REFINED MODEL OF CONTINGENT EFFECTS

Key Messages:

- While much effort has advanced our understanding of the production of judgmental UI, the same cannot be said for the downstream, utilisation of UI.
- The little work (eleven studies in forty years) that has been put into understanding whether UI is beneficial to decision-makers has produced inconclusive results.
- Only three out of eleven studies have found an overall effect due to the provision of UI in decision-making, and only two of those studies went on to show that there was significant benefit in providing UI to decision-makers.

- Eight out of the eleven studies received at least some result (however minor) indicating that UI has a very contingent role in decision-making.
- Most of this research has been spawned by efforts to reform traditional accounting practices, but no reform has occurred.
- Literature review highlighted two factors that were not part of the preliminary model, but may still have significant influence on UI utilisation: UI must have diagnostic or decision-making value for any given task; and, for UI to have utility a task's loss function needs to be asymmetric, and the direction of the asymmetry is also important.

Significant effort has been put into understanding the production of judgmental UI and the calibration of confidence in judgment. This effort has at times created divergent understandings of calibration, but advances have been made and our knowledge of calibration has begun to converge. While such a large amount of effort has advanced our understanding of the production of judgmental UI, the same cannot be said for the downstream, utilisation of UI⁴⁴. This lack of effort put into understanding the utilisation of UI is surprising given that the calibration literature tells us people are poor at judging uncertainty, and modify their judgments of confidence to be informative rather than accurate. That is, people make judgments about their confidence based on how that information is to be received and utilised by another decision-maker. As a result the how and why of utilising UI (such as judgments about confidence in decisions) needs to be better understood.

⁴⁴ This holds true regardless of whether UI is judgmentally produced by experts or statistically based.

The little work that has been put into understanding whether UI is beneficial to decision-makers has produced mixed results. This incomplete and inconclusive area of research will now be analysed using the model of contingent effects describing the contingent utility of UI developed in section 2.2.5. Afterwards this preliminary model will be augmented into a refined model describing the contingent utility of UI.

Table 2-1 summarises the main studies to have investigated the utility of UI in decision-making. The table helps to highlight several points of interest. First, there are only eleven studies that have investigated UI utility. Second, half of these studies were based on dissertations⁴⁵ indicating that there has been little follow-up research in this field and that the field is immature. Third, only three out of eleven studies have found an overall effect due to the provision of UI in decision-making, and only two of those studies went on to show that there was significant benefit in providing UI to decision-makers. Fourth, finally and contrastingly, eight out of the eleven studies received some result indicating that UI has utility in decision-making. Points one and two show the need for more research in this area, while points three and four clearly indicate that whatever impact UI will have on decision-making, it will most likely be highly contingent.

Most of this research has been spawned by efforts to reform traditional accounting practices. Studies by Oliver (1972), Hawkins (1974), Keys (1978) and Chen & Summers (1981) were all investigating whether decisions based on financial reports would change if the reports contained probabilistic rather than

⁴⁵ Chen & Summers (1981), Foong (1992), Hawkins (1974), Oliver (1972) and the current study are all either unpublished dissertations or are publications arising from dissertation research.

traditional, deterministic information. Birnberg & Slevin (1976), Johnson (1982) and Eckel (1983) did not use financial tasks within their studies, but used abstract task settings. From an abstract setting these studies made an inference about the utility of UI in a financial setting, hoping to influence accounting practices. No such influence has occurred.

Hirst, Koonce & Miller (1999) represents another group of papers investigating the utility of UI in a financial setting, although this is specifically targeted at the communication of management earnings forecasts to investors and analysts. Of most interest to this collection of research is the impact of forecast precision (for example, see also Baginski, Conrad & Hassell, 1993; Pownall, Wasley & Waymire, 1993; Kennedy, Mitchell & Sefcik, 1998). Hirst, Koonce & Miller (1999) investigated the impact of forecast precision in combination with prior forecast accuracy. This combination is effectively the same as providing UI, except without the explicit identification of uncertainty.

Ho, Keller & Keltyka (2002) contribute to this area through their research into ambiguity. They investigate whether the commonly observed behaviour of ambiguity aversion (Ellsberg, 1961) holds for both loss and gain conditions, when ambiguity applies to outcomes only. They achieve this by comparing range and point forecasts against a target level, and therefore their results have direct consequences for UI utilisation.

Herman, Ornstein & Bahrack (1964) provide the most unusual contribution. They investigated different probabilistic and deterministic presentation formats for displaying object location. Despite its uniqueness it is one of the strongest contributors to this field.

Foong (1992, 2003) took a more holistic approach to investigating UI utility by commencing the development of a contingent model of UI utility. While all the other studies were concerned with whether UI altered decision behaviour or improved decision performance, Foong (1992, 2003) delved into the reasons why past research had obtained mixed, highly contingent results. She attributed this primarily to the lack of financial incentives available to the participants of earlier studies.

All of these studies have been summarised, and aspects relating to each part of the model of contingent UI utility (see section 2.2.5) are presented in Table 2-1 to Table 2-6. Figure 2-9 describes the key used in tables 2-1 to 2-6.

The rest of this section includes brief descriptions of each study, some constructive comments outlining their strengths and weaknesses, and concluding comments on what has been learnt about the utility of UI.

○	Very Low
◐	Low
◑	Medium
◒	High
●	Very High
Y	Yes
N	No
Blank	Not applicable or unknown

Figure 2-9: The Key for Tables 2-1 to 2-6

Paper	Task	Overall results		Contingent results	Results - Effect of UI
Birnberg & Slevin	(1976) General questions - set target and decision rule	N	N		Minor effect moving from low to high variance, but not due to UI.
Chen & Summers	(1981) Investment decision & cost determination	N	Y		Significant results only when directly comparing between different treatments. Confidence in decisions was also affected.
Eckel	(1983) Business game: maximising profits	Y	Y		Probabilistic information improved decision performance. Results impacted by a learning effect, and by the source of UI.
Foong	(1992) Inventory forecasting	N	Y		UI beneficial with tabular display and asymmetric loss function.
	(2003) Forecast & investment decision	N	Y		UI beneficial with incentives.
Hawkins	(1974) Investment allocation decision	Y	N		The provision of UI significantly altered decision behaviour.
Herman, Ornstein & Bahrick	(1964) Search & attack game	Y	N		Discrete probabilistic displays improved object location judgments, and also improved calibration, beyond those with probability density function or point estimate display.
Hirst, Koonce & Miller	(1999) Earnings predictions	N	Y		The interaction of forecast precision and prior forecast accuracy influenced confidence, but not judgments.
Ho, Keller & Keltyka	(2002) Ambiguous outcome investment	N	Y		Behaviour was ambiguity prone (used UI) in loss condition, and ambiguity averse (used point estimate) in gain condition.
Johnson	(1982) Nonfinancial, predecision probability judgment	N	Y		UI beneficial at high task uncertainty, but worse at low and medium levels of task uncertainty.
Keys	(1978) Loan decision	N	N		UI was of no benefit.
Oliver	(1972) Loan decision	N	N		Those with UI exhibited a slight reluctance to lend, however this was not statistically significant.

Table 2-1: Results of empirical investigations into the utility of uncertainty information.

Paper	Task	UI Proxies	Uncertainty Levels	UI Format: CI	UI Format: Other	Error Distribution	Forecast Provided	Graphical/Tabular	History	# of trials	Optimal Decision	Data Credibility	Task Complexity	# of Participants	Overall results	Contingent results	Comment
Birnberg & Slevin	(1976) General questions - set target and decision rule	H	2	95	S	Y	T	5	N	☐	●	100	N	N	N	N	Asymmetric distributions were tested in a pilot study: no difference.
Chen & Summers	(1981) Investment decision & cost determination		V	R			T	N	8	N	☐	●	70	N	Y	Y	Perceived data credibility analysed. Point, Range, and CI compared.
Eckel	(1983) Business game: maximising profits				DD	A	N	T	N	6	Y	☐	●	109	Y	Y	UI provided to and produced by subjects.
Foong	(1992) Inventory forecasting	H		50	S	Y	T, G	20	36	Y	☐	●	60	N	Y	Y	Investigating display mode and loss function.
	(2003) Forecast & investment decision	H		50	S	Y	T, G	18	20	Y	☐	●	80	N	Y	Y	Investigating impact of display mode and incentives.
Hawkins	(1974) Investment allocation decision	H		90	S	N	T	2	1	N	☐	●	72	Y	N	N	Cues included 2 years of financial reports and industry ratios.
Herman, Ornstein & Bahrick	(1964) Search & attack game			50, 90	DF	S	Y	G	N	40	Y	☐	●	18	Y	N	Comparing point estimate, two confidence contours, probability distribution.
Hirst, Koonce & Miller	(1999) Earnings predictions	H			R	S	Y	T	3	1	Y	☐	●	121	N	Y	Investigation of forecast precision and prior accuracy.
Ho, Keller & Keltyka	(2002) Ambiguous outcome investment				R		Y	T	N	1	N	☐	●	39	N	Y	Comparing ranges and forecasts against target levels.
Johnson	(1982) Nonfinancial, predecision probability judgment		3	V	DD	A	N	T	N	8	Y	☐	●	65	N	Y	Intuitive probability revision for stick length. Environmental (8 levels) and informational (3 levels) uncertainty.
Keys	(1978) Loan decision	H, I		V	S	Y	T	3	3	N	☐	●	38	N	N	N	Short, long-term loans. Subjective UI cue came with point estimate.
Oliver	(1972) Loan Decision	H, I		90	S	N	T	2	1	N	☐	●	120	N	N	N	Intermediate-term loan decisions. Subjective confidence interval cue.

of history points provided
 G=Graphical, T=Tabular (or verbal)
 An e.g. of redundant information
 (S)ymmetric, (A)symmetric
 Discrete Distribution, Density Function, Range
 Confidence Interval Size (%), V=Varied
 The different levels of task uncertainty
 H=History, I=Intuition
 Whether data is believable

Table 2-2: Task factors with potential to impact UI utilisation.

Paper	Task	Participants Expertise/Experience Trained in the task		Overall results		Contingent results	Comment
		P	E	N	Y		
Birnberg & Slevin	(1976) General questions - set target and decision rule	P	○	○	N	N	
Chen & Summers	(1981) Investment decision & cost determination	P	○	○	N	Y	Personal traits were used as a means for analysis.
Eckel	(1983) Business game: maximising profits	U	○	●	Y	Y	Interaction with computer greatly increased training for this task.
Foong	(1992) Inventory forecasting	P	○	●	N	Y	
	(2003) Forecast & investment decision	P	○	●	N	Y	
Hawkins	(1974) Investment allocation decision	E	●	○	Y	N	Task was completed by volunteers from a mailout.
Herman, Ornstein & Bahrick	(1964) Search & attack game	U	○	●	Y	N	Significant training took place prior to the experiment.
Hirst, Koonce & Miller	(1999) Earnings predictions	P	○	○	N	Y	
Ho, Keller & Keltyka	(2002) Ambiguous outcome investment	P	○	●	N	Y	
Johnson	(1982) Nonfinancial, predecision probability judgment	U	○	●	N	Y	For training participants had instructions, and also completed trial decisions.
Keys	(1978) Loan decision	E	●	○	N	N	Participants were given instructions, but it is unclear whether these were used.
Oliver	(1972) Loan Decision	E	●	○	N	N	

Indicating prior knowledge of task
(E)xperts, (P)ostgraduates, (U)ndergraduates

Table 2-3: Person factors with potential to impact UI utilisation.

Paper	Task	Incentives	Feedback	Loss Function	Time Pressure	Realistic Task	Overall results	Contingent results	Comment
Birnberg & Slevin	(1976) General questions - set target and decision rule	N	N		○	N	N		
Chen & Summers	(1981) Investment decision & cost determination	Y	N		○	N	Y		
Eckel	(1983) Business game: maximising profits	Y	Y	S	◐	◐	Y	Y	Motivated by performance being 15% of grade.
Foong	(1992) Inventory forecasting	N	Y	S, A	○	◐	N	Y	The loss function was modified between symmetric and asymmetric.
	(2003) Forecast & investment decision	Y, N	Y	A	○	◐	N	Y	Risk averse loss function. Incentives were investigated.
Hawkins	(1974) Investment allocation decision	N	N		○	●	Y	N	
Herman, Ornstein & Bahrick	(1964) Search & attack game	Y	Y	A	●	◐	Y	N	Considerable time pressure, each decision being made in 15 seconds.
Hirst, Koonce & Miller	(1999) Earnings predictions	Y	N	S		◐	N	Y	
Ho, Keller & Keltzka	(2002) Ambiguous outcome investment	Y	N		◐	◐	N	Y	Mildly motivated since 2% of their final grade was dependent upon their score.
Johnson	(1982) Nonfinancial, predecision probability judgment	Y	N	S	◐	○	N	Y	
Keys	(1978) Loan decision	N	N	A	○	●	N	N	Loss function was implied asymmetric, due to risk averse nature of loan decisions.
Oliver	(1972) Loan Decision	N	N	A	◐	●	N	N	Loss function was implied asymmetric, due to risk averse nature of loan decisions.

(Symmetric, (A)symmetric
Such that strategy may change
(Y)es, (N)o

Table 2-4: Environment factors with potential to impact UI utilisation.

Paper		Task	Expectation of Risk Risk Propensity Calibration Ambiguity				Overall results	Contingent results	Comment
Birnberg & Slevin	(1976)	General questions - set target and decision rule	N	N	N	N	N	N	
Chen & Summers	(1981)	Investment decision & cost determination	N	Y	N	N	N	Y	One group of participants were selected based on risk-taking tendency.
Eckel	(1983)	Business game: maximising profits	N	N	N	N	Y	Y	
Foong	(1992)	Inventory forecasting	N	N	N	N	N	Y	
	(2003)	Forecast & investment decision	N	N	N	N	N	Y	
Hawkins	(1974)	Investment allocation decision	N	N	N	N	Y	N	
Herman, Ornstein & Bahrack	(1964)	Search & attack game	N	N	Y	N	Y	N	They recorded operator estimates of the hit probability at each stage. Calibrated with confidence contours.
Hirst, Koonce & Miller	(1999)	Earnings predictions	Y	N	N	Y	N	Y	High expectations of risk due to general knowledge of company earnings. Ambiguity present because accuracy of range forecast was not explicit.
Ho, Keller & Keltyka	(2002)	Ambiguous outcome investment	N	Y	N	Y	N	Y	Ambiguous outcomes, and some probabilities. Risk propensity affected use of UI and results.
Johnson	(1982)	Nonfinancial, predecision probability judgment	Y	N	N	N	N	Y	Task involved pre and post probability judgments, and therefore identified expectations of risk.
Keys	(1978)	Loan decision	Y	N	N	N	N	N	Previous experience and expertise may have set a strong expectation of risk by the loan officers.
Oliver	(1972)	Loan Decision	Y	N	N	N	N	N	Previous experience/expertise may have set a strong expectation of risk by the loan officers.

Presence of ambiguous cues
Were the participants calibrated?
Did risk propensity affect results?
Expected and task risk mismatch?

Table 2-5: Interactions of task and person with potential to impact UI utilisation.

Paper	Task	Losses weigh heavy	Target Level	Ruinous Losses	Threatening Situation	Overall results	Contingent results	Comment
Birnberg & Slevin	(1976) General questions - set target and decision rule	N	Y	N	N	N	N	Participants created their own target level and then a decision rule.
Chen & Summers	(1981) Investment decision & cost determination	N	N	N	N	N	Y	Decisions involved calculation of various accounting figures.
Eckel	(1983) Business game: maximising profits	Y	N	Y	N	Y	Y	The goal of maximising profits increased the salience of potential losses.
Foong	(1992) Inventory forecasting	Y	N	N	N	N	Y	The risk averse inventory decision with feedback highlights possible losses.
	(2003) Forecast & investment decision	Y	N	N	N	N	Y	The risk averse investment decision with feedback highlights possible losses.
Hawkins	(1974) Investment allocation decision	N	N	N	N	Y	N	
Herman, Ornstein & Bahrack	(1964) Search & attack game	Y	Y	N	Y	Y	N	A threatening, loss possible situation which had a target (location of enemy) for comparison with the UI.
Hirst, Koonce & Miller	(1999) Earnings predictions	N	N	N	N	N	Y	
Ho, Keller & Keltzka	(2002) Ambiguous outcome investment	Y	Y	N	N	N	Y	The presence of a target level and point estimate dramatically effected UI use.
Johnson	(1982) Nonfinancial, predecision probability judgment	N	N	N	N	N	Y	
Keys	(1978) Loan decision	Y	N	N	N	N	N	Loan amount was large for the small- to medium-sized banks the subjects were from.
Oliver	(1972) Loan Decision	N	N	N	N	N	N	

Participants had to react to 'threats'
 Decisions had possible ruinous outcomes
 Task involved achieving a target level
 Task had salient probability or magnitude of loss

Table 2-6: Risky behaviours with potential to impact UI utilisation.

Research investigating the impact of UI on accounting decisions

Oliver (1972) was the first to investigate UI utility with the main aim being to help understand whether reform of traditional accounting practices would be worthwhile, however no significant results were obtained. The study required 120 professional bankers to make a one-off intermediate-term loan decision that required the allocation of a limited amount between the loan requests of two hypothetical companies. They had 30 minutes to perform the task. Half of the participants received deterministic (point estimate) financial statements for the two companies while the other half received probabilistic financial statements (90% confidence intervals which were calculated subjectively). The format of the financial statements did not alter resulting loan decisions, although there was a (non-significant) slight reluctance by those receiving the UI (probabilistic financial statements) to lend.

Keys (1978) conducted a similar loan decision experiment to Oliver (1972) with some key modifications. Participants (38 loan officers from small- to medium-sized banks) completed a mail-out survey asking them to make short- and long-term loan decisions for each of three actual companies (with company names withheld). Participants were provided with three years of financial statements for each of the three companies, with approximately half of the participants receiving probabilistic financial statements while the other half received traditional financial statements. The probabilistic financial statements contained confidence intervals of varying widths and trend that had been calculated subjectively (although subjects were told they were measures of uncertainty in the cost system) and were accompanied by a point estimate. The

amount of the loan was large enough such that small- to medium-sized banks would carefully consider the possibility of loan default. An education section explaining confidence intervals was included in the research instrument, although there was no evidence that this was used. Like Oliver (1972) the format of the financial statements in Keys (1978) had no effect on the loan decisions.

Chen & Summers (1981) continued the investigation into the potential for probabilistic accounting figures by conducting a laboratory experiment with postgraduate students that included two types of preliminary decisions (an investment/stock price decision and a cost determination), and a capital allocation decision following each of the preliminary decisions. Accounting information was divided into four treatments, being deterministic (point estimate), range, confidence interval (a pair of interval-scaled accounting figures with associated level of confidence), and triple confidence interval (three pairs of interval-scaled accounting figures with associated levels of confidence). The size of the ranges and confidence intervals varied. Personal traits were identified, as a means of analysis (including risk propensity), but did not signal differing decision behaviour.

Participants were randomly assigned to one of the two preliminary decisions (price or cost) and completed the task under all four treatments. Under each treatment, which was experienced separately and in a random order, the participant received information on a different company and was required to produce both point and range estimates. After completing all four treatments, the same four tasks with answers were returned to the participant

so that the capital allocation task could be commenced. This involved simultaneously comparing all four companies, each with a different accounting format treatment, and allocating capital accordingly. Subjects completed the same procedure for both the price and cost decisions.

The two preliminary price and cost decisions, where participants received different accounting formats separately, were not influenced by the format of the accounting information. However, in the capital allocation task, when participants were able to compare the different treatments simultaneously, a significant difference in decision behaviour was observed due to treatment effects. The authors do not elaborate on exactly how behaviour differed under the capital allocation task.

Chen & Summers (1981) also found that decisions made with probabilistic data left participants feeling more confident and more satisfied. And in particular participants were more satisfied with singular confidence interval data (one pair of interval-scaled accounting figures) than with triple confidence interval data (three pairs of interval-scaled accounting figures).

Like Chen & Summers (1981), Hawkins (1974) also conducted an experiment that included an investment allocation task. The aim of the experiment was to see whether investment decisions were altered by the format of the financial information, regardless of whether those decisions were better or worse. A survey was mailed to more than 400 randomly selected financial analysts. Subjects received 2 years of financial reports and industry ratios for two hypothetical companies. Half of the sample received conventional deterministic financial information while the other half received information in a

probabilistic format. Their task was to allocate \$10,000 to either or both of the companies. The survey generated 72 responses. Results indicated that those respondents receiving probabilistic financial statements made significantly different investment decisions to those in the control group who received conventional financial statements.

The studies by Oliver (1972), Hawkins (1974), Keys (1978) and Chen & Summers (1981) were all motivated by the desire to improve the information content of traditional accounting reports, but their results were inconclusive and little was learnt about whether or not probabilistic financial information could improve or even modify decision-making⁴⁶. However, detailed analysis of the tasks used in each of these studies formed a basis from which future research could better pinpoint differences in behaviour due to UI.

Lack of results in Oliver (1972) and Keys (1978) are most likely due to two task characteristics. First, the provision of history information, although limited, provided an alternative for UI. Second, the presence of an intuitive expectation of task uncertainty by the experienced loan officers may have rendered the control group's task equivalent to that of the experimental group. These inadvertent UI proxies may have led to the UI providing no benefit to the experimental group. A second, and more likely, reason for the UI not causing different behaviour may have been due to the nature of the task which encouraged the use of the mid-point of the confidence intervals rather than using the probabilistic information conveyed by the confidence intervals. Loan

⁴⁶ Hawkins (1974) was the only one of these four studies to find an overall effect due to UI, and even this result was not very strong and did not investigate why there was a difference in decisions made by those in the control group and those receiving UI.

decisions involve taking a number of quantitative variables and combining them with some form of risk calculation. It is highly likely that given this form of task people would look for cognitive shortcuts in order to reduce the required effort. One such shortcut would be to take the midpoints of the confidence intervals, rather than to multiplicatively combine confidence intervals for multiple variables. Because the confidence intervals in these studies were based on symmetric error distributions, the midpoints of the confidence intervals were exactly the same as the point estimates received by the control group. This cognitive shortcut therefore led the experimental and control group tasks to be equivalent. Nakano (1976) questioned Oliver's task in a similar manner, explaining that the data were not additive by nature, and the addition of probabilistic information into historical costs simply made the information more difficult to comprehend.

The nature of the tasks in both Oliver (1972) and Keys (1978) promoted simplification strategies because the UI did not provide obvious diagnostic or decision-making value in its given form (a confidence interval). The mid-point of the provided confidence interval (essentially a forecast) was of greater value to the participants, and was easy to determine, therefore this was the information that was utilised.

Results in Chen & Summers (1981) may have also been impacted by this same lack of UI value, leading to a cognitive shortcut. Their preliminary price and cost calculation tasks would have encouraged the use of the confidence interval midpoint in order to simplify the task. Questions are also raised about the experimental procedure of the investment allocation task used in Chen &

Summers (1981) where participants received all treatment formats at the same time and were required to select between different companies under different treatment formats. However, it is interesting to note that the significant results of this study and that of Hawkins (1974) relate to similar tasks.

Other criticisms of the studies by Oliver (1972), Hawkins (1974), Keys (1978) and Chen & Summers (1981) include: no provision of feedback, which could have reduced the opportunity to learn how to utilise the UI⁴⁷; UI determined subjectively may have had low credibility and may not have been trusted by participants; and, there was little or no training in the meaning and potential use of probabilistic information cues (refer to Tables 2-1 to 2-6 for additional task characteristics).

The effect of UI outside accounting

Beach (1983) found that provision of probability information had little impact on decision-makers, and the results outlined above would appear to reflect this finding. Lack of significant results in an accounting setting led other researchers to investigate the impact of UI on decision-making in a more controlled environment, hoping to infer any significant results back to a financial setting. Birnberg & Slevin (1976), Johnson (1982) and Eckel (1983) all investigated UI in a non-financial setting and most obtained more significant and more interesting results than prior research into UI utilisation.

Birnberg & Slevin (1976) were the first wanting to infer results from a non-financial setting, but unfortunately they too did not obtain significant results. 100

⁴⁷ The lack of feedback in these studies was particularly emphasised in Foong (1992).

postgraduate students completed a choice problem that required them setting a target level and establishing a decision rule. As in earlier studies, participants were equally divided between those receiving implicit information regarding uncertainty (history information including 5 observations detailing the performance of other students completing the same task) and those receiving explicit UI (a confidence interval in addition to the same history information). Unlike earlier studies Birnberg & Slevin (1976) varied task uncertainty in order to investigate whether UI would have greater impact when there was a mismatch between an individual's expectation of uncertainty and underlying task uncertainty. The task was divided into two parts with the only difference between the two parts being the task uncertainty. Half of the participants received high task uncertainty first followed by low task uncertainty. The other half received low task uncertainty followed by high task uncertainty. The provision of UI made no difference to the decisions being made by participants. Those participants moving from low to high variance appeared to overreact to the increased uncertainty, although this was not due to the UI.

Johnson (1982) also investigated the impact of changing task uncertainty on the impact of UI on decision-making, but found a significant result when task uncertainty was high. Johnson (1982) devised a Bayesian revision task that required participants to make an initial probability judgment (pre-test phase) regarding a sample of dowel (wooden stick) length, a length estimation task, trials to investigate task uncertainty (a form of task training), followed by the post-test phase which required another judgment of dowel length subject to experimental treatments including environmental uncertainty (within subjects),

informational uncertainty⁴⁸ (3 levels between subjects) and report format (four levels between subjects). The reports, which showed a sample of dowel lengths from a population, came in four formats: point estimate; point estimate with probability of being correct (called an augmented point estimate), confidence interval; and a probability distribution. Participants, 65 undergraduate accounting students, were equally divided among twelve experimental groups across the varying levels of report format and informational uncertainty. Results showed that as informational uncertainty increased the accuracy of judgments got worse, except for those receiving the confidence intervals (a significant interaction effect). Those provided with confidence intervals actually improved their judgments as task uncertainty increased. Singling out the point estimate and confidence interval groups Johnson (1982) showed that at low and medium task uncertainty those provided with a point estimate had significantly lower judgment errors, while at the high level of informational uncertainty those with confidence interval information were superior in judgment. The confidence interval group appeared to be following a fundamentally different strategy to all other treatment groups.

Eckel (1983) used an experimental game in order to investigate the impact of probability information on both the decision performance within the game and also on the decision process itself. Eckel (1993) was keen to investigate not only whether probabilistic information altered decisions, but whether decisions were improved. He also wanted to see the influence of feedback and the source

⁴⁸ Informational uncertainty is another term for task uncertainty.

of UI (generated internally or supplied). His results are the most conclusive evidence that UI improves decision-making to date.

The task was an experimental game completed by 109 undergraduate students who were heavily motivated by 15% of their managerial accounting course grade depending on their performance in the game. The task involved participants acting as president of a hypothetical firm, with the objective being to maximise profits over 6 periods of business activity. Participants first obtained comprehensive instructions and discussed the experiment in detail. The game proceeded with the distribution of general economic forecasts to all participants, the optional purchase of more accurate economic forecasts and finally the submission of decisions regarding various organisational variables, such as selling price for the firm's product and the setting of advertising spending. There were four experimental groups, made up of two levels of forecast information (probabilistic or deterministic) and two levels of access to the computer budget (access or no access). The only difference in procedure was for those individuals allowed access to the computer budgeting tool. Prior to submitting decisions for the controllable organisational variables those individuals with access to the budgeting tool were allowed to interact with it as many times as they liked in order to help them make their decisions. The experimental group with probabilistic data and access to the computer algorithm was effectively creating its own UI.

Results showed those with access to probabilistic information made more accurate price and advertising decisions, and their operating results for the hypothetical company were much higher than the results of participants

provided with deterministic information. This was particularly true for those with access to the computer algorithm. Participants with probabilistic forecasts also purchased more information. A learning effect was also apparent, with performance improving over the length of the experiment, and the gap between probabilistic and deterministic groups getting larger. Of the groups with access to the budgeting algorithm, those with probabilistic information used the budgeting tool significantly more and considered more decision inputs. Eckel (1983) claimed that the budgeting algorithm had such a significant effect because it allowed the participants to generate their own cues as input to their final decisions, which would have increased reliability on those information cues, but particularly on UI.

More structured, but less realistic, experiments helped to better explain the potential role of UI on decision-making. Although Birnberg & Slevin (1976) did not obtain significant results, they led the way in using a non-financial setting for this field of research, and they investigated some interesting aspects of experimental task design, such as task uncertainty, which were successfully followed up by Johnson (1982) and Eckel (1983). There are two possible reasons why Birnberg & Slevin (1976) did not find an effect due to UI. First, participants received no training regarding UI. Second, the control group received history information (5 data points) while the experimental group received both history information and UI. There may have been enough history information given to the control group so that they had an adequate understanding of task uncertainty and the experimental group obtained no advantage by receiving UI. Also, since those in the experimental group received no training in understanding or using UI they may have disregarded the UI and

attempted to determine uncertainty by analysing the history information, again rendering their task the same as the control group's. In contrast, Johnson (1982) and Eckel (1983) benefited by providing direct and indirect training to participants in the use of UI. Johnson (1982) included trials to investigate task uncertainty between the pre-test and post-test phases of his experiment that acted as training in how to use UI. Similarly Eckel (1983) allowed half of his participants to access a computer budgeting algorithm (some with UI some without), which enabled a much better understanding of how decisions in an experimental game would impact the performance of a hypothetical firm. While Eckel felt that access to the computer algorithm improved the performance of those participants with UI because it effectively made the UI self-generated rather than provided by an external source, it is equally, if not more, likely that those with UI benefited from accessing the budgeting tool because it acted as training and enabled them to comprehend the UI and use it to advantage.

In a completely different area of endeavour, but much earlier than any of the studies discussed so far, Herman, Ornstein & Bahrick (1964) compared conventional (point estimate) displays of object location (similar to that appearing on a radar screen) with two forms of probabilistic displays. The probabilistic displays included one with two discrete confidence contours (like confidence intervals) at the 50% and 90% level centred on the point estimate and another display format which included a two-dimensional probability density function featuring different degrees of shading and again centred on the point estimate. Participants took part in a search and attack game where they had to determine at which point they would fire at a target through 10 stages as the object location information (and therefore probability of successfully hitting the

target) changed from stage to stage, with only 15 seconds being available for each stage. 40 trials were repeated in this fashion. Those participants with the confidence contours outperformed all other participants, including those with the more informative probability density functions, and were more calibrated in their decisions. Significant training took place prior to the experiment. Disaggregation of the 50% and 90% confidence contours into separate treatment groups in a follow up experiment had no effect on results. Despite its unusual application this study provides some of the strongest results in UI utility.

UI and the reporting of company earnings

More recently, a number of researchers have investigated the effects of company earnings forecast precision on the forecast revisions of analysts and the decisions of investors (for example Baginski, Conrad & Hassell, 1993; Pownall, Wasley & Waymire, 1993; Kennedy, Mitchell & Sefcik, 1998; Hirst, Koonce & Miller, 1999). Hirst, Koonce & Miller (1999) investigate the impact of forecast precision in combination with prior forecast accuracy. This combination is effectively the same as providing UI, except without the explicit identification of uncertainty. They found this combination to influence confidence in judgments, but not the judgments themselves. 121 postgraduate students completed an earnings prediction task. Participants assumed the role of an investor evaluating a hypothetical company, whose financial information was based on several actual companies from the same industry (manufacturers of snow mobiles). There were four treatment groups, accounting for two levels of forecast precision (point estimate or range) and two levels of prior forecast accuracy ('low' or 'high'). After receiving 3 years of financial performance for the

company participants were asked to make an initial earnings forecast for the next year and record their confidence in this forecast. Participants were then first provided with information on the prior accuracy of this company's earnings forecasts (simply 'high' or 'low'), and were then given the company's earnings forecast (point estimate or range forecast) for the next year. The company's point forecast was the midpoint of the range forecast. Finally, participants were asked to make revised earnings forecasts and record their confidence in these final decisions. While there was no impact on forecasts, prior forecast accuracy and forecast precision did influence investor confidence. Specifically, if prior forecast accuracy was high, and those forecasts were point estimates, participants' confidence in their own forecasts was significantly higher.

Hirst, Koonce & Miller (1999) reference Birnberg & Slevin (1976), Keys (1978) and Oliver (1972) in claiming that individuals faced with a range will use the midpoint. However, Johnson (1982) clearly showed that this was not necessarily the case. In Johnson (1982) those provided with a confidence interval followed a completely different strategy, and did not use the midpoint. Hirst, Koonce & Miller (1999) share a common characteristic with the earlier accounting related studies that failed to observe any difference in decision behaviour due to the presence of UI: the task involved the combining of several variables in order to compute some answer, in this case a forecast. As mentioned earlier the use of the midpoint would have been a strategy used to reduce information processing effort, thus rendering the task for those receiving a range or confidence interval equivalent to those receiving a point estimate. Also, the experiment was designed so that the range (and therefore the forecast which was the midpoint of the range) was always above the last actual, and

therefore the range added little value. One final point regarding this paper: their contention that point estimates imply greater accuracy than range estimates may well be more psychological than factual, particularly given the performance of organisations all over the world during the past few years including new-economy failings and major corporate collapses.⁴⁹

Recent research into the influence of UI on judgment

Ho, Keller & Keltyka (2002) contribute to the literature on UI indirectly, through their investigations of behaviour under ambiguity. They carried out an experiment looking at whether the framing of choices as gains or losses influenced attitude towards outcome ambiguity.⁵⁰ Their significant results have direct repercussions for the utility of UI. 39 MBA students in a 2 x 2 within subjects experiment completed a capital budgeting task for a hypothetical department, where the experimental treatments included the type of case study (internal rate of return or return on investment) and the likely return condition (gain or loss). Participants were given four separate projects to consider, one for each of the experimental treatment combinations. In each case the participant was provided with three pieces of information: a target rate of return for the hypothetical department; a point forecast provided by personnel for the

⁴⁹ For additional commentary on Hirst, Koonce & Miller (1999) see Kennedy (1999).

⁵⁰ In the same paper Ho, Keller & Keltyka (2002) also investigated the impact that framing choices in terms of gains or losses has on attitudes towards probability ambiguity. The results of this study are not directly relevant to this research and are not reported in detail here.

rate of the return for the particular project; and, a range forecast⁵¹ provided by personnel for the rate of return for the same project. The point forecast was the midpoint of the range forecast. Point forecasts were above the target level in the gain condition, while they were below the target level in the loss condition. The target level was always within the range forecast. For each of the four combinations of case and gain/loss condition (which were presented in random order) subjects had to decide on whether they believed the point forecast (referred to as a certain outcome) or the range forecast (an ambiguous outcome). After making each of the four decisions subjects rated perceived risk and perceived ambiguity for both the point forecast and range forecast alternatives. The type of case (internal rate of return or return on investment) had no impact upon results, however the return condition had a significant impact. Subjects believed the point forecast under a gain condition (ambiguity averse) and the range forecast in the loss condition (ambiguity seeking). Ho, Keller & Keltyka (2002) conducted a follow-up experiment that introduced risk to their original experiment, applying a probability distribution over the outcome range (for example, in the gain condition this said a 50% probability of the ROI being between 18% and 22%, and a 50% probability of the ROI being between 14% and 17.9%) and also to the point estimate (for example, in the gain condition it was stated there was a 50% probability of ROI being 20% and a 50% probability of ROI being 16%). As in the earlier experiment these

⁵¹ The range forecast was essentially a 100% confidence interval.

information cues were compared to a target level (which was 16%). In the gain condition the unambiguous, point outcomes were again preferred over the ambiguous range outcomes. However, results were slightly different under the loss condition: under risk there was equal preference for the ambiguous (range) and unambiguous (point) forecasts for return on investment.⁵²

Like Herman, Ornstein & Bahrack (1964), Ho, Keller & Keltyka (2002) clearly demonstrate the significance of target levels: creating a comparison point for contrasting range (or UI) and point information. Their other contribution is linking UI utility to whether the prevailing conditions signify gains or losses. However, the results of their follow-up experiment, which investigated propensity towards outcome ambiguity under risk, could have been more significant. First, the structure of the information cues was not very realistic, and second they did not create a task that was equivalent to the task in their first experiment, which looked at ambiguous outcomes. In the first experiment the unambiguous (point estimate) outcome was always clearly above or below the target level while in the follow-up experiment the unambiguous outcome always had one of its two values equal to the target level. This must have influenced the attractiveness of the unambiguous option in the loss condition, and may explain why the unambiguous range option had reduced popularity under risk: the preference reversal may hold under risk and under certainty.

⁵² There are other studies that investigate the effects of vague or ambiguous payoffs (for a recent example which shows the standard result of outcome ambiguity avoidance see González-Vallejo, Bonazzi & Shapiro, 1996) where preference for ambiguous payoffs (presented as confidence intervals) are compared to unambiguous payoffs (point estimates), however these studies mainly focus on whether or not ambiguity is avoided and generally do not help to understand how UI can be better utilised.

The final paper to be discussed in this section is Foong (1992, 2003). Foong (1992, 2003) reviewed all of the prior literature and attempted to systematically identify characteristics of task, environment and decision-maker that may have influenced the utility of UI. The factors that were of most interest to Foong and were investigated included the mode of display (graphical or tabular), the use of incentives, and the form of the loss function used to evaluate performance. Table 2-4 lists environmental factors of past studies, and it can be seen that the presence of incentives was commonly associated with those studies obtaining significant results. Review of the previous literature also convinced Foong that some additional characteristics must be present in order for UI to provide any benefit to decision-makers: feedback must be provided so that decision-makers may learn how to use UI; and, tasks must be repetitive and longitudinal to increase the benefits of the feedback.

Foong (1992, 2003) also found one of the key determinants for finding an effect due to UI was the presence of a task in which UI had diagnostic or decision-making value. Many of the earlier studies investigating UI utility did not demonstrate that their decision task benefited those with UI, providing one explanation for those studies' lack of significant results. This may have explained the tendency for subjects in earlier studies to use the mid-point of the confidence intervals rather than using the probabilistic information conveyed by the confidence intervals: the subjects could not see how they would benefit from the UI, and in fact the UI may not have added any benefit.

Foong (1992) conducted two experiments, the first an inventory forecasting task investigated the impact of forecast form (point estimate forecast

versus a combination of forecast and confidence interval), display mode (graphical or tabular) and loss function (symmetric, 'understock' biased or 'overstock' biased). The second experiment was a stock price forecast and investment decision, with the same structure as the first experiment except that incentives were manipulated rather than loss function (reported in Foong, 2003).

In the first experiment 60 participants were evenly divided among the 12 experimental groups. It was their task to forecast inventory for 36 periods in a row, starting with access to 20 periods of history. A forecast was provided regardless of whether the participants received UI (which was in the form of a confidence interval). The data was based on an artificially generated time series.

Interestingly, no significant results were found under the graphical display mode. However, significant results were found under the tabular mode of display. Although penalty cost (one of the measures used to evaluate performance) was no different between those groups that did and did not receive UI, decision accuracy was significantly improved in the group with UI using a tabular display. UI was only of benefit with one of the three loss functions, the 'understock' biased loss function, and only appeared to provide improvements during the early stages of the task. The 'understock' biased loss function may have elicited these results because of the upward trended time-series used in this task.

In the second experiment by Foong (1992, 2003) 80 postgraduates completed a stock price forecast and investment decision, spread amongst 8

treatment groups (2 levels of information, 2 types of display mode and 2 types of incentive condition). Participants made 20 consecutive investment decisions, representing 20 months' of decisions. During each month participants played the role of an investment adviser, selecting between 1 of 3 stocks. The same stocks were used throughout the 20 trials, and each time series started with 18 periods of historical data. Prior to selecting the preferred stock each month, the investment adviser was forced to forecast each of the 3 time series. The loss function was biased towards conservative decision-making. The aim was to yield a minimum target rate of return each month. Those provided with UI outperformed the control group only when incentives were provided, but were significantly worse than those without UI when no incentives were offered. These results were independent of display mode.

Foong showed us that the effective utilisation of UI is highly contingent upon factors such as loss function, incentives, and possibly mode of display. However, her results are not strong and it is still unclear whether UI utility is as contingent as it appears. The weakness of results in Foong (1992, 2003) may be due to some of the same factors already identified in other studies. For example, Foong provided her participants with a large amount of history information for the time series that decisions were based on. This history data could have acted as a proxy for UI, diminishing the potential benefit of the UI. Making participants produce a judgmental forecast may have led them to anchor on that information cue, distracting attention away from the UI. Finally, the use of confidence intervals based on a symmetric error distribution makes it possible and likely for individuals to utilise the midpoint of the confidence intervals, although this did not appear to be the case in these studies.

Regardless of any weaknesses Foong (1992, 2003) has shown that any benefit in decision-making due to UI is highly contingent (1992, p. 57):

“Prior studies of the impact of the CI (Confidence Interval) information on decision performance were based on the premise that risk information is the foundation for optimal decision making and the explicit disclosure of risk information, such as the CI information, should improve decision performance. No consideration had been given for the potential interaction effects that the displayed risk information might have with other task and/or contextual variables.”

Summarising the contributions of prior research into UI

Table 2-7 summarises the contributions of each study to our understanding of what contributes to UI utility. The table is based on the contingent model of UI utility and includes judgments regarding the importance of each factor from the model and lessons learnt, and our current overall level of understanding of that factor. The ‘Importance’ of each factor was judgmentally determined by considering the normative importance in conjunction with the descriptive importance (according to frequency of evidence) of the factor. ‘Our Understanding’ was determined by the strength and frequency of contributions.

As expected, Table 2-7 highlights the potential ‘Importance’ of task factors in determining whether or not UI will be utilised in decision-making, but it also confirms that these factors have not been adequately investigated relative to their importance (denoted by the fact that ‘Our Understanding’ is low relative to the ‘Importance’ of this factor). Taking ‘Level of task uncertainty’ as an example of a significant task related factor, it has been pointed out that Johnson (1982) showed task uncertainty could influence UI utility while Birnberg & Slevin (1976) found no such result. This is a factor that may be crucial to UI utility, but its

influence is not well understood and accordingly there is a strong impetus to investigate further (see Table 2-7).

Normative Model	Importance	Lessons Learnt	Contributions to our understanding										
			Our Understanding	Birnberg et al. (1976)	Chen et al. (1981)	Eckel (1983)	Foong (1992, 2003)	Hawkins (1974)	Herman et al. (1964)	Hirst (1999)	Ho et al. (2002)	Johnson (1982)	Keys (1978)
Task													
• UI format	●	Midpoint of a range can be used.	●	●	●			●	●	●	●	●	●
• Presence of UI proxies	●	UI can be ignored.	●	●		●	●		●			●	●
• Level of task uncertainty	●	UI beneficial at high uncertainty.	●	●								●	
• UI adds obvious value	●	Otherwise, suboptimal strategies.	●	●	●				●			●	●
Person													
• Task training	●	Increases UI understanding and use.	●	●	●	●	●		●			●	●
• Expertise/Experience	●	May reduce UI utility.	●									●	●
Environment													
• Payoffs/Incentives	●	Appear necessary for UI use.	●	●	●	●	●		●	●	●	●	●
• Feedback	●	Possibly confusing with UI.	●	●	●	●	●		●			●	●
• Time pressure	●	May reduce UI utility.	●						●				
• Realistic task context/story	●	Required for UI utilisation.	●										
• Loss function	●	UI utility when loss function is asymmetric.	●			●							
Interactions													
• Intuitive expectations of risk	●	UI may be useful when intuition and reality differ.	●						●		●	●	●
• Individual perceptions of risk	●	UI beneficial when uncertainty perceived as high.	●	●							●		
• Calibration of confidence	●	Calibration may influence belief in UI.	●						●				
• Probabilistic Ambiguity	●	Point estimates preferred.	●							●	●		
Risky Behaviour													
• Losses weigh heavy	●	Losses increase concentration.	●		●	●			●		●		●
• Avoid risks where target level jeopardised	●	Target levels increase UI salience.	●						●		●		
• Avoid ruinous losses	●	Risk-averse behaviour with large losses.	●		●								
• Risk seeking in threatening situations	●	Threats may alter decision strategies.	●						●				

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

Contributions to our understanding The extent to which each individual study contributed to our overall understanding of that factor

* Not derived from the model of contingent effects

○ Very Low

◐ Low

◑ Medium

◒ High

◓ Very High

Table 2-7: A comparison of normative/descriptive importance and our understanding of factors with influence over UI utility.

Research investigating UI highlighted two factors that were not part of the preliminary model of contingent effects, but may still have significant influence on UI utilisation. First, UI will only be of benefit when it has diagnostic or decision-making value for a given task. The potential value of UI also needs to be fairly obvious and easy to extract, otherwise decision-makers will ignore UI or use simplification strategies. For example, when UI is provided in the form of a confidence interval and the task does not make the value of UI obvious to decision-makers, simplification strategies will be encouraged, such as the use of the confidence interval's midpoint, and there will be no benefit to the provision of UI. This situation appears to commonly occur in tasks that require calculations combining variables that are reported as confidence intervals because there are obvious cognitive benefits to minimising information processing. It also seems more common when there is no point-of-comparison, such as a target level, for comparing the confidence interval with. Foong (1992, 2003) picked up on this need for UI task value, and commented that the tasks used in some of the earlier studies did not have an optimal solution, the value of the UI was therefore difficult to determine, and as a result this may have reduced the utility of UI.

The second, less significant, factor that emerged was the loss function. Work by Foong (1992, 2003) indicates that the loss function needs to be asymmetric (which clearly highlights the decision value of the UI), and the direction of the asymmetry is also important. Both of these factors have been added to Table 2-7, and they are denoted with an asterisk in the far right of the table to indicate that they are not derived from the preliminary model of contingent effects.

Reviewing literature that investigates the potential utility of UI has highlighted the greater significance of some parts of the preliminary model and two additional factors. In particular, it appears that the following factors are of most importance:

- UI format (task)
- Presence of UI proxies (task)
- Level of task uncertainty (task)
- UI adds obvious value (task)
- Task training (person)
- Intuitive expectations of task (interactions)
- Losses weigh heavy (risky behaviour), and
- Avoid risks where target level jeopardised (risky behaviour).

While it is still possible that many of the other factors not listed above are just as important in determining UI utilisation (for example, an individual's experience or expertise), empirical findings currently draw attention to the eight factors listed above.

2.3.1 A Refined Model of Contingent Effects for Decision-Making under Uncertainty with UI

Key Messages:

- This section refines a model of contingent effects describing the contingent utilisation of UI, based on the preliminary model and augmented by empirical observation.

- Two factors have been added to the preliminary model: 'UI adds obvious value' and 'Loss function'.

Having developed a preliminary model outlining those factors most likely to influence UI utility, and then reviewing the literature that has directly investigated the potential utilisation of UI, a refined model describing UI utility will now be specified: a contingent model of behaviour based on the preliminary model and augmented by experimental observation. Figure 2-10 describes the refined model, and it can be seen that it is identical to the preliminary model apart from the addition of two factors: 'UI adds obvious value' in task, and 'Loss function' in environment (reasons for their inclusion can be found at the end of section 2.3).

This refined model of contingent effects will be used to guide research questions aimed at understanding the contingent utilisation of UI, and will also guide the design of any experiments resulting from the research questions.

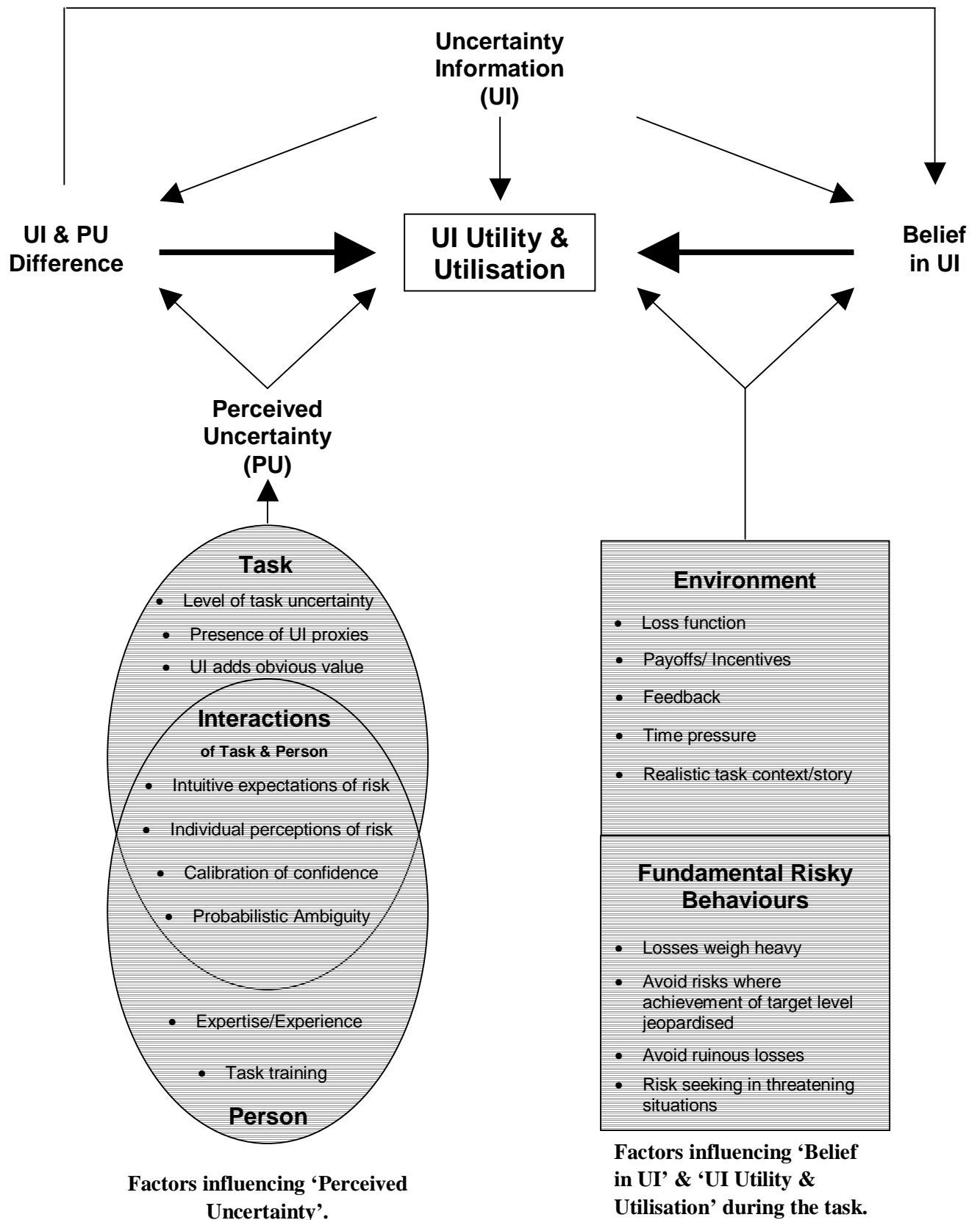


Figure 2-10: A refined model of contingent effects describing the contingent utility of UI

2.4 RESEARCH QUESTION FOR EXPERIMENT 1

Key Messages:

- The refined model of contingent effects creates a map for future empirical research investigating UI utilisation.
- Numerous studies will be required before a contingent model describing what impacts UI utilisation will have strong predictive powers.
- This study can only hope to define a model for future research efforts and contribute to our understanding of just a small part of the contingency model.
- Experiment 1 will investigate whether UI influences decision performance, when the UI has diagnostic or decision-making value.

The refined model of contingent effects defined in section 2.3.1 (see Figure 2-10) creates a map for empirical research investigating UI utilisation, to the extent that this area is currently understood. This model, supplemented by our understanding of each of its components (see Table 2-7), should drive ongoing research efforts into understanding UI utilisation. As can be seen in Table 2-7 our understanding of the contingent utility of UI is low. Although there are some hints, the exact influence of each factor and the contexts within which they apply are yet to be fully understood. Numerous studies will be required before a contingent model describing what impacts UI utilisation will have strong predictive powers.

This study can only hope to define a model for future research efforts and contribute to our understanding of just a small part of the contingency model.

Only a small, select group of factors from the refined model of contingent effects will be investigated while other factors are held constant in order to maximise the chances of a significant finding: using what has been learnt in prior studies.

The main research question for Experiment 1 is:

- When proxies for UI are not present, will the provision of explicit, objective UI that has diagnostic or decision-making value improve decision performance?

The following chapter will expand upon this research question, define hypotheses, describe an experiment to investigate the hypotheses and report on the results of the experiment. The focus of the first experiment will be to develop an appropriate task for demonstrating the utility of UI.