Investigating the Influence of Outcome Utility on Estimates of Probability

Adam J. L. Harris

A thesis submitted for the degree of Doctor of Philosophy

September 2009

School of Psychology

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

UMI Number: U584402

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U584402
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed .... (candidate)   Date ...24/1/2010.........

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed .... (candidate)   Date ...24/1/2010.........

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed .... (candidate)   Date ...24/1/2010.........

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)   Date ...24/1/2010.........

Acknowledgements

This thesis would not have been possible without funding from the Economic and Social Research Council.

I should like to thank the School of Psychology's Cognitive Group and SPC (Social Psychology Club). The feedback received from both greatly benefited this thesis. I should also like to thank Harriet Over and Jacky Boivin for feedback on written work.

Adam Corner, James Close, Carl Hodgetts and Andreas Jarvstad (the 'Hahn lab') provided assistance with data collection, numerous intelligent contributions during discussions, as well as even more numerous less-intelligent contributions during lab-group drinks outings. In particular, I would like to thank Adam Corner. Not only were the preliminary studies in Chapter 2 part of a joint project that included him, but he has constantly been a sounding board for ideas and a source of advice.

I should like to thank all the technical support staff in the School of Psychology, both those in the I.T. department, and the graphics department.

Most importantly, I am pleased to be able to thank my supervisor, Ulrike Hahn. Throughout the three years that I have spent on this thesis, she has been a source of knowledge, advice and encouragement. She could not have done more to make my PhD experience more rewarding or more enjoyable. I very much hope that we will continue to collaborate together on future projects.

Thesis Summary

This thesis is concerned with the rationality of human probability estimates, specifically the potential influence of outcome utility on estimates of probability. Intuitively, and normatively, the desirability of an outcome should not make that outcome seem more or less likely to occur. Chapter 1 provides a background to the subsequent empirical work by addressing some general issues surrounding a probabilistic approach to human reasoning. The major questions addressed are whether people represent uncertainty quantitatively and their competence at doing so, considered with reference to the status of extant biases in the literature on human probability judgment. Chapter 2 presents seven studies investigating the effect of negative utility on estimates of probability, using a minimal paradigm in which there is a visually defined, objective, probability. When there was an indication that future human action could influence the likelihood of an outcome (the outcome was, in some way, controllable), negative outcomes were rated as more likely to occur than neutral outcomes.. This moderating effect of control can be given a decision-theoretic explanation in terms of loss function asymmetry (e.g., Weber, 1994). Consequently, these effects can be understood as rational reactions to the recognition of the uncertainty of human cognition. Chapter 3 investigates the effects of positive utility using the same visual representation of probability. Across four studies, no effect of positive outcome utility was observed, a result consistent with the asymmetric loss function explanation proposed for the findings in Chapter 2. Chapter 4 presents a statistical-based critique of the 'unrealistic optimism' phenomenon (e.g., Weinstein, 1980). Additionally, two empirical studies failed to find any evidence that the 'unrealistic optimism' phenomenon is more than just a statistical artifact. The results from all three experimental chapters provide support for the contention that people's probability estimates are not systematically biased by utility considerations.

Contents

Introduction

"Uncertainty is the only certainty there is"

(John Allen Paulos)

To a greater or lesser extent, every aspect of human life is characterised by uncertainty. Having asked a colleague whether or not he shall see them in the office the following day, Andy should not be certain that he will see them on the next day upon receiving an affirmative reply. They, or he, might be taken ill and not make it to work, for example. Of course, some uncertainties are more uncertain than others. The likelihood of the 12:30 train from Cardiff to London leaving at 12:30, for example, is more uncertain than the likelihood of Andy seeing his colleague the next day.

In order to function successfully, people must be able to live with uncertainty in their lives. When Andy leaves the house to travel to work, he must make a decision about whether or not to carry an umbrella based on uncertain information from the weather forecaster, who may tell him that there is a 30% chance of rain. The standard way of representing uncertainty and quantifying it so as to guide rational action is through probabilities. Indeed, the standard normative model of decision making, 'Subjective Expected Utility' theory (SEU) (Savage, 1954) posits that, in order to guide rational action, people should combine the utility (the subjective 'goodness' or 'badness') of a possible outcome with the probability of that outcome occurring. Returning to the earlier example, assume Andy assigns a utility value of -100 to the outcome 'getting wet', a utility value of -10 to the outcome 'carrying the burden that is an umbrella' and a utility value of 0 to the outcome of 'not carrying an umbrella and not getting wet' (i.e., it does not rain). SEU prescribes that for each of the four combinations of events (Table 0.1), Andy should weight the utility of that outcome by the probability of that outcome, in order to calculate the expected utility (EU). To calculate the EU of carrying an umbrella, Andy should combine the utility associated with carrying an umbrella given that it rains with the probability of it raining *and* the utility associated with carrying an umbrella given that it does not rain with the probability of it not raining. The EU of carrying an umbrella is then computed by summing across the two uncertain outcomes (rain and no rain). Andy should then calculate the EU for not carrying an umbrella in the same manner and choose the action with the greatest EU. In this case, this corresponds to carrying an umbrella, as:

$$EU_{umbrella} = (-10 \times 30\%) + (-10 \times 70\%) = -10$$

$$EU_{no\ umbrella} = (-100 \times 30\%) + (0 \times 70\%) = -30$$

Table 0.1
The four key possible outcomes, and their utilities, under consideration when deciding whether or not to carry an umbrella

		Weather Event (probability)	
		Rain (30%)	No rain (70%)
Possible Actions	Umbrella	-10	-10
	No umbrella	-100	0

A key aspect of many of the real-world probabilities that humans care about is that, in contrast to the majority of laboratory based judgment contexts, the probabilities associated with the potential outcomes are not well specified. Rather, the decision maker must engage in a process of reasoning in order to derive subjective estimates of the probability of different events. Thus, whilst a patient suffering from gangrene is aware that she does not want to die and she does not want to lose her leg, but she would rather lose her leg than die, the probabilities associated with dying if the patient does not choose to have her leg amputated may not be immediately apparent. It is likely that the patient will seek information from a variety of sources in order to help her make a judgment of the relevant probabilities and thus help guide her decision.

Similarly, a juror for a criminal case is aware that she does not want to find an innocent man guilty, or a guilty man innocent. Blackwell's maxim, that it is better for ten guilty men to go free than one innocent man to be convicted (e.g., Nagel, Lamm, & Neef, 1981), furthermore suggests that the former is a more negative outcome than the latter. To derive estimates as to the likelihood of the defendant's guilt, the juror must make a judgment based on the evidence presented in court.

The examples presented above demonstrate that in real-world decision problems, whilst probability judgments are crucial, they cannot typically simply be 'read off' the environment, rather they must be constructed from the available evidence. Such construction might require complex reasoning or merely a simple estimate. However, once probabilistic information is not readily available to inform decision-theoretic calculations, there are a number of factors that might bias these probability *judgments*. A thorough understanding of potential biases is necessary to fully understand human judgment and decision making.

Although there are a number of important issues related to such probability judgments that have emerged from past research, the majority of judgment research has not specifically taken other aspects of the decision context into account, although reasoning researchers have begun to consider the potential for decision-theoretic concerns influencing reasoning (e.g., Bonnefon, in press). One salient feature of any decision context is the utility associated with different possible outcomes. In line with this, the systematic laboratory-based investigation of the effect of utility on probability estimates presented in this thesis therefore represents an emphasis on a key situational characteristic of real-world judgments.

For any important probability judgment, there will be utilities present. This is because a judgment is only 'relevant' if it is made in order to inform a decision, and decisions always involve utilities. The more consequential a decision, the more extreme the utilities are likely to be. If utility does bias judgments of probability, then such a bias is likely to be fundamental and prevalent throughout human life, as people are constantly making decisions in order to guide their actions. Moreover, given that more consequential decisions are associated with more extreme utilities, such a bias is likely to be amplified in precisely those decision-making contexts in which people are most concerned with accuracy – important and consequential decisions.

In SEU, probabilities and utilities should be combined to derive expected utilities. One assumption intrinsic to this model is that the processes of deriving these component parts, the probabilities and utilities, are independent. Edwards (1962) points out that whilst this assumption is not critical to the mathematical content of the model, "it is very difficult to see how the model could be applied to real decisions unless some such assumptions were made" (p. 43). This difficulty in applied settings is demonstrated when one considers the equation for SEU. If utility biases judgments of probability then the calculation of either component in isolation becomes an incredibly complex task. It is our belief that this was what Edwards meant by the importance of the independence assumptions for the model to be applied to real decisions. Despite the seeming practical importance of the independence of judgments of probability and utility, typical studies of human judgment have been made in situations where utilities are not readily apparent (see e.g., Tversky & Kahneman, 1974). Chapters 2, 3 and 4 of this thesis present a systematic investigation of the potential interdependence of probabilities and utilities.

The importance of whether outcome utility biases estimates of probability does not rest on the normative theory of SEU alone. This research question is equally important given any current mainstream descriptive theory of decision making. All theories of decision making assign both probability and utility a key role. Thus, the question of whether utility systematically biases estimates of probability is a fundamental question in human cognition.

Descriptive theories that have sought to capture people's deviations from the prescriptions of SEU are mostly variants on the normative model. These theories include: Rank-dependent expected utility models (Quiggin, 1993); Prospect theory (Kahneman & Tversky, 1979a); Cumulative prospect theory (Tversky & Kahneman, 1992); Configural weight models (e.g., Birnbaum, 2004) and Regret theory (Loomes & Sugden, 1982). The most famous and influential of these, Prospect theory, was purposely designed in an attempt to achieve "the minimal set of modifications of expected utility theory that would provide a descriptive account of...choices between simple monetary gambles" (Kahneman, 2000, p. x, as cited in Brandstätter, Gigerenzer, & Hertwig, 2006, p. 411). Consequently, the practical issues of a possible interdependence between utilities and subjective probabilities is a key concern given any of these descriptions of human decision making.

There are two descriptive models of decision making that do not take the normative prescriptions of SEU as a starting point (see e.g., Oaksford, Chater, & Stewart, in press, for a review). However, both require judgments of probability in utility laden contexts. Decision by Sampling (DbS) (Stewart, Chater, & Brown, 2006; Stewart & Simpson, 2008) draws on psychophysical research demonstrating that people are better at relative judgments than absolute judgments. For example, it is easy to determine which of two lines is longer, but somewhat more difficult to specify the lengths of the lines according to some absolute measure (e.g., in centimetres). Stewart and colleagues propose, therefore, that people evaluate both the utility and probability of an uncertain outcome relative to a comparison sample, which can comprise of items in the current problem space, or other items from working memory. Consequently, the relative rank (r) of an attribute (e.g., utility of an outcome) is given by $r = (R - 1) / (N - 1)$, where R is the rank within the sample of N comparison outcomes (which includes the target outcome). Thus, for example, within a comparison sample of gains {2, 5, 7, 100, 150, 20000}, the subjective value of a gain of 7 is (3 – 1) / (6 – 1) = 2/5. Note that this subjective value is unaffected by the absolute size of the gains incorporated in the sample. The subjective value of a probability is calculated in the same way as for utilities, but if the probabilities refer to losses then they are assigned a negative value.

DbS is not derived from the normative framework of SEU and it does not require the multiplicative combination of subjective probabilities and utilities, but it does propose a consideration of both probabilities and utilities (for further details, see Stewart & Simpson, 2008). Consequently, the question of whether they are subjectively interdependent is important from a DbS perspective.

Generalising Gigerenzer and colleagues' research program into fast and frugal heuristics of cognition (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999), Brandstätter et al. (2006) proposed the priority heuristic as a descriptive process model of how choices are made. In the case of a choice between two gambles (or actions), it is proposed that people evaluate the prospects in the following order: Minimum gain, probability of minimum gain, maximum gain (for losses, simply replace 'gain' with 'loss'). If the difference between the two prospects differs by more than a criterion amount at any stage in this process then the process is stopped and the decision maker chooses the more desirable prospect. In addition to assigning a key role to judgments of both utility and likelihood, Brandstätter et al. observe that the heuristic does not overcome considerations of expected utility when there is a marked difference in the expected utility associated with one choice versus the other. Rather, "the heuristic performed best when the ratio between expected values was about 2:1 or smaller" (p. 429).

The debate over the best descriptive model of human decision making is not one which will be addressed in this thesis. Having provided only a very brief overview of the core tenets of two decision making models that do not have their roots in SEU, it should be clear that they nevertheless assign crucial roles to both probability and utility. Thus, judgments situated in the context of real-world decisions must be made in the context of both probabilities and utilities. The empirical work presented in this thesis is therefore of relevance to decision making, whichever descriptive model is endorsed.

The possibility of a biasing effect of utility on probability judgments intuitively also raises more basic level concerns over the rationality of human cognition. Such a bias would be analogous to believing the world to be a certain way simply because we do (or do not) want it to be that way. Consequently, the presence of such a bias might suggest a flaw in the most basic process of how probability judgments are derived. Given the prevalence of a decision-theoretic context for most consequential probability judgments, a utility bias on probability estimates would be near universal in its prevalence, and thus a far more critical challenge against human rationality than many of the other classic biases (e.g., the conjunction fallacy, base rate neglect, framing effects), which will be discussed in Chapter 1.

There are two important features of many real-world probability judgments that have yet to be highlighted explicitly, but which it is necessary to highlight in order to place the next chapter in context. Firstly, in all the judgment situations outlined above, including the gangrene patient, the juror, and even the case of deciding whether to take an umbrella to work, the judgment concerns that one particular occasion or event. Thus, this thesis is primarily concerned with *single-event* probabilities. When deciding whether to convict or acquit a defendant, the probability of guilt should be a subjective judgment based on the evidence presented in this single case. In deriving single-event probability estimates, people will, however, (and indeed, should) also make use of frequency data. Frequency information can inform single-event probability estimates as a guide to prior degree of belief (for example). In estimating the likelihood of a particular neighbour being a lawyer, if I know that half of my neighbours are lawyers and half doctors, I should use this frequency information to inform my estimate of this particular neighbour being a lawyer (50%) (more on this below). Secondly, as real-world judgments are potentially associated with extreme consequences, these are the judgments for which it is most important that people get them *right*. For this reason, the framework theories used in this thesis will be normative ones: the issue of what constitutes a right decision *is* a normative question.

Overview of the Thesis

To supplement the description of SEU above, Chapter 1 will introduce the normative theory of Bayesian probability. In writing a thesis on the topic of probability judgments, it is important to acknowledge those research traditions that have questioned people's ability to reason using probabilities. By way of providing a background to the present work, some key findings from this literature, as well as their counter-arguments are summarised, demonstrating that Bayesian probability is far from being a superhuman normative theory that mere mortals are utterly unable to approximate. Rather, given the right tasks and facilitators, people are often able to perform remarkably well on probability judgment tasks, as exemplified in a recent study of our own (Harris & Hahn, 2009). Chapter 1 thus provides an introduction to the literature on human probability judgment, in order to place the subsequent empirical work in a firm theoretical context.

Having set the theoretical context for this work, Chapters 2, 3 and 4 provide a systematic investigation of the potential biasing effect of utility on estimates of probability. Despite a great deal of research often cited as evidence that utility does influence probability (e.g., Irwin, 1953), in the introduction to Chapter 2, we show that such a conclusion is premature. Furthermore, in Chapters 2 and 3 we undertake a systematic investigation of the possible interdependence of event utility and probability estimates. From the research in these chapters, we conclude that whilst utility might lead to a biasing of probability estimates in practice, it will only do so through mediating mechanisms. Increased understanding of these mechanisms will allow the identification of those situations where probability estimates are most likely to be biased. In Chapter 2, the proposed mediating mechanism is loss function asymmetry. Hence, people bias their estimates of probability in one direction in order to guard against the consequences associated with a more costly error in the opposite direction. Such a bias can be considered rational in many situations (e.g., Batchelor & Peel, 1998). Chapter 4 extends this research by critiquing the ubiquitous finding of unrealistic optimism in the Social Psychology literature, the phenomenon whereby people estimate negative events as less likely to happen to themselves than to others (e.g., Weinstein, 1980). We conclude that the existing evidence is not sufficient to enable the conclusion that this is a universal human bias. The conclusion from these three chapters is that there is presently insufficient evidence to suggest that probability estimates are routinely biased by utility, although there exist real-world situations in which a bias may emerge through various mediators.

Chapter 1 - Theoretical Background

As outlined in the Introduction, this chapter will introduce the most established normative framework for probability judgments, as well as address various critiques of the normative frameworks of judgment and decision making. Specifically, this chapter aims to address two critiques that might be levied against the line of research pursued in this thesis. Firstly, why investigate yet another bias in human probability judgment? Do we not already know that people are poor at probabilistic reasoning? Errors such as the conjunction fallacy (e.g., Tversky & Kahneman, 1982) and base rate neglect (e.g., Kahneman & Tversky, 1973), as well as biases of overconfidence (e.g., Lichtenstein, Fischhoff, & Phillips, 1982), conservatism (e.g., Phillips & Edwards, 1966) and framing effects (e.g., Tversky & Kahneman, 1981) might suggest that people are unable to even approximate the prescriptions of probability theory. We shall address this issue in the present chapter by arguing that people's probability judgments might not be as susceptible to error as such research has often suggested and that these research lines are far from extinct (see also, e.g., Kynn, 2008). The second critique is more fundamental and was stated by Gerd Gigerenzer when he was the discussant in the 'reasoning and uncertainty' session at 'EUROCOGSCI 07,' following the first presentation of work from this thesis (Studies 1 and 2). Gigerenzer seemed to claim that people simply do not represent single-event probabilities.

Having introduced the normative theory of Bayesian probability, we shall address each of the outlined critiques in turn, beginning with the more fundamental second question. This chapter will conclude with an example from our own research demonstrating good probabilistic reasoning by naïve participants in a novel experimental paradigm (Harris & Hahn, 2009). The results from this study provide support for the contention that people are able to aspire to the normative prescriptions of Bayesian probability in many contexts.

Bayesian Probability

Bayesian probability is derived directly from the fundamental mathematical axioms of probability theory and is concerned with the internal consistency and coherence of subjective probabilities. Within the Bayesian framework, therefore, probabilities are conceptualised as subjective degrees of belief, rather than as objective frequencies existing in the external environment.

Bayesian probability is not the only means with which it has been proposed that people can deal with uncertainty. Indeed, the fundamentality of probability has been questioned by some researchers (see e.g., Politzer & Bonnefon, 2009). Alternative rational theories for reasoning under uncertainty include (non-exhaustively): Plausible reasoning (e.g., Rescher, 1976), a "rational instrument" (p. 1) in which "the conclusion of a piece of reasoning takes its status from that of the "weakest" premiss" (p. xi), the Shafer-Dempster school of non-additive beliefs (Shafer, 1976), fuzzy set theory (Zadeh, 1965; see also, e.g., Schum, 1988, 1994), rough set theory (Pawlak, 1982), certainty factors (Shortliffe & Buchanan, 1975), epistemic belief theory (Spohn, 1990), possibility theory (e.g., Dubois & Prade, 1988), and logical, argumentation-based approaches to uncertain reasoning and decision making (e.g., Amgoud, Bonnefon, & Prade, 2005; Fox, Krause, & Ambler, 1992). A further alternative proposed in the literature is that of explanatory coherence (e.g., Thagard, 1989, 2000) by which hypothesis evaluation is a constraint satisfaction problem. Thagard proposes Explanatory Coherence as both a descriptive and normative theory. His argument, however, is made entirely on descriptive grounds citing factors such as the multitude of conditional probability judgments required for a probabilistic analysis of a problem, in conjunction with his argument that explanatory coherence is "psychologically natural in that it views inference as analogous to neurological processes in which multiple neurons interact in parallel" (Thagard, 2004, p. 236).

Probability theory, however, is a well established normative framework. Indeed, Lindley (1982) argues that "only probability is a sensible description of uncertainty" (p. 1). Specifically, Lindley (1982, 1994) demonstrates the 'inevitability' of the basic probability axioms, starting only with the assumptions that uncertainty can be represented by a number and that an urn containing some white and some black balls can represent a standard measurement of uncertainty (in the same way as length is measured with relation to the metre). The fundamental axioms of probability theory are (see also, e.g., Howson & Urbach, 1996): Probabilities are constrained to be real numbers between 0 and 1; tautologies are assigned probabilities of 1; the joint probability of exclusive events is equal to the sum of their individual probabilities. From these three axioms[1], all the mathematical laws of probability necessarily follow. Furthermore, as shown by de Finetti (1974), given a scoring rule by which a person incurs a penalty of $(1-p)^2$ if an event is found to be true and p^2 if an event is found to be false (where p denotes a numerical value previously assigned by the person to the likelihood of the event in question), a person will *necessarily* incur a larger penalty if their likelihood estimates do not obey the probability axioms (the so-called 'Dutch book' argument). Lindley (1982, 1994) argues that if other scoring rules are used then either people should provide responses that are, in reality, only transformations of probability (e.g., odds), or people should only estimate 0 or 1 (demonstrating the inadequacy of such a scoring rule). "All sensible rules lead back, via a possible transformation, to probability. Probability is inevitable" (Lindley, 1994, p. 6; see also, e.g., Cox, 1946; Horvitz, Heckerman, &

---

[1] Some authors (e.g., Howson & Urbach, 1993) consider the definition: $P(a \mid b) = \dfrac{P(a \,\&\, b)}{P(b)}$, where $P(b) \neq 0$, to be a fourth axiom of probability.

Langlotz, 1986; Snow, 1998). Lindley's argument is explicitly related to subjective probability, and he stresses the importance of considering it as a function both of the event being contemplated and of the knowledge of the person contemplating that event. Consequently, his argument is one in support of Bayesian probability.

Lindley's (1982, 1994) arguments are examples of 'Dutch book' arguments, which are frequently used to support the normative status of Bayesian probability (see e.g., Howson & Urbach, 1996). Dutch book arguments relate to a person's betting tendencies, and rely on the relatively uncontroversial principle that it is undesirable to engage in a betting strategy by which your opponent, without having any special knowledge, is able to guarantee your loss. If an individual's subjective beliefs are coherent with respect to the probability axioms, then this protects them against engaging in betting strategies that would guarantee a loss (Howson & Urbach, 1996), such as the following example: Suppose Nancy believes that there is a .75 chance of England winning the 2010 FIFA world cup. As a non-Bayesian, Nancy also believes there is a .55 chance of England not winning the 2010 FIFA world cup (thus violating the complementarity rule [$P(\text{not } a) = 1 - P(a)$]). According to her subjective probabilities, Nancy should be willing to wager £75 against a bookmaker's £25 that England will win the cup. However, she should also be willing to wager £55 against a bookmaker's £45 that England will not win the cup. Having bought both these bets, Nancy is doomed to lose money. Whether England win or don't win the world cup, Nancy will collect £100 from the bookmaker. However, her total stake will have been £75 + £55 = £130, resulting in a net loss of £30 (see Howson & Urbach, 1996, for mathematical proofs as to the presence of Dutch book gambles for bets that do not obey the probability calculus). By contrast, if Brenda the Bayesian also believed England to have a .75 chance of winning the cup, as a Bayesian she would be constrained to assigning a .25 chance to them not winning the cup. Consequently, she would be willing to place the same bet as Nancy on the possibility of England winning, but she would only be happy to bet £25 against a bookmaker's £75 on the possibility of England not winning. The consistency in this betting pattern therefore saves Brenda from a guaranteed loss. For if she is persuaded to bet on both the complementary events, she is guaranteed to break even, whilst Nancy would ensure a guaranteed loss for herself.

It is the root of Bayesian probability in the axioms of probability with their established normative status that makes it such an attractive framework within which to investigate human cognition. Given that we desire important real-world likelihood judgments to be *right*, we take the established norms of Bayesian probability as the normative framework within which we undertake our empirical investigation.

Do People Represent Uncertainty Quantitatively?

Having outlined the normative credentials of Bayesian probability, we now turn our attention to the more fundamental of the two critiques outlined above: Do people reason with probabilities? As outlined above, probabilities have been described as the inevitable normative way to represent uncertain beliefs (Lindley, 1982). In order for Bayesian probability to be a suitable normative framework within which to investigate human cognition, it would seem necessary for people to possess some sort of quantitative representation of uncertainty, otherwise probability could not be a meaningful psychological construct. At a more basic level, the methods used in this thesis will require participants to estimate probabilities. Such a methodology necessarily assumes that people are able to represent uncertainty quantitatively.

In order to address this question, we will consider evidence that is independent from the decision making literature discussed thus far. The first evidence for people's sensitivity to uncertainty and, indeed, different degrees of uncertainty, comes from the number of verbal probability expressions that exist to express it. Table 1.1 shows 69 possible such expressions. In

Table 1.1

A sample of verbal probability phrases in the English language.

Core terms	Possible prefixes
Definitely	most; almost
Definite	
Likely	un; very; highly; not; most
Probable	im; most; very; quite; highly
Probably	
Sure	very; completely
Chance	slim; slight; sure; great; good; very good; there is a…; very low; poor; low; small; non-negligible; reasonable; meaningful; high; very high; big
Possible	very; almost
Impossible	absolutely; practically
Might	
May	
Perhaps	
Toss-up	
Even odds	
Even chance	
Certain	almost; absolutely; nearly; close to; not
Doubtful	
Can't rule out entirely	
Chances are not great	
Not inevitable	
One must consider	
It could be	
One can expect	
Reasonable to assume	
It seems	
It seems to me	
One should assume	
To be expected	
One chance out of two	

Note. Phrases in this table are taken from: Beyth-Marom (1982); Budescu & Wallsten (1995); Mullet & Rivet (1991); Smits & Hoorens (2005); Wallsten, Budescu, & Zwick (1993).

the presence of such a rich corpus of language to convey uncertainty, it seems likely that people should have a good understanding of its quantitative nature. This is further supported by the acknowledgement that, despite such a rich corpus, people understand the vagueness associated with such terms and consequently they generally prefer to receive probabilistic information in numerical (as opposed to linguistic) form (Budescu, Weinberg, & Wallsten, 1988; Erev & Cohen, 1990; Wallsten, Budescu, Zwick, & Kemp, 1993). Were uncertainty an inherently qualitative construct psychologically, it would be very strange for people to demonstrate such a preference for numerical information.

Recent research in neuroeconomics (whose goal is expressed as being "to better understand decision-making behavior by taking into account the cognitive and neural constraints on this process, as investigated by psychology and neuroscience, while also utilizing the mathematical decision models and multiplayer tasks that have emerged from the field of economics" [Sanfey, 2007, p. 151]), has shown areas of brain activation that appear to be sensitive to manipulations of probability (e.g., Fiorillo, Tobler, & Schultz, 2003; Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004). Fiorillo et al. and Morris et al. used single cell recordings of neuronal activation in monkeys and found sensitivity of dopamine neurons to reward probability such that neurons showed minimal responding to positive outcomes that were perfectly predicted (probability = 1). When, however, positive outcomes occurred with probabilities less than 1, the magnitude of dopamine neurons responses increased as the probability of the reward decreased. Essentially, therefore, "the dopamine neurons' response reflects mismatch between expectation and outcome in the positive domain" (Morris et al., 2004, p. 133). Moreover, Tobler, O'Doherty, Dolan, and Schultz (2007; see also, Knutson, Taylor, Kaufman, Peterson, & Glover, 2005) effectively demonstrated the disassociation of reward magnitude and probability in an experimental design that shed new light on decision making under risk. In a human fMRI study, Tobler et al. showed that brain activation in the caudate and ventro-medial putamen correlated with increases in both reward magnitude and probability. A medial prefrontal region showed specific sensitivity to the probability manipulation whilst remaining insensitive to variations in reward magnitude. Most impressively, this study showed that activity in the medial and posterior striatum was specifically correlated with changes in expected value whilst remaining invariant to changes in reward magnitude and probability that did not produce a corresponding change in expected value. That is, for example, "activations differed insignificantly between stimuli predicting 100 reward points with $P = 1.0$ and 200 points with $P = 0.5$ (expected value 100 points), but activations were higher than for stimuli associated with an expected value of 50 points and lower than for stimuli associated with an expected value of 150 points" (Tobler et al., 2007, p. 1626). The discovery that brain activation patterns in the medial and posterior striatum were sensitive to the multiplicative interaction of probability and utility provides some support for the idea that the brain encodes probabilities quantitatively. Whether the manipulations of outcome utility and probability used in this study enable quite such a bold conclusion to be drawn is not clear, but it is nevertheless an intriguing piece of evidence.

This thesis will not make further mention of research from the emerging field of neuroeconomics. It suffices to say that, as the results of Tobler et al.'s (2007) study demonstrate, this is a research area with great promise and it may indeed hold the key to definitively deciding between competing descriptive accounts of human decision making behavior: "Demonstrating that brain areas do indeed weight and sum probabilities and values is an important piece of evidence that the family of utility-theory models may well be an accurate representation of how the brain decides between alternatives" (Sanfey, 2007, p. 153).

More evidence suggesting that humans can represent uncertainty quantitatively, in a format that would be amenable to probabilistic reasoning comes from recent studies with infants. Téglás, Girotto, Gonzalez, & Bonatti (2007) presented evidence that 12 month old infants are able to quantify uncertainty, even in the absence of long run frequency data. Téglás et al. argued that infants are able to distinguish different probability levels (simple 'more likely' versus 'less likely' distinctions) for single-event probabilities dependent on the specific event characteristics. The studies that Téglás et al. used involved presenting infants with movies showing objects bouncing around inside a container. One object subsequently exited the container and infants looked significantly longer at the movie when the exit was improbable than when it was probable, thus demonstrating a recognition that this was an unexpected occurrence. For example, Téglás et al.'s Study 1 involved four objects inside a transparent container, which had an open pipe at its base, "as in a lottery game" (p. 19156). Three of these objects were identical, whilst one was of a different colour and shape. Infants looked at the movie significantly longer when the 'different' object exited the container than when any of the three other objects exited the container. This and similar studies led Téglás and colleagues to conclude that infants were sensitive to the prior probabilities of the various outcomes.

Evidence that infants can distinguish between a likely and an unlikely outcome for a single event probability can be taken as evidence that humans are predisposed to represent uncertainty quantitatively. The evidence is not, however, as conclusive as Téglás et al.'s (2007) study suggests. Girotto and Gonzalez (2008) conducted a series of studies to investigate whether children are able to revise their probabilistic beliefs in the light of new information. Girotto and Gonzalez tested children between the ages of 3 years 8 months and 11 years 3 months. In Studies 1 and 2, the youngest children (all under 5 years old) were unable to perform above chance in choosing the more likely of two alternatives in a prior probability condition similar to that used in Téglás et al.'s studies with infants. In Girotto and Gonzalez's prior probability condition, the children were shown two puppets with names that corresponded to their colour (e.g., Mr. Black and Mr. White). Children were told that Mr. Black owned the black chips and Mr. White owned the white chips. They were then shown four chips, three of which were one colour (e.g., black) and one of which was the other colour (e.g., white) (as a memory aid, they were also given a piece of card depicting the four chips). The chips were subsequently put in a bag. The experimenter told the children that they were going to pick a chip out of the bag and that the owner of that chip would win a chocolate. Children were required to indicate which puppet they would like to be in order to win the chocolate (Study 1), or asked which puppet was more likely to win the chocolate (Study 2). If children understood the probability distribution in this task, they should have chosen to be the puppet who owned the predominant number of chips in the bag. Conceptually, this task is similar to that in Téglás et al., and yet the (approximately) 4 year old children in this study performed at chance, suggesting that they were unable to explicitly use the representation of ordinal probability that the infants appeared to display in Téglás et al.'s study. This inconsistency does not invalidate Téglás et al.'s result, as the inconsistency may simply highlight the extra difficulty associated with verbally explicating such a representation. It does, however, suggest the need for further research in this area.

Girotto and Gonzalez (2008) recognised that the drawing of a chip from a box, although a single event, is nevertheless a random, repeatable event. In their Study 3 they therefore used a non-repeatable, single, event produced by an intentional agent. Kindergartners (mean age 5 years 10 months, range 5 years 5 months to 6 years 5 months) and older children were able to reliably use the probabilistic information provided to guide their decisions (again between a less likely and a more likely event). This study further supports human comprehension of single-event probabilities, even in the absence of educational experience. The weight of evidence does, therefore, seem to support the contention that people represent uncertainty quantitatively, although further research is required to establish this contention beyond doubt.

Human Biases in Probability Judgment

A thesis investigating a potential bias relating to human probability judgment would not be complete without a discussion of other biases previously highlighted in the literature (e.g., Kahneman, Slovic, & Tversky, 1982; Phillips & Edwards, 1966). Consequently, in this section we will provide a brief, critical summary of research demonstrating other failures of human probability judgment. Although the described biases might suggest that future work investigating the rationality of human probability judgment is obsolete, in reality such biases might be largely constrained to the laboratory and not accurately represent the competency of real-world human probability judgment. The tasks in which these biases manifest themselves require participants to aggregate unfamiliar probabilistic information, in a manner akin to a mathematics test. Even if people were completely unable to approximate the prescriptions of probability theory in such tasks, the degree to which this research reflects human probabilistic reasoning in the real-world would remain unclear. The brief review presented next provides further support for this argument.

Throughout the 1970s and 1980s a substantial literature developed, questioning the rationality of people's probabilistic reasoning. This literature was dominated by

Kahneman and Tversky's 'heuristics and biases' research program (e.g., Kahnemann, et al., 1982; Tversky & Kahnemann, 1974), and also by research demonstrating that people's belief updating is conservative with respect to Bayesian prescriptions (e.g., Edwards, 1968; Fischhoff & Beyth-Marom, 1983; Peterson & Miller, 1965; Peterson, Schneider, & Miller, 1965; Phillips & Edwards, 1966; Phillips, Hays, & Edwards, 1966; Slovic & Lichtenstein, 1971). The conclusions of these research programs hindered the potential growth in popularity of Bayesian approaches in applied domains. The resistance to Bayesianism displayed by some applied researchers is summed up in a quote from Pennington and Hastie (1993, p. 213):

> "It is generally known that the Bayesian system is an inadequate description of human behavior under most conditions".

It has, in fact, been claimed that the general conclusion of researchers and students without an expertise in judgment and decision making research echoes that of Pennington and Hastie (1993) (as noted in Christensen-Szalanski & Beach, 1984). Past research is not, however, as unequivocal as it is often portrayed (see e.g., Kynn, 2008). Although the heuristics and biases program in particular has received much publicity, there is a considerable amount of research that actually suggests people's reasoning is able to approximate Bayesian prescriptions (see Kynn, 2008, for a concise review). Conclusions such as those expressed by Pennington and Hastie may arise partly from the citation bias in judgment research, whereby articles demonstrating poor human judgment are cited more often than those demonstrating good judgment. Although Christensen-Szalanski and Beach, in the wake of the heuristics and biases research, found that there were 37 articles demonstrating good human judgment between 1972 and 1981 and 47 demonstrating poor human judgment, the latter were calculated as being between six (Christensen-Szalanski & Beach, 1984) and three times (Robins & Craik, 1993) more likely to be cited than the former. In the following sections, we present a review of the most famous 'non-Bayesian' biases and the research they have prompted. We will introduce the biases and summarise some of the major critiques against the existence of the individual biases. At the end of the two sections that follow, the reader may conclude (as we do) that, although humans are not perfect Bayesian reasoners, they are not as inherently irrational as is often believed. There are certainly enough instances of good probabilistic reasoning to motivate further research within this framework, as Bayesianism appears to be a normative standard that people can at least aspire to (see also, Parsons, 2001, p. 29).

Conservatism and Overconfidence

Conservatism was the predominant finding of early research investigating whether Bayes' Theorem could be considered a good descriptive model of human reasoning, that is, a model of what people actually do (e.g., Edwards, 1968; Peterson & Miller, 1965; Peterson, Schneider & Miller, 1965; Phillips & Edwards, 1966; Phillips, Hays, & Edwards, 1966). Typical tasks in which conservatism is observed are bookbag and poker chip tasks. In these tasks, different coloured chips (e.g., red and blue) are drawn from a bag in front of the participant. Participants are told that the bag could consist of one of a number of different proportions of red and blue chips. For example, the bag from which the chips are being drawn could be an 80/20 bag, a 60/40 bag, a 40/60 bag, or a 20/80 bag (red/blue chips). For each possible bag, participants must estimate the probability that the chips are actually being drawn from that bag. In a typical, 'online', judgment task, participants revise these probability estimates following each draw from the bag. When compared against the prescriptions of Bayes'

Theorem, participants' probability estimates typically do not change as much as they should (i.e., they are conservative).

Later research using a different paradigm yielded a strong, replicable, result that appears opposite to the conservatism finding. This result was one of overconfidence (see e.g., Lichtenstein et al., 1982, for a review). Probability estimates are said to be overconfident if they are anti-regressive with respect to the true, objective, probability. That is, if the true objective probability, $P(x)$, lies below .5, an overconfident estimate will be less than $P(x)$. Similarly, if $P(x)$ is greater than .5, an overconfident estimate will be greater than $P(x)$. The relationship between objective probability (x) and overconfident subjective probability (y) can therefore be expressed as:

$$y = mx + c$$

where $0 < m < 1$ and $c > 0$ (Figure 1.1). Overconfidence is the dominant finding from studies employing a calibration paradigm. In a calibration study, participants are typically required to answer a series of true-false general knowledge questions, or multiple choice questions, and also to provide a confidence level, as a subjective degree of belief in the correctness of their answer. To determine the calibration of the participants' performance, the researcher places participants' answers into 'bins' depending on the confidence level given. That is, all answers reported with a confidence between 85% and 95% are placed in a single 'bin'. A well calibrated individual should therefore answer between 85% and 95% of these questions correctly. Typically, however, participants are overconfident (see Figure 1.1) in their confidence ratings. That is, of those questions which participants claim to be 90% sure of the answer, they answer less than 90% of them correctly. Likewise, if (in yet another instantiation of the paradigm) participants are judging the probability that a statement is true and they report a probability of 10%, such an estimate is overconfident if more than 10% of statements assigned such a probability are true. The intuitiveness of this labelling as overconfidence is clear when one considers that such a response is identical to one in which participants answer a true/false question as false and then report 80% confidence in their answer.

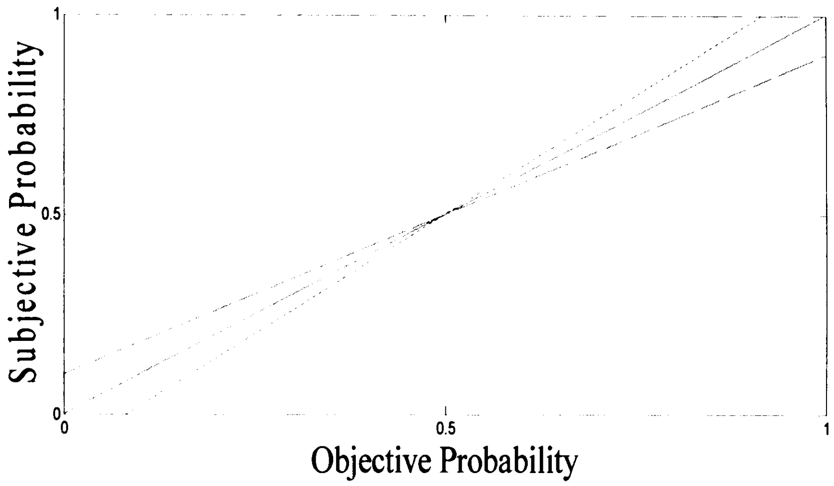


Figure 1.1. Probability estimates that are anti-regressive from 0.5 (dotted line) represent overconfidence, whilst those regressive from 0.5 (dashed line) represent underconfidence. The solid line represents perfectly calibrated probability judgments.

The finding of overconfidence has been much critiqued in the literature, leading to the conclusion that judgment researchers have "been overconfident in their conclusion that probability forecasters are overconfident" (Pfeifer, 1994, p. 203; see also, Soll, 1996). Erev, Wallsten and Budescu (1994) extended the scope of this critique by further accounting for results demonstrating conservative belief revision and demonstrating that both effects could be accounted for as resulting from the same statistical mechanism. Erev et al. noted that the data analysis in calibration studies (typically showing overconfidence) was 'back to front' from that in belief revision studies. That is, in calibration studies objective probabilities are analysed as a function of subjective probabilities to determine the match between the two. In revision-ofopinion studies, in which the dominant finding is one of underconfidence, or 'conservatism' (see, Edwards, 1968; Fischhoff & Beyth-Marom, 1983; Rapoport & Wallsten, 1972; Slovic & Lichtenstein, 1971), subjective probabilities are analysed as a function of objective probabilities. Erev et al. demonstrated that data from the same experiments could be re-analysed to show either over- or underconfidence depending on the analysis chosen. Furthermore, a model assuming an accurate underlying representation of the objective probability, but with an error component added to the response, led to responses that were regressive to the midpoint of the scale, which resembled overconfidence if objective probability was analysed as a function of subjective probability and underconfidence if subjective probability was analysed as a function of objective probability. In other words, there was no real fact to the matter of whether responses were over- as opposed to underconfident. Furthermore, both might simply reflect unbiased, random error. Thus both overconfidence and conservatism may simply be results of an error prone, but systematically unbiased judgment process with the opposing findings reflecting the different methods of data analysis employed (for a concise review of further critiques of 'conservatism' see Ayton & Wright, 1994).

This line of research raises question marks over the validity of the methods and analyses used in studies demonstrating overconfidence and conservatism in human probability judgment. Subsequently, the extent of either bias in human everyday reasoning is unclear, and it may even transpire that people are neither overconfident or conservative in their probability judgments.

'Heuristics and Biases'

The heuristics and biases research program relates to a substantial body of work undertaken by Kahneman and Tversky (e.g., 1973, 1979b; Tversky &

Kahneman, 1973, 1974, 1982, 1983). Their research showed, using pencil-and-paper laboratory tasks, that people were almost universally susceptible to a variety of fundamental judgment biases, including: Framing effects (e.g., Tversky & Kahneman, 1981), base rate neglect (e.g., Kahneman & Tversky, 1973), and the conjunction fallacy (e.g., Tversky & Kahneman, 1982). All three of these biases have generated a great deal of subsequent research.

Framing effects

Framing effects reflect the phenomenon that phrasing identical information in different ways affects the choices that people make. For example, if people are informed that an experimental drug treatment has a 20% mortality rate within five years, they are less likely to choose the treatment than if they are told it has an 80% survival rate (Marteau, 1989; McNeil, Pauker, Sox, & Tversky, 1982; Wilson, Kaplan, & Schneidermann, 1987). Prima facie, this inconsistency in people's choices when presented with logically equivalent information appears to be fallacious. McKenzie and colleagues (McKenzie & Nelson, 2003; Sher & McKenzie, 2006, 2008), however, have argued that although the information in the two 'framing' conditions is logically equivalent, it is not informationally equivalent. In natural language, there are a myriad of ways that this same information can be conveyed. That the speaker has chosen one means of conveyance is not random, and therefore provides information relevant to the hearer's choice. McKenzie and colleagues have demonstrated that the way in which a speaker frames a decision provides information both about the action that they would recommend, and about its position relative to an implied reference point. In the above example, by framing the treatment in terms of its 80% survival rate, a physician is recommending that they would take the treatment, as well as providing information that 80% is better than the 'standard', and is thus a good statistic. By contrast, framing the treatment in terms of the 20% mortality rate implicitly conveys the information that a 20% mortality rate is higher than you should accept. According to the reference point hypothesis, and empirically supported in McKenzie and Nelson (2003) and Sher and McKenzie (2006), descriptions are more likely to be framed in terms of an attribute that is above the reference point, than one that is below the reference point. Sher and McKenzie (2006), for example, presented participants with one full glass of water and one empty glass of water. They asked participants to transfer water from one glass to the other and then present 'a half-full cup', or a 'half-empty cup.' Assuming that the starting state for each cup represents its reference point, the reference point hypothesis predicts that participants should present the previously empty cup as the 'half-full cup' and the previously full cup as the 'half-empty cup.' These were precisely the results Sher and McKenzie observed.

McKenzie and colleagues' critique of previous interpretations of the framing effect demonstrates the importance of a full consideration of the information available to the participant during an experimental task. Furthermore, it demonstrates that framing effects do not necessarily imply that a participant is not rational. Similar, pragmatic-based, critiques have been levied against other tasks supposedly demonstrating the irrationality of people's probability judgments.

The conjunction fallacy

The most pervasive and well-researched of the biases identified by Kahneman and Tversky is the conjunction fallacy (e.g., Kahneman & Tversky, 1982; Tversky & Kahneman, 1983). The significance of the conjunction fallacy is demonstrated by the fact that a Google Scholar search with the term "conjunction fallacy" yielded 2,240 hits (12/08/2009). Example materials that have led to the demonstration of this bias are:

> "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations" (Kahneman & Tversky, 1982, p. 126).

In the simplest version of this task, participants are asked which is more probable?

(i) Linda is a bank-teller;

(ii) Linda is a bank teller who is active in the feminist movement.

Between 80 and 90% of statistically naïve participants typically rate (ii) as more probable than (i) (Hertwig & Gigerenzer, 1999; Kahneman & Tversky, 1982; Tversky & Kahneman, 1983), thus violating the mathematical principle that the conjunction cannot be more probable than either one of its constituent elements.

In order to report the 'correct' answer (that (i) is more probable), participants can ignore all the information presented in the materials and simply base their answer on the response options, as (i) must always be more likely than (ii) according to probability theory. Note, therefore, that these materials carry an element of pragmatic deceit. According to Grice's (1975/2001) cooperative principle of conversation, the maxim of relation requires speakers to make their contribution relevant. If the 'correct' answer is most easily achieved by ignoring potentially distracting information then clearly the experimenter is flouting the maxim of relation by including this irrelevant information (see also, Hilton, 1995, p. 250; Kahneman & Tversky, 1982). If participants expect the experimenter to be co-operative in his utterances, then they are entitled to infer that information provided is relevant to their judgment, or why would it be provided? Consequently, they will erroneously try to make use of it in the judgment task, potentially leading them to commit a conjunction error through an overweighting of its diagnosticity.

In addition to the issue outlined above, the conjunction fallacy has also been critiqued on a number of other grounds including: The ambiguity of how to interpret 'Linda is a bank-teller' (e.g., Agnoli & Krantz, 1989; Dulany and Hilton, 1991; Macdonald, 1986; Markus & Zajonc, 1985; Morier & Borgida, 1984); the ambiguity of the word 'probable' (Hertwig & Gigerenzer, 1999); and the ambiguity of the word 'and' in versions of the problem using 'and' to specify the conjunction, 'Linda is a bank teller and is active in the feminist movement' (Hertwig, Benz, & Krauss, 2008). The ambiguity in this task is further demonstrated by a much higher percentage of 'don't know' responses in a Linda task than in a variety of other judgment tasks, both simple games of chance involving dice and coins and Tversky and Kahneman's (1974) maternity ward study (Hertwig, Zangerl, Biedert, & Margraf, 2008). A further argument against the claim that the conjunction fallacy is evidence of human irrationality is presented in Bovens and Hartmann (2003). They recognise that the conjunction fallacy is not a violation of probability theory if the information about Linda (that she is a bank teller etc.) is interpreted as being a report from a partially reliable source, rather than a statement of fact. There is, therefore, debate concerning the degree to which a violation of the conjunction rule in the 'Linda' task truly represents a reasoning error (in addition to references above, see e.g., Chase, Hertwig, & Gigerenzer, 1998; Politzer & Noveck, 1991; Wolford, Taylor, & Beck, 1990; but see also, e.g., Bar-Hillel, 1991; Mellers, Hertwig, & Kahneman, 2001; Tentori, Bonini, & Osherson, 2004).

Such research, along with McKenzie and colleagues' research into framing effects (see above), has highlighted the importance of understanding the pragmatic implications involved in such judgment tasks. Hilton (1995) has also suggested a pragmatic explanation for results of studies demonstrating base rate neglect (see also, Birnbaum, 1983). It should, however, be noted that Kahneman and Tversky themselves did not conclude from the heuristics and biases program that people are inherently irrational in their probability judgments. Their conclusion is more process oriented, arguing that people make their probability judgments through useful, but fallible heuristics such as representativeness and availability. The experiments were designed precisely in such a way so as to 'bring out' the fallibility of these judgment processes (as hinted at above). Consequently, "it is not surprising that useful heuristics such as representativeness and availability are retained, even though they occassionally lead to errors in prediction or estimation" (Tversky & Kahneman, 1974, p. 1130). In this thesis we are not concerned with the *process* by which people make their probability judgments. Rather, it is important to conclude at this point that, despite the straw-man based conclusions of (mostly) less experienced researchers (as noted in Christensen-Szalanski & Beach, 1984), the heuristics and biases research program does not demonstrate that people's probability judgments are inherently irrational (see also, Kynn, 2008).

One result arising from the 'heuristics and biases' debate is that there are often ways of improving people's performance on such tasks. One way in which performance is greatly facilitated in such tasks is by the presentation of the information in a frequency format, rather than as a probability (e.g., Cosmides & Tooby, 1996; Gigerenzer, 1994; Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Gigerenzer & Hoffrage, 1995).

Frequency representations are not, however, the only ways in which reasoning performance can be improved on such tasks. A number of researchers have proposed, for example, that the improved performance on pencil-and-paper judgment tasks brought about by a frequency representation results not from the specific nature of this representation per se, but from its clarification of certain, relevant aspects of the reasoning problem (e.g., Agnoli & Krantz, 1989; Ayton & Wright, 1994; Barbey & Sloman, 2007; Evans, Handley, Perham, Over, & Thompson, 2000; Girotto & Gonzalez, 2001; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999; Mellers & McGraw, 1999; Sloman, Over, Slovak, & Stibel, 2003; see also, Hattori & Nishida, in press). The "nested sets" hypothesis (e.g., Barbey & Sloman, 2007; Sloman & Over, 2003; Sloman et al., 2003) implies that frequency descriptions improve performance on traditional reasoning tasks, on which people are typically fallible (such as the conjunction fallacy and base rate neglect), because the importance of considering category instances (an 'outside' view of probability judgment) versus category properties (an 'inside' view of probability judgment) is highlighted. This account assumes that the representation of instances (as in a frequency description) makes the set inclusion relations between them transparent. The best example to demonstrate the importance of a recognition of the set inclusion relations is the conjunction fallacy. Once the set inclusion relations inherent in a conjunction fallacy scenario are understood, specifically the recognition that the conjuncts are entailed by the conjunction, the fallaciousness of judging the conjunction as more likely than either of its conjuncts is made clear.

Agnoli and Krantz (1989; see also, Fisk & Pidgeon, 1997) demonstrated that training with Euler circles (Figure 1.2) and highlighting that a category is smaller when extra properties are added to its definition, served to reduce instances of the conjunction fallacy. Furthermore, Sloman et al. (2003) demonstrated that no facilitation in reasoning performance was observed for a frequency presentation when the nested-sets relation was made opaque, further supporting the view that it is the transparency of the nested-sets relations rather than the frequency format per se that results in less reasoning errors.

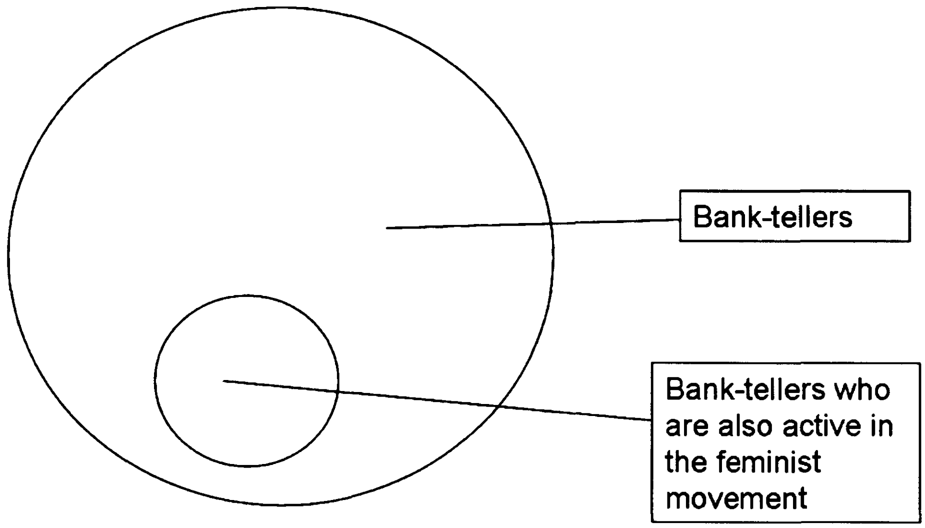


Figure 1.2. A Euler circle representation of Kahneman and Tversky's (1982) 'Linda' problem.

Base rate neglect

The nested sets hypothesis has also been proposed to account for improved reasoning performance brought about by frequency representations of base rate neglect tasks (Sloman et al., 2003). Kahneman and Tversky (1973) provided the first systematic investigation of base rate neglect, which is best demonstrated with reference to two of their experimental conditions. Kahneman and Tversky presented participants with the following cover story:

> "A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70

> lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100." (Kahneman & Tversky, 1973, p. 241).

When presented with only this information and asked to judge the probability that an individual chosen at random from this sample was an engineer, the median response was the normatively correct response of 30%. However, if also presented with an uninformative description of the randomly chosen individual (Dick), the median estimate of Dick being an engineer or a lawyer were both 50%. Participants clearly appear to ignore the base rate information when provided with additional information, even though it was non-diagnostic. Kahneman and Tversky explained such responses in terms of the representativeness heuristic. The nondiagnostic description is equally representative of both engineers and lawyers and hence participants provide an estimate of 50%. This represents a normative violation of the prescriptions of Bayes' rule, by which subsequent information should be combined with prior information to determine posterior probability. When subsequent information is non-diagnostic, the normatively correct posterior probability is equal to the prior probability.

Cosmides and Tooby (1996) demonstrated that a frequentist representation of a traditional base rate neglect problem resulted in a large reduction in the instances of base rate neglect errors made by participants. Their investigation was based on an experimental scenario first used in Casscells, Schoenberger, and Graboys (1978), which demonstrated participants tending to completely neglect base rate information. Cosmides and Tooby ran a number of experiments based on this scenario, including a version of the original Casscells et al. experiment designed to be conceptually identical to the frequency versions Cosmides and Tooby used (aside from its use of probabilities rather than frequencies [Cosmides & Tooby, Experiment 5]). Using this scenario, Cosmides and Tooby did not observe total base rate neglect, but the normatively correct response was still rare. In Experiment 5, Cosmides and Tooby used the following scenario:

> "The prevalence of a disease X is 1/1000. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, 5% of all people who are perfectly healthy test positive for the disease.
>
> What is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?_______%" (Cosmides & Tooby, 1996, p. 39).

The correct Bayesian solution to this problem, *P(D|T)* (probability that the person has the disease given a positive test result) is given by Bayes' Theorem:

$$P(D \mid T) = \frac{P(D)P(T \mid D)}{P(D)P(T \mid D) + P(\neg D)P(T \mid \neg D)}$$

where *P(D)* is the prior probability that the person has the disease, *P(T|D)* is the sensitivity of the test – the likelihood of a positive test result given the person has the disease. *P(T|¬D)* is the false positive rate – the likelihood of a positive test result when the person does not have the disease. Inserting the numbers from Cosmides and Tooby's example:

$$P(D \mid T) = \frac{.001 \times 1}{(.001 \times 1) + (.999 \times .05)} = .02$$

Thus, the normatively correct response is 2%. In this version of the problem, only 36% of participants gave this response (32% of participants demonstrated total base rate neglect by reporting 95%). When, however, the problem was presented in a frequency format, 72-80% of participants gave the normatively correct response (Cosmides & Tooby, Experiment 2 Condition 1 and Experiment 3 Condition 2).

As stated above, Sloman et al. (2003) argue that it is not the frequency representation per se that improves reasoning performance on tasks such as this. Rather, one by-product of a frequency representation is that it makes set inclusion relations clear. Sloman et al. found no difference in responses between problems represented in a frequency format, a probability format in which the nested-sets relations were made clear, or a probability format which was accompanied by Euler circles (Figure 1.3). The lack of a difference between the probability format in which the nested-sets relations were made transparent and a frequency format demonstrated the merit of the nested-sets hypothesis. The Euler circles manipulation served to strengthen the story and to demonstrate that there are a variety of manipulations which can improve participants' reasoning performance once nested-sets relations are made transparent. Furthermore, Sloman et al. reported that the use of Euler circles did not provide an additive improvement in reasoning performance across conditions. Rather, it only improved performance in the probability condition in which the nested sets relations were not made transparent, providing evidence for the contention that it is the transparency of these relations that is critical in improving reasoning performance on such word problems. These results provided more support for Sloman et al.'s assertion that: "Arithmetic operations that follow from transparent nested-set relations are easy to perform generally and not just in frequency problems" (Sloman et al., 2003, p. 298).

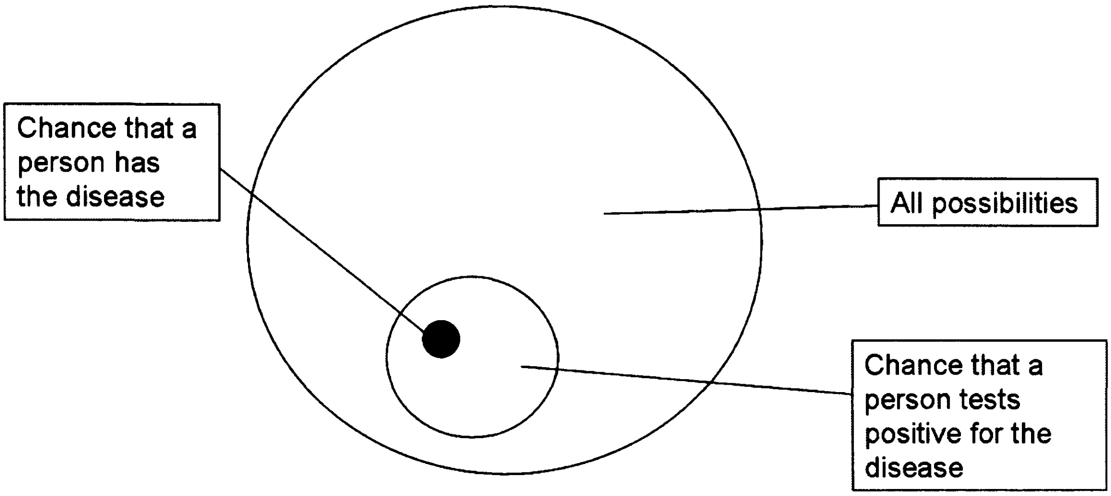


Figure 1.3. Euler circles used in the probability condition of Sloman et al.'s (2003) Experiment 2.

Hattori and Nishida (in press) have provided a similar, but more specific account for base rate neglect, based on the principle of equiprobability. We will not go into the details of their account here, but it suffices to say that they too emphasise the importance of making task structure clear to participants. Furthermore, they provide yet more examples of ways in which to improve performance on base rate neglect tasks, including the use of stimuli with which participants have greater real-world knowledge. Such demonstrations further question the degree to which traditional pencil-and-paper judgment tasks reflect the competence of everyday probabilistic reasoning.

Summary

Recent research investigating the competency of human probability judgment has raised question marks over the conclusion that people are inherently poor Bayesian reasoners. When tasks are constructed in such a way as to facilitate the correct understanding of a problem (even if this is a simple rephrasing in terms of everyday events with which people are familiar [e.g., Hattori & Nishida, in press]), people are often able to approximate the Bayesian response. Such results suggest

Bayesianism as a standard to which human reasoning should aspire, and therefore as a suitable normative framework within which to undertake psychological research.

The 'Probabilistic Turn'

The importance of a thorough understanding of real-world probability biases is further supported by the recent revival of interest in the question of the competence of human probability judgment relating to a variety of phenomena. In addition to the work mentioned above relating to the conjunction fallacy (e.g., Bovens & Hartmann, 2003) and framing effects (e.g., Sher & McKenzie, 2008), we can add recent work questioning the fallaciousness of people's supposed misperceptions of randomness (Hahn & Warren, 2009). This renewal of interest in human probability judgment has coincided with a 'probabilistic turn' in cognitive science generally (see Chater & Oaksford, 2008; Chater, Tenenbaum, & Yuille, 2006; Oaksford & Chater, 2009). Within this 'probabilistic turn', Bayesian models have been applied to a variety of areas of human cognition, including: Causal cognition (e.g., Griffiths & Tenenbaum, 2005; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003), vision (e.g., Weiss, Simoncelli, & Adelson, 2002), language acquisition and processing (e.g., Tenenbaum, Griffiths, & Kemp, 2006; for a review see, Chater & Manning, 2006) and sensorimotor control (for a review see, Körding & Wolpert, 2006). In these areas, human cognition has been shown to be well matched to the prescriptions of optimal models. In the area of sensorimotor control, for example, people seemingly demonstrate a good understanding of the uncertainty associated with movement, to be able to adopt a strategy that maximises reward, as prescribed by Bayesian Decision Theory (e.g., Trommershäuser, Maloney, & Landy, 2003, 2008). Given the ability of people to approximate the Bayesian standard in these tasks, the failure to do so in many probability judgment tasks (see above) represents something of a paradox. That people's performance on these tasks can be improved (e.g., Hattori & Nishida, in press; Sloman et al., 2003) represents one piece of evidence suggesting that, perhaps, such a paradox is more illusory than real. As Chater et al. (2006) point out, it is not surprising that people struggle with probabilistic tasks, as people generally struggle with much of mathematics. However,

> 'the fact that, for example, Fourier analysis, is hard to understand does not imply that it, and its generalizations, are not fundamental to audition and vision. The ability to introspect about the operations of the cognitive system are the exception rather than the rule – hence, probabilistic models of cognition do not imply the cognitive naturalness of learning and applying probability theory' (Chater et al., 2006, p. 288)

The quote above further questions the degree to which demonstrations of probability judgment errors are troublesome for a probabilistic approach to cognition.

Despite the argument that 'the cognitive naturalness of learning and applying probability theory' (Chater et al., 2006, p. 288) is not necessary for a probabilistic account of human cognition, all theories of decision making do assign expectancies a key role. People must therefore be able to *represent* these expectancies in some way, and if such representations are susceptible to biases then so will the resulting decisions be. Thus, prevalent real-world biases of probability judgments are a fundamental research topic within a probabilistic account of human cognition.

An Example of Competent Probabilistic Reasoning

Thus far in this chapter, we have argued that there is evidence that people are likely to be able to represent uncertainty quantitatively. The established normative basis for a Bayesian approach to human reasoning supports continued research within this framework. Although a number of judgment biases have been identified, we have argued that the scale of these biases is unclear, and they are typically most apparent in pencil-and-paper laboratory tasks that resemble a mathematics test to participants. That performance can be improved by clarifying certain aspects of these problems further questions their prevalence and relevance in real-world judgment situations. By contrast, for more fundamental cognitive abilities, the Bayesian framework appears to provide a good descriptive model of human behaviour (see e.g., Chater et al., 2006).

Having provided above a brief review of research suggesting that people are poor at probabilistic reasoning tasks, in this section we will describe a study in which people appear to be very proficient, thus providing further support for a probabilistic approach to human reasoning. The described study (Harris & Hahn, 2009) demonstrates rational belief updating on the basis of multiple witness testimonies, and consequently has applied consequences for real-world decision making.

In real-world decision making contexts, including formal contexts such as the courtroom, people must often aggregate information they receive from different sources. Bovens and Hartmann (2003) demonstrated how, in Bayes' Theorem, coherence, prior belief and source reliability combine to determine how likely a set of testimonies is to be true. Two simple assumptions are required: Firstly, individual testimonies are assumed to be conditionally independent of each other, that is, the witnesses are conveying their own observations and have not, for example, influenced each other. Formally,

$$P(R_1 \mid F) = P(R_1 \mid F, R_2)$$

where R_i is a report from source i and F is the fact about which they are reporting. Secondly, witnesses are assumed to be partially reliable, that is they are not setting out to lie, but do not necessarily report the truth. Formally,

$$p > q > 0$$

where p = the true positive rate (chance of the witness stating F is true given that F is indeed true) and q = the false positive rate (chance of the witness stating that F is true when it is not). This also seems reasonable. If the witnesses are already known to be fully reliable (i.e., that what they say is the indubitable truth) then their reports are fully believed and no other feature of the information set can influence the believability of that information. In addition, it is necessary for their reports to bear some relation to the truth (i.e. $p \neq q$). Otherwise, the fact that they concur can be nothing other than a coincidence.

Consider that three witnesses provide reports relating to the culprit of a burglary: The first testifies that the burglar spoke French, the second testifies that the burglar was wearing a French football shirt, and the third, that the burglar was waving the Tricolore flag. Figure 1.4 shows the proportions of people with such attributes in the population of possible suspects. Figure 1.4 also provides information relating to the co-occurrence of these attributes (their joint probability distribution). From Figure 1.4, we can read from the central part of the diagram that 10% of this hypothetical population are Tricolore waving, French football shirt wearing French speakers, whilst just 5% speak French without wearing either the football shirt or waving the flag (top left section). The different regions of overlap (the various 'a' regions indicated in Figures 1.4) are included in the so-called probabilistic weight vector $\langle a_0, a_1, a_2, \ldots, a_n \rangle$ (abbreviated in the following as a_i). a_0 captures the prior probability that all witnesses are correct, a_1 the probability that all but one witness is correct (regardless of which one) and so on (see Figure 1.4).

In addition to the weight vector, a_i, a further source of influence on the believability of an information set is the reliability of the witnesses. Bovens and Hartmann (2003) define their reliability parameter, r, directly from the Bayesian likelihood ratio, as $1 - q/p$. Given the assumptions above, Bovens and Hartmann (pp. 131-133) simplify Bayes' Theorem for the posterior degree of belief (P^*) in the information set ($F_1,\ldots,F_n$) having received reports ($R_1,\ldots,R_n$):

$$P^*(F_1,\ldots,F_n) = \frac{P(R_1,\ldots,R_n \mid F_1,\ldots,F_n)P(F_1,\ldots,F_n)}{P(R_1,\ldots,R_n)} \qquad \text{Equation 1.1}$$

to

$$P^*(F_1,\ldots,F_n) = \frac{a_0}{\sum_{i=0}^{n}(a_i \bar{r}^i)} \qquad \text{Equation 1.2}$$

where $\bar{r}$ ("1- r") equals q/p. This is a normative prescription for the updating of degree of belief in the truth of a conjunction of facts ($F_1 \wedge F_2 \wedge \ldots \wedge F_n$) reported by multiple witnesses. The equations take into account witness reliability, prior probability judgments, and the degree to which the reports fit together. The influence of this latter factor on the posterior degree of belief is evident from the a parameters' interaction with reliability in the denominator of Equation 1.2. Different weight is thus given to information dependent on its degree of consistency with the other information received. Equation 1.2 can be illustrated using the probabilistic information in Figure 1.4.

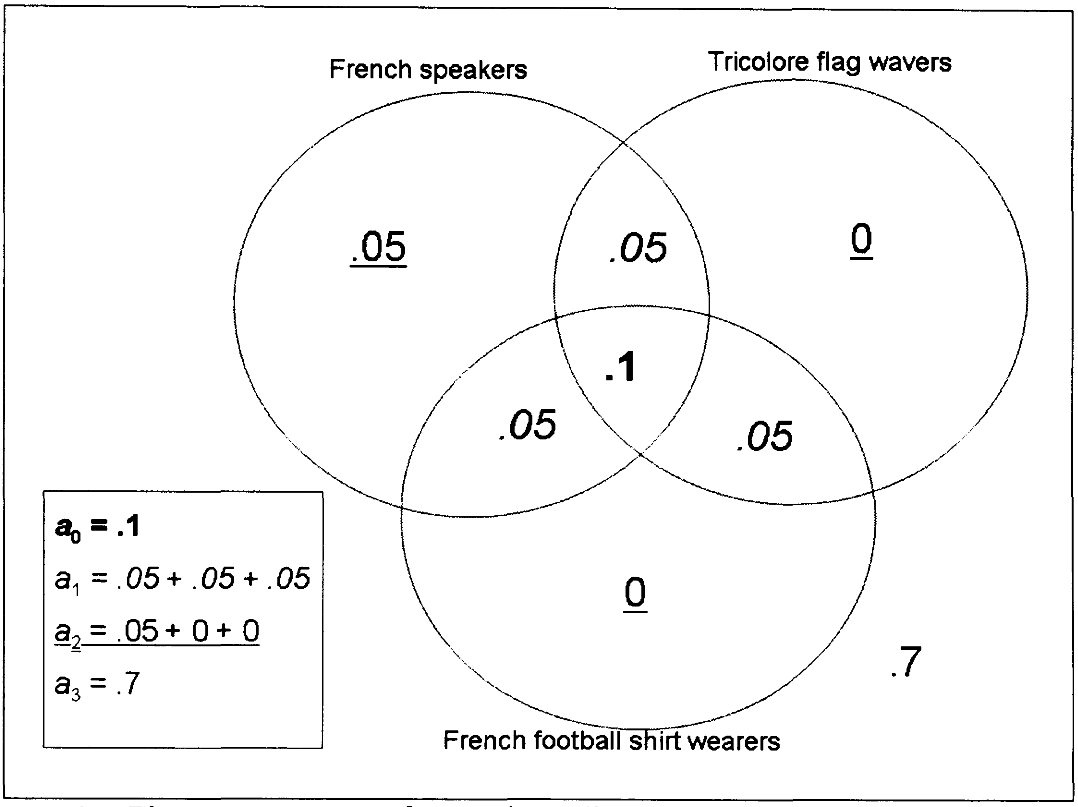


Figure 1.4. The co-occurrence of example attributes in the population of suspects (the joint probability distribution)

Implementing Equation 1.2 for Figure 1.4:

$$P^*(F_1,\ldots,F_n) = \frac{a_0}{\sum_{i=0}^{n}(a_i\bar{r}^i)}$$

$$P^*(F_1,\ldots,F_3) = \frac{.1}{(a_0\bar{r}^0) + (a_1\bar{r}^1) + (a_2\bar{r}^2) + (a_3\bar{r}^3)}$$

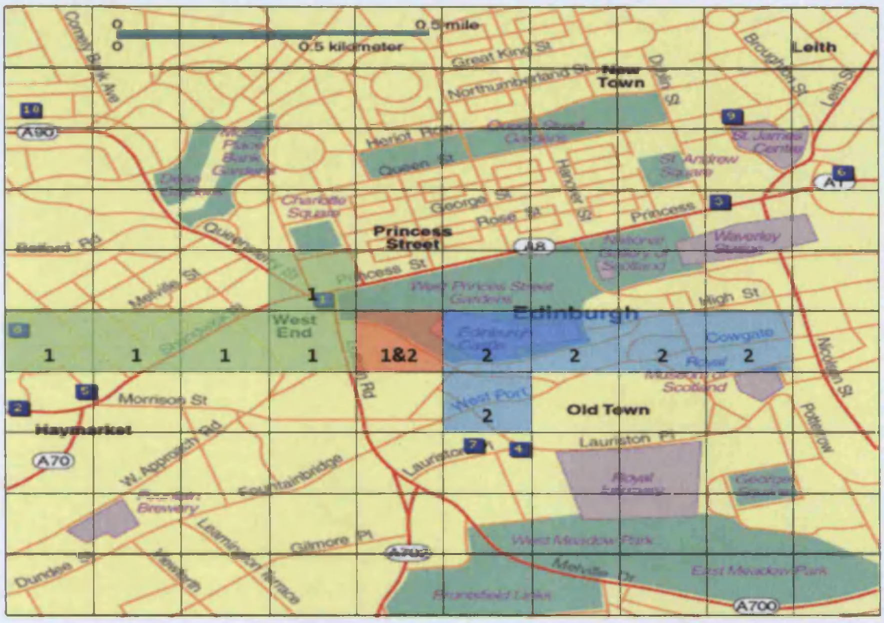
$$P^*(F_1,\ldots,F_3) = \frac{.1}{(.1\times 1) + (.15\times .25) + (.05\times .25^2) + (.7\times .25^3)} = .66$$

In Harris and Hahn (2009), we undertook an empirical investigation to determine the extent to which people's intuitions matched the prescriptions of the Bayesian formalisation. Participants read a cover story stating that a man had been murdered and police were searching for the body. The police had received tip-offs from a number of witnesses (either two or three) as to the location of the body. These witnesses were described in such a way as to meet the conditional independence and partial reliability assumptions specified above. The tip-offs were represented visually on a map of the city, which participants were free to scrutinise at their leisure. In order to define participant reliability, participants were first presented with a map illustrating that the same witnesses had provided identical reports (i.e., they agreed perfectly on the possible location of the body) and participants were told the police's posterior degree of belief in the truth of those reports. As participants were informed that the police knew the reliability of the two witnesses, this provided them with the information necessary to infer witness reliability. Subsequently, participants were asked to imagine that the same witnesses had in fact provided reports that were not identical, but which did overlap to some extent. Participants were asked to indicate how convinced, on a 21 point numerical scale, the police should be that the body is in the area shaded 'red', an area corresponding to the conjunction of the witness reports.[2] Once again, the reports were presented in a visual format, and different coloured shading was used to indicate regions of different probabilistic overlap (e.g., Figure 1.5).

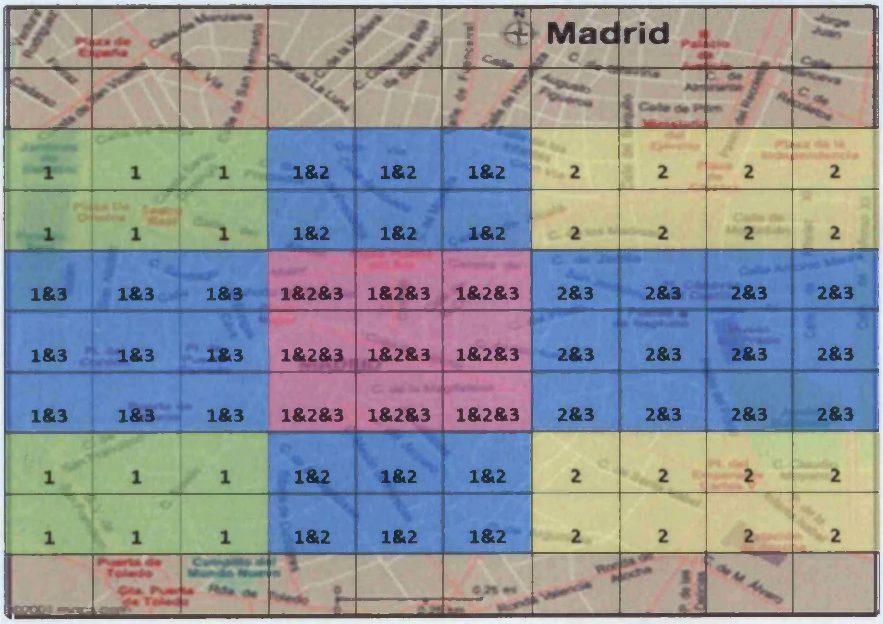
Our results showed that participants' posterior belief ratings were close approximations of the Bayesian norm, with the Bayesian model able to account for 83% of the variance in participants' ratings across the different maps used in the experiment. Moreover, participants' ratings were much better predicted by the Bayesian model than they were by a cognitively simpler averaging model, which was only able to account for 50% of the variance in participants' ratings (see Figure 1.6).

---

[2] This area was 'pink' on some maps (e.g., map L2 in Figure 1.5).



A2



L2

Figure 1.5. Two examples of the maps used in Harris and Hahn (2009). Different coloured shading, and appropriate numbers inside the shaded grid squares illustrated how many (and which) witnesses had indicated a particular grid location as a possible location of the body.

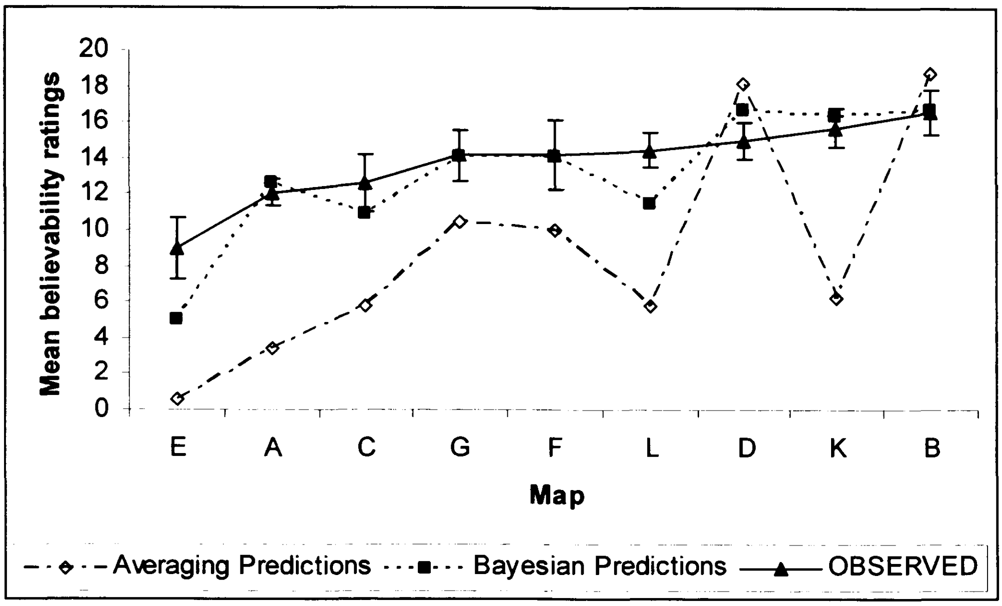


Figure 1.6. A comparison of the fits of a simple averaging model and the Bayesian model with the observed ratings in Harris and Hahn (2009). Maps are arranged in order of increasing observed ratings. Error bars are plus and minus 95% confidence intervals.

The results of this study again suggest that the Bayesian probability calculus is a normative standard for human reasoning to which people may be able to aspire. In this study, which used naturalistic materials and the sort of scenario with which people may already be familiar with from detective novels and television programmes, participants were able to process and aggregate complex information in a manner consistent with the prescriptions of Bayesian probability. It seems very unlikely that participants could have performed so well on this task were they unable to represent uncertainty quantitatively.

In conclusion, a considerable body of research (see also, e.g., Griffiths & Tenenbaum, 2006) is building up, suggesting that people are more rational processors of probabilistic information than has previously been believed.

Chapter Summary

The primary aim of this chapter was to address two fundamental critiques relating to a probabilistic approach to human reasoning: That people do not represent uncertainty quantitatively, and that past research demonstrating judgment errors in probabilistic reasoning has rendered such research pointless. These critiques were addressed with a critical review of the existing literature, and the illustration of proficient probabilistic reasoning in a recent study of our own. We conclude that the extant evidence continues to suggest Bayesian probability as a suitable normative framework for human reasoning, and one which people should strive to aspire to. Consequently, it seems an appropriate framework within which to cite our empirical investigation. In the remainder of this thesis, we turn our attention to the focal issue of whether estimates of probability are fundamentally biased by considerations of outcome utility. In real-world decision making contexts, people are typically only interested in making probability judgments in order to guide decisions. When there are decisions to be made, there are also consequences to be envisioned, and hence utility-laden outcomes to be considered. The possibility of a biasing effect of utility on probability estimates is therefore a potentially ubiquitous judgmental bias and one which would affect people throughout their daily lives in both routine decision making, and more formal decision making contexts, such as within the courtroom.

Chapter 2 - Estimating the Probability of Negative Events

Chapter Overview

As outlined above, we are interested in the question of whether probability judgments are independent of event utility. It seems particularly important to be able to provide accurate assessments of the probability with which *negative* events occur so as to guide rational choice of preventative actions. The question addressed in this chapter is whether or not our probability estimates for negative events are systematically biased by their severity. In a minimal experimental context including an unambiguous, objective representation of probability, it is found that participants judge a controllable event as more likely to occur when its utility is extremely negative than when it is more neutral. No effect is observed when the event is not controllable. This result suggests a decision-theoretic explanation based on loss function asymmetries and supports the claim that probability estimates are not intrinsically biased by utilities.

Introduction

As already introduced, SEU (Savage, 1954) posits that when selecting between alternative courses of action, individuals should select the alternative with the greatest expected benefit – that is, individuals should seek to maximise the subjective expected utility of their choice. The normative principles of SEU dictate that the assessment of an outcome's expected utility should be based on how probable that outcome is perceived to be (the expectancy component), and the subjective value attached to that outcome (the utility component). Our decision about whether or not to carry an umbrella, for example, should be based on how likely we think it is that it will rain and how bad it would be if we were to get wet, compared to the irritation of carrying an umbrella with us unnecessarily if, in fact, it did not rain. Within this framework, probabilities and utilities are assumed to be independent constructs. Intuitively, one might not expect an individual's estimate of the *chance* of rain to be based on their judgment of how *bad* it would be if they got caught without an umbrella. However, there is a long history of research querying whether probabilities and utilities are in fact assessed independently.

Estimating Probabilities

Early research on decision-making (Crandall, Solomon & Kellaway, 1955; Edwards, 1953, 1962; Irwin, 1953; Marks, 1951; Morlock & Hertz, 1964) gave some grounds for believing that people's estimates of an event's probability are influenced, to some extent, by the event's utility. However, these initial studies typically used choice paradigms, and thus assessed probability judgments only indirectly. Given that choice is governed by both probability and utility, and that both of these factors can simultaneously and subjectively vary, it is very hard to isolate either factor using such an approach. As such, the results from studies utilising decision-making paradigms could generally be explained in terms of non-linear utility functions. One such example of a phenomenon observed in decision making is that people seem to be unduly influenced by the threat of negative consequences when assessing the best course of action; in other words, 'losses loom large' (e.g. Kahneman & Tversky, 1979a). Within Prospect theory, the 'losses loom large' phenomenon is explained with reference to the non-linear utility function (typically convex for losses and concave for gains). Consequently, no interdependence between probability and utility is necessary to account for this (and related) findings in decision making or choice paradigms (see also, Edwards, 1962; Kadane & Winkler, 1988). In addition, with respect to the specific 'marked-card' paradigm used in many of these studies, Windschitl, Smith, Rose and Krizan (in press) have provided support for a 'biased-guessing' account, which does not imply a biasing effect of utility on subjective probabilities. The 'marked-card' paradigm requires participants to guess whether they will draw a marked or non-marked card from a deck of cards. The drawing of a marked card is associated with a value, which is either positive or negative. The typical finding is that there are more 'yes' responses to the question 'will you choose a marked card' when the marked card is associated with a positive outcome (e.g., Marks, 1951). Windschitl et al. argued that this does not imply that participants hold genuinely biased estimates of probability. Rather, they offer support for the contention that participants choose to provide an optimistic guess, whilst recognising that that guess is indeed optimistic. Consequently, although their *guesses* appear optimistic, participants maintain realism in their subjectively held probability estimates.

Some support for the idea that utilities might influence probability estimates emerges from research into the subjective interpretation of probability words (e.g., Weber & Hilton, 1990). The concept of probability is inherently numerical, yet we often communicate probabilities through verbal descriptors such as 'unlikely', 'possible' and 'probable'. Several empirical studies have attempted to investigate how such verbal statements are selected and interpreted. In these experiments, participants are typically instructed to respond with a single numerical probability, or a probability range, to questions like the following:

> "You have a wart removed from your hand. The doctor tells you it is *possible* it will grow back again within 3 months. What is the probability it will grow

back again within 3 months? _________" (Wallsten, Fillenbaum, & Cox, 1986, p. 574, italics added).

Weber and Hilton (1990; see also Verplanken, 1997) found that verbal probability expressions were assigned higher numerical probabilities when they referred to a severe (i.e. very negative) event as opposed to a more neutral event. In contrast, Fischer and Jungermann (1996) found that probability expressions referring to more severe events were given *lower* numerical values than those referring to more neutral events. Within this area of research there is, therefore, conflicting evidence as to exactly *how* probability and utility interact.

Crucially, however, most of these findings seem to be examples of context effects inherent in natural language use (see e.g., Grice, 2001). Context effects on the interpretation and selection of vague terms are ubiquitous. There exist, for example, studies demonstrating the effect of context on people's interpretations of verbal expressions of quantity. Borges and Sawyers (1974) and Cohen, Dearnley, and Hansel (1958) demonstrated that people's interpretation of the exact numerical meaning of quantifiers depends, in part, on the absolute magnitudes of the quantities involved. When participants were asked to select 'a few', 'some', or 'several' marbles from a tray, the absolute number of marbles selected increased linearly with the total number of marbles in the tray. The base rate of negative events also typically decreases with their severity (Weber & Hilton, 1990). Hence, corresponding linguistic conventions for vague quantifications of probability such as 'rare' or 'likely' already predict the pattern found by the majority of studies in this area – namely a decrease in the numerical values assigned to probability expressions in the context of more severe events. Moreover, evidence for such decreases has been found in both the *interpretation* (Weber & Hilton, 1990; Fischer & Jungermann, 1996) and *production* of verbal probability expressions (Merz, Druzdzel, & Mazur, 1991), suggesting a shared linguistic understanding. One cannot infer from such contextually bound variation in the numerical interpretation of verbal probability statements that people's *actual estimates* of probability are distorted by the utility of the outcome.

A different aspect of the richness of natural language is demonstrated by Bonnefon and Villejoubert (2006). They proposed that probability expressions are often used pragmatically to decrease the impact of acts that threaten the 'face' (projected sense of positive identity and public self-esteem) of an individual (Brown & Levinson, 1987; Goffman, 1967). In the following example, *possibly* is used not to communicate uncertainty, but rather out of politeness to reduce the impact on the individual's 'face': "Your bad breath is possibly the reason people shun you" (Bonnefon & Villejoubert, 2006, p. 748). In fact, in this example, *possibly* denotes a high likelihood. Within the medical domain, the act of informing a patient that they might develop a certain condition is a face threatening act whose threat increases as condition severity increases. Bonnefon and Villejoubert (2006, p. 748) therefore proposed that "the more severe the patient's condition, the more likely a probability expression will be interpreted as a face-management device", rather than as an expression of likelihood. With such an interpretation, the probability associated with the expression is increased. Crucially, however, both speaker and listener will be aware of the discrepancy between the underlying and expressed probabilities. Resulting 'biases' are therefore not genuine biases, as the listener is essentially correcting for the face-saving act of the speaker. Rather, the effect is simply a product of conversational convention and further demonstrates the pragmatic richness of natural language in conversational contexts.

Staying with the medical domain, Wallsten (1981) reanalysed data presented in Fryback and Thornbury (1976, as cited in Wallsten, 1981) and found that human diagnosticians (radiographers) overrated the probability of a "space-occupying lesion" (Wallsten, 1981, p. 147) being a malignant tumour as opposed to a benign cyst or simply a normal variation. Such a result may be seen as an instance of event severity (clearly a malignant tumour is a very severe event) biasing estimated probability. However, as recognised in Levy and Hershey (2008), event severity in these data is confounded with real-world prevalence as malignant tumours are relatively rare events. Thus, there is nothing to suggest that this result is anything more than a further instance of the well established finding that low frequencies are over-estimated (e.g., Attneave, 1953; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978). The same reasoning can explain the results from weather forecasting experiments in which forecasters have been shown to overestimate the probability of occurrence of severe weather events (Murphy & Daan, 1984; Murphy & Winkler, 1982).

A further source that might suggest that people's utilities systematically bias their probability judgments is research into estimates of personal risk for negative life events. A sizeable literature reports that people are prone to 'unrealistic optimism' (e.g., Kirscht, Haefner, Kegeles & Rosenstock, 1966; Weinstein, 1980, 1982, 1984). Individuals seemingly regard their own personal risk to be less than that of the average person, displaying a kind of 'invulnerability bias'. On the assumption that one's own illnesses are subjectively more negative events than another's illnesses (especially if the 'other' is simply the 'average person') this would suggest that the increased severity of an event leads to a (protective) depression of estimated probability. As the range of negative life events included in these studies typically varies, examinations of correlations between the degree of unrealistic optimism and event severity can test this interpretation more directly. It turns out that there is no evidence for such a relationship once other relevant variables (e.g. prior experience) are controlled for (Eiser, Eiser & Pauwels, 1993; Heine & Lehman, 1995; van der Velde, Hooykas & van der Pligt, 1992; van der Velde, van der Pligt & Hooykas, 1994; Weinstein, 1982, 1987, Weinstein, Sandman & Roberts, 1990)[3].

The most direct evidence for the independence of probability and utility in the negative domain, to date, comes from a study by Pruitt and Hoge (1965); however, their study suffers from other methodological difficulties. Participants were presented with a sequence of 24 flashes, each from one of two lights. Participants were tested on an unseen 25th flash. Participants were asked to rate the probability of a 'Light A' (as opposed to 'Light B') flash. Participants were also told that they would either lose or gain money if the flash came from 'Light A' on this trial, with the value of a 'Light A' flash ranging from –50 cents to +50 cents. Pruitt and Hoge observed a positive linear relationship between the utility of the outcome and participants' probability ratings. It is possible, however, that the pragmatics of the situation (i.e., the fact that they were taking part in an experiment) led participants to believe that it was unlikely that they would emerge from the study having to pay money to the experimenter. As such, participants may have reasonably assumed that rewarded outcomes would occur more

---

[3] [1]The only exception is an experimental study by Taylor and Shepperd (1998) who led participants to believe they were being tested for a medical condition with either severe or non-severe consequences. An effect of severity was found such that when participants were told that test results were imminent, optimism was *eliminated* in the severe condition. This effect seems attributable to a desire not to 'jinx' things. No effect of severity was found in participants who did not expect feedback.

frequently than penalised outcomes. This hypothesis would predict the same linear trend observed in their data. In summary, there is presently no direct evidence for an effect of negativity on probability estimates.

There is, however, also a literature investigating whether outcome utility biases estimates of probability in the case of *positive* outcomes. Indeed, there have been more (and more direct) tests of interdependence between utility and probability in the positive domain than in the negative domain (see Krizan & Windschitl, 2007, for a review).

Price (2000) divided his participants into two teams and required them to estimate the probability that a member of Team A would throw a dart closer to the bullseye than a member of Team B. He found that members of Team A gave significantly higher estimates than members of Team B. This, coupled with a manipulation check that participants desired their own team members to win the contest, was taken as evidence for a wishful thinking effect. However, within the social psychological literature on groups there is an abundance of studies reporting such effects in contexts of intergroup competition (e.g. Blake & Mouton, 1961; Jourden & Heath, 1996; Sherif & Sherif, 1956), and these are well-explained by motivational and cognitive factors *other* than wishful thinking, such as the protection of the group's positive self-image (Jourden & Heath, 1996). Consequently, Price's study cannot be considered to be a satisfactory test of a general wishful thinking bias.

In five empirical studies, Gendolla (1997) demonstrated that failures on important exams were rated as more surprising than failures on unimportant exams. Gendolla concluded that this effect was a result of increased outcome desirability increasing outcome expectancy, which in turn made failure more surprising. Whilst he did produce evidence to support these relationships, this evidence does not support the contention that increased positive utility routinely biases expectancies. In his Study 3, Gendolla observed a "significant main effect of outcome importance on rated effort expenditure…this indicated higher effort ratings in the important conditions than the unimportant ones" (Gendolla, 1997, p.179). The resulting (marginal) main effect of outcome importance on expectancy can therefore be attributed to perceived greater effort, which would reasonably be associated with higher expectations of success, rather than a direct biasing effect of utility on expectancy.

The most extensive, direct, test of the relationship between positive utility and subjective probability estimates (Bar-Hillel & Budescu, 1995) found no evidence for an effect of positive outcome utility on probability estimates. Bar-Hillel and Budescu observed a wishful thinking effect (such that good outcomes were rated as more probable than neutral outcomes) in only 30% of approximately 1300 probability judgments, leading them to title their paper, "The elusive wishful thinking effect." They also highlighted that previous observations of the wishful thinking effect *outside* controlled laboratory conditions (e.g. Babad & Katz, 1991) can be well-explained as "an unbiased evaluation of a biased body of evidence" (Bar-Hillel & Budescu, 1995, p. 100, see also Gordon, Franklin, & Beck, 2005; Morlock, 1967). Bar-Hillel, Budescu, and Amar (2008), for example, found potential evidence of wishful thinking in the prediction of results in the 2002 and 2006 football World Cups. However, a further experiment showed that these results were more parsimoniously explained as resulting from a *salience* effect rather than a "magical wishful thinking effect" (Bar-Hillel et al., 2008, p. 282), that is, from a shift in focus that biases information accumulation rather than an effect of desirability *per se*. Moreover, tests of the wishful thinking phenomenon have reported conflicting results with some finding evidence for wishful thinking (Price, 2000; Pruitt & Hoge, 1965), others finding the opposite, a pessimism bias (Dai, Wertenbroch, & Brendl, 2008; Mandel, 2008), while still others found little effect of outcome utility at all (Bar-Hillel & Budescu, 1995; Erev & Cohen, 1990). Consequently, as Krizan and Windschitl (2007) conclude in their extensive review of the literature on biasing effects of positive outcomes, there is little evidence that desirability directly biases estimates of probability.

In summary, despite a long history of research potentially suggesting an influence of outcome utility on probability judgments, this issue remains unsettled. Moreover, the lack of any direct tests in the negative domain means that the issue remains entirely open for probability estimates of negative events.

Overview

In the following, we describe seven studies testing the proposition that the severity of negative events directly influences their perceived probability[4]. Study 1 provides a demonstration that severe (extremely negative) events are assigned higher probability estimates than neutral events, a finding that is replicated twice (Studies 1-3). By contrast, Studies 4 and 5 fail to replicate this effect in a different scenario. Finally, Studies 6 and 7 support an explanation for these differences in terms of loss asymmetry.

---

[4] Studies 1, 2, 3, 6 & 7 were published in Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition, 110,* 51-64. The idea for Study 1 was conceived in collaboration with Adam Corner and Ulrike Hahn. The cover story for Study 1 was conceived by Adam Corner. The JAVA program that generated the different matrices was written by Ulrike Hahn

A Direct Test of Severity Influence

Are severe outcomes perceived as more probable, or less probable, than neutral outcomes? In attempting to answer this fundamental question, it seems necessary to dispose of as many potential confounds as possible, and avoid the ambiguities that trouble the interpretation of verbal probability expressions. We therefore wanted a task in which participants provided numerical estimates (see also, Pruitt & Hoge, 1965, on the desirability of numerical estimates). The main difficulty in choosing appropriate materials for such estimates is that, as noted, severity and probability are typically confounded in the real world (see also, e.g., Weber & Hilton, 1990), such that 'really bad' events are less frequent than 'moderately bad' or neutral ones. At the same time, certain severe real-world events (e.g., accidents and fires) are judged as more prevalent than they truly are as a result of, for example, media coverage (e.g., Slovic, Fischhoff, & Lichtenstein, 1982). This is typically construed as an example of the availability heuristic (e.g., Tversky & Kahneman, 1973), which could potentially confound the results of any experiment eliciting probability estimates of real-world events. Simply comparing estimates across events of different severity would consequently be insufficient as a test for bias. What is required is an objective measure of the probabilities involved. Since such measures are difficult to obtain, and because differences in knowledge between people could furthermore give rise to rational deviations from these objective probabilities, we developed fictitious scenarios. Crucial to our experimental design is the fact that participants are supplied with an objective basis for their subjective estimates and that this objective basis is *identical* across the severity manipulations. Any systematic difference that arises in participants' estimates of probability across conditions is consequently directly attributable to the manipulation of severity.

Study 1

The purpose of Study 1 was to provide a direct demonstration of the effect of outcome severity on estimates of outcome probability using a paradigm in which these estimates are anchored to an objective probability to which all participants have equal access. Specifically, the relevant probabilities were provided in a visual display. The use of visual displays as a means of presenting probabilistic information to participants has considerable precedent (e.g., Bar-Hillel & Budescu, 1995; Cohen & Wallsten, 1991; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986), but has not been used to directly investigate the relationship between the severity of negative events and their probability. Participants saw cell matrices in which different coloured cells represented different outcomes. To make the interpretation of these matrices more natural, the cover story was chosen such that the spatial arrangement of the cells had a straightforward real-world correspondence. Specifically, the cells were presented as a graphical representation of a large apple orchard. Yellow cells corresponded to apple trees bearing 'bad' apples; grey cells corresponded to 'good' apple trees. The matrix was made sufficiently large that counting the number of cells would have been extremely time consuming, thus ensuring that participants would be giving *estimates* even though they were being presented with an objective probability. The cover story associated with the display varied the significance of the 'bad' apples such that they were either fatally poisonous (the severe outcome) or tasted unpleasant (the neutral outcome). Participants were allocated to either the severe or the neutral cover story and asked to provide a probability estimate for the event in question. Crucially, however, all participants saw exactly the same visual displays. The paradigm therefore provided a direct test of the hypothesis that outcome utility may alter the subjective probability of an event's occurrence.

Method

Participants

100 participants took part in Study 1. The study was conducted remotely using an internet host, iPsychExpts.com (Brand, 2005). 55 female and 45 male participants with a mean age of 30 completed the study, in an average time of 2.54 minutes. 50 participants provided probability estimates of severe outcomes, and 50 provided probability estimates of neutral outcomes.

Design

Study 1 was designed to test the hypothesis that probability estimates of severe outcomes differ from probability estimates of neutral outcomes. This hypothesis was tested using visual response matrices containing varying proportions of grey and yellow cells. Outcome severity was manipulated between participants, such that the yellow cells in the display matrices corresponded to outcomes of either extremely negative or neutral utility. The number of yellow cells in the display matrices was manipulated within participants, such that everyone gave three estimates of probability (low/medium/high outcome probability).

Materials

A visual display containing 2236 cells with a random distribution of grey and yellow cells was constructed with a simple JAVA program designed specifically for the study (see Figure 2.1). In the low probability condition of the study, the randomly distributed yellow cells were constrained to occupy less than 5% of the display. In the medium probability condition, 50% of the cells in the display were yellow. In the high probability condition, more than 90% of the cells in the display were yellow. All participants viewed the same three matrices.

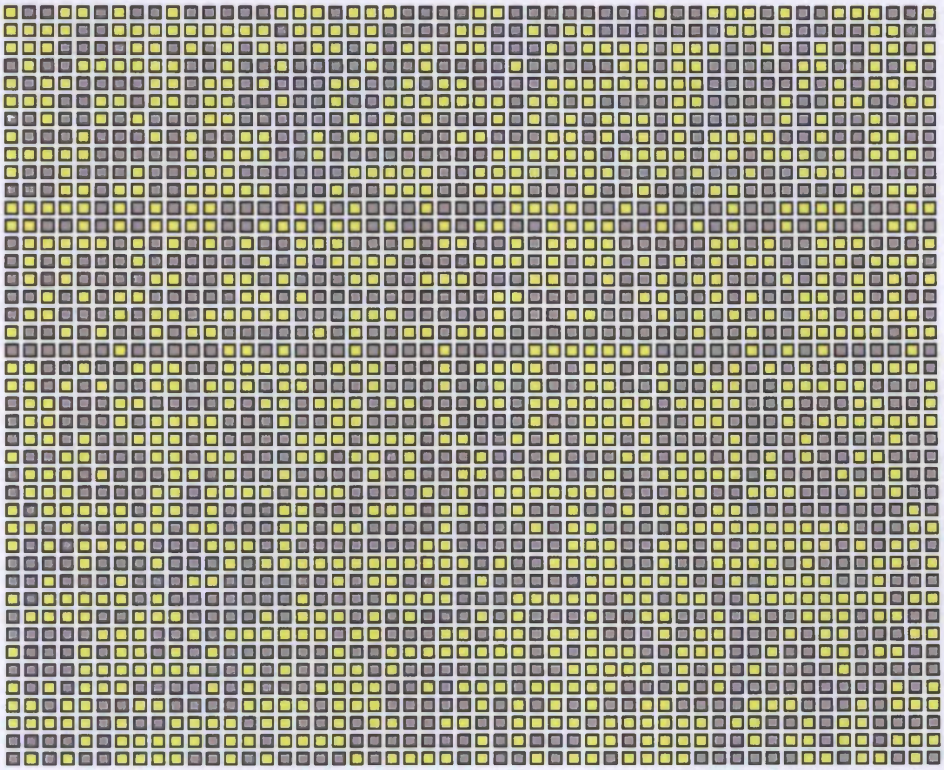


Figure 2.1. An example of a cell matrix (from the medium probability level)

Depending on the outcome severity condition participants were randomly assigned to, they read one of the following cover stories:

Severe outcome:

'A farmer has just bought an orchard that contains over 1000 apple trees. The picture below shows the layout of the orchard, with each coloured circle representing an apple tree. Trees that are coloured GREY bear fruit that is tasty, and delicious to eat. Trees that are coloured YELLOW have been sprayed with a particularly potent type of pesticide, and bear fruit that is **fatally poisonous** to humans. The farmer's young daughter is always playing in the orchard, and despite her father's warnings, she often picks apples to eat from the trees in the orchard.

Unfortunately, however, there is no way of knowing whether an apple tree bears edible or inedible fruit without trying an apple from the tree (the colours grey and yellow simply represent the different types of apple). The safety of his daughter is extremely important to the farmer, who is very concerned that she might eat a poisonous apple by mistake.'

Neutral outcome:

'A farmer has just bought an orchard that contains over 1000 apple trees. The picture below shows the layout of the orchard, with each coloured circle representing an apple tree. Trees that are coloured GREY bear fruit that is tasty, and delicious to eat. Trees that are coloured YELLOW bear fruit that is sour, and unsuitable for eating. Unfortunately, however, there is no way of knowing whether an apple tree bears edible or inedible fruit without trying an apple from the tree (the colours grey and yellow simply represent the different types of apple).'

In the severe outcome condition, participants were asked by the farmer to "estimate the chance of his daughter choosing an apple from a tree that bears fatally poisonous fruit (YELLOW), if she were to randomly pick an apple from any of the trees in the orchard". In the neutral outcome condition, participants were asked to estimate the chance of the daughter picking a sour and inedible apple.

Probability estimates were made on a 21-point numerical scale from 0% (Absolutely Impossible) to 100% (Absolutely Certain). Participants responded by clicking on a radio button.

Procedure

The study was run through ipsychexpts.com. Having chosen to participate in the study, participants first viewed the consent screen, which was followed by a screen containing the general instructions for the study. The next three screens contained the experimental materials. Having completed the study, participants were required to enter their age and sex before finally being presented with a debriefing screen.

Results

As the study was conducted remotely using an internet host, we followed Birnbaum (2004b) and performed several basic checks prior to data analysis. All participants under the age of 18 were excluded (in line with departmental ethical guidelines), data from the same Internet Protocol (IP) address were excluded (in order to guard against multiple entries from the same individual), and participants with demographic details that aroused suspicion of fabrication (an age entry of 90 or over) were eliminated from subsequent analysis. In addition, we excluded participants who had obviously failed to understand the instructions in that they had provided estimates of the three, clearly distinct, levels of probability that deviated from their basic rank order. Participants who failed to complete such a basic task in less than 15 minutes were also excluded, to ensure that people were estimating, and not counting the cells. Following these exclusions, 73 participants were included in the analysis, 40 in the severe outcome condition, and 33 in the neutral outcome condition.

A preliminary analysis was conducted to establish that the probability manipulation (i.e., the proportion of yellow cells in the display matrices) had in fact produced different probability estimates. Collapsing across both outcome severity conditions, a significant main effect of probability in the expected direction was observed, $F(2, 142) = 1149.0, p<.001, MSE = 87.9$. More importantly, Figure 2.2 displays these probability estimates, but split by outcome severity. At each level of the probability manipulation, the estimated proportion of yellow cells in the display matrices was higher in the severe outcome condition, producing an overall main effect of outcome severity, $F(1, 71) = 7.36$, $p<.01$, $MSE = 174.60$, $eta_p^2 = .09$. There was no interaction between probability and severity, $F(2, 142) = .75$, $p>.05$, $MSE = 87.90$.

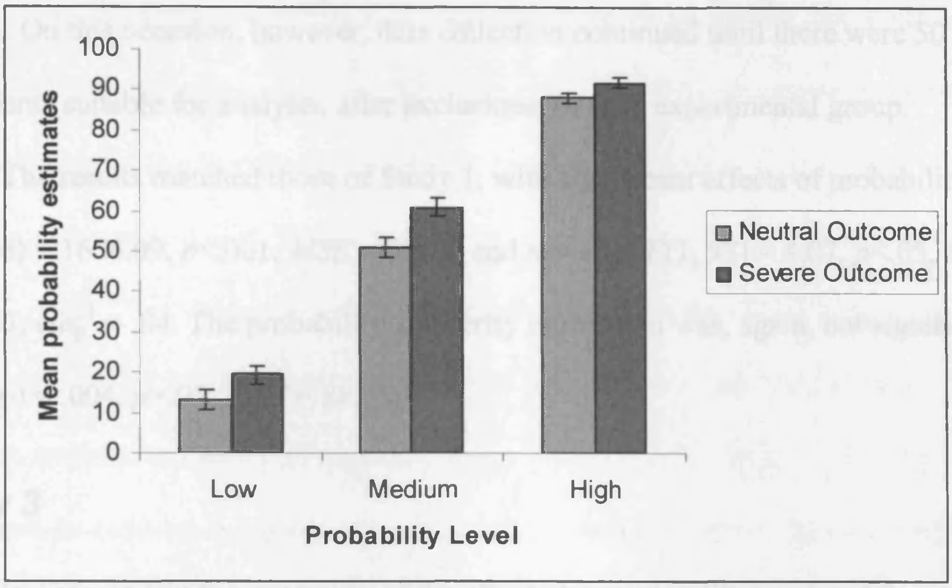


Figure 2.2. The effect of outcome utility on probability judgments. Error bars are plus and minus 1 standard error.

Study 2

Finding a statistically significant effect of outcome severity on judgments of probability in such a minimal paradigm with a patently fictitious story of no personal relevance to participants was sufficiently surprising that we sought to replicate this result. Study 2 was a direct replication of Study 1 with a different set of participants.

Method

Participants

52 female and 48 male participants with a mean age of 26 completed the study, in an average time of 2.56 minutes.

Design, materials and procedure

Study 2 was an exact methodological replication of Study 1.

Results

The same basic checks were undertaken prior to analysis as were performed in Study 1. On this occasion, however, data collection continued until there were 50 participants suitable for analysis, after exclusions, in each experimental group.

The results matched those of Study 1, with significant effects of probability, $F(2, 196) = 1656.09$, $p<.001$, $MSE = 88.75$, and severity, $F(1, 98) = 4.07$, $p<.05$, $MSE = 127.67$, $eta_p^2 = .04$. The probability x severity interaction was, again, not significant, $F(2, 196) = .004$, $p>.05$, $MSE = 88.75$.

Study 3

In order to test the generality of the effect observed in Studies 1 and 2, we repeated the study with new matrices using different colours and different probability levels.

Method

Participants

An internet sample of 89 males and 182 females, aged between 19 and 64 (median = 33 years) completed Study 3, in an average time of 2.91 minutes.

Design

The same mixed 3x2 design was employed as in Studies 1 and 2.

Materials

Three blue and black matrices were constructed using the JAVA program. In this study, the colour blue was used to represent 'good' apples, whilst black was used to represent 'bad' apples. The percentage of black cells in these matrices was approximately 20%, 50% and 80% for the three probability levels. As in Study 1, all participants viewed the same 3 matrices.

The same basic orchard premise was used in the cover stories, but some minor changes were made to the text to maximise the similarity between severe and neutral conditions: In the neutral outcome condition, the 'bad' trees had 'been sprayed with a contaminated pesticide that, though not dangerous to humans, leaves the fruit tasting **horribly sour**.' This change ensured that in both conditions the apples were sprayed with a pesticide (which was also 'contaminated' rather than a 'particularly potent type of' in the severe condition) and that the effect of the pesticide was in bold font in both conditions. In addition, a sentence was added to the end of the cover story stating that 'The happiness of his daughter is important to the farmer, who is very concerned that she might eat a sour apple by mistake.' The final modification made to the cover story was that in the severe condition the words 'edible or inedible' were replaced with the words 'delicious or poisonous'.

Procedure

The procedure was identical to that in Study 1.

Results and Discussion

Participants were excluded prior to data analysis using the same criteria as in Studies 1 and 2. Following participant exclusions, there were 75 males and 152 females, with 112 participants in the neutral outcome condition and 115 in the severe outcome condition.

The results for these participants are summarised in Figure 2.3. Again, there was a main effect of probability, $F(1.9, 389.6) = 1819.12, p<.001, MSE = 94.81$, and a main effect of severity, $F(1, 206) = 4.13, p<.05, MSE = 403.89, eta_p^2 = .02$. The probability x severity interaction was, once again, not significant, $F(1.9, 389.6) = 2.95$, $p>.05$, $MSE = 94.81$ (Greenhouse-Geisser corrections applied). The results replicate exactly the findings of Studies 1 and 2 despite changes to colours and probabilities associated with the matrices, further suggesting that this is a robust effect, despite the minimal nature of this paradigm.

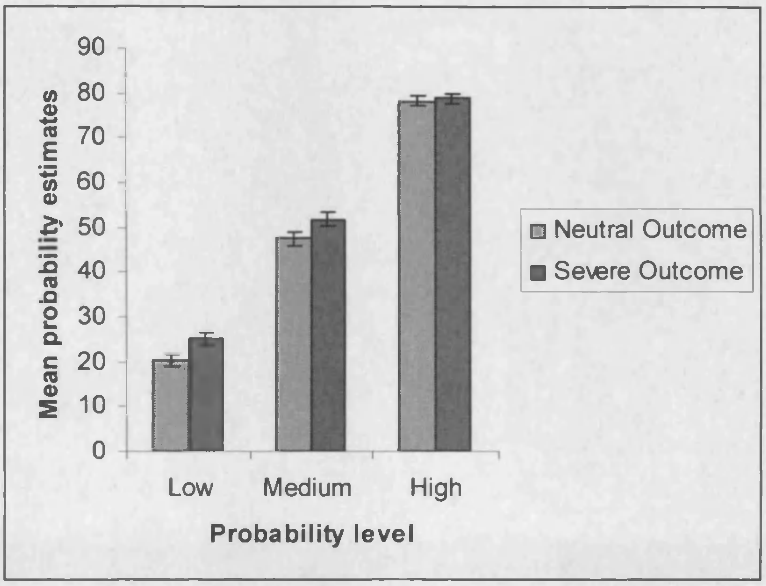


Figure 2.3. Mean probability estimates made in Study 3. Error bars are plus and minus 1 standard error.

Study 4

So far, the effect has only been demonstrated with a single paradigm. Consequently, we designed both a new cover story and a different visual representation of probability for Study 4. The cover story introduced a bomber plane that had to drop the bomb it was carrying. The bomb would either fall safely in a river, or would explode on the land. From a visual representation of the landscape into which the bomb could fall, participants had to judge the probability that the bomb would fall on the land. In the severe condition the land was a densely populated city, whilst in the neutral condition it was uninhabited farmland.

Method

Participants

An internet sample of 44 males and 56 females aged between 18 and 60 (median = 27 years) completed the study, in an average time of 2.52 minutes.

Design

The same mixed 3x2 (probability x outcome severity) design was used as in Studies 1-3.

Materials and procedure

Three visual displays were created in Microsoft's "Paint" application. These visual displays consisted of a grey rectangle with a wavy blue line (representing a river) crossing the diagonal from the bottom left corner to the top right corner (see Figure 2.4). A circle, incorporating an area of the rectangle containing both grey and blue was superimposed over the centre of the display. The three visual displays differed only in the thickness of the blue line representing the river. The low probability condition had the broadest river, whilst the high probability condition had the narrowest river.

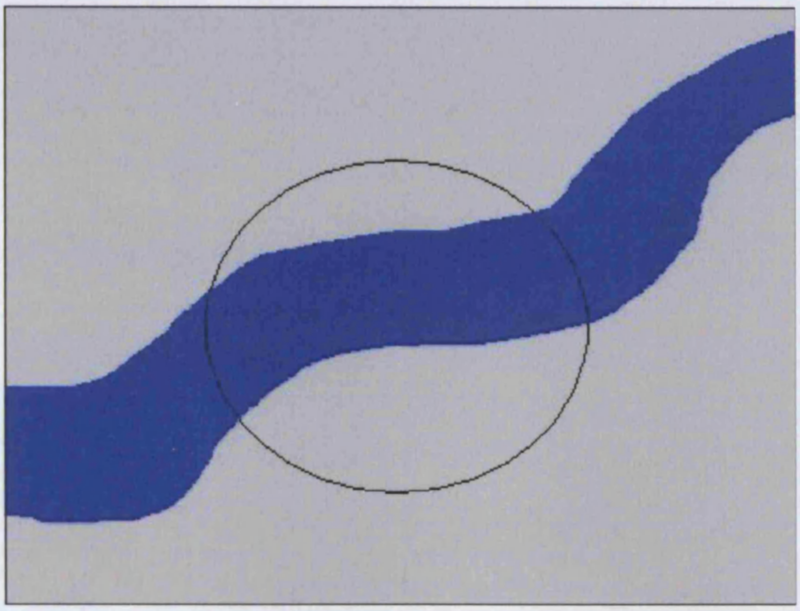


Figure 2.4. An example visual stimulus (from the medium probability level)

Depending on the outcome severity condition participants were randomly assigned to, they read one of the following cover stories:

Severe outcome:

'A British bomber plane is running low on fuel. The only way the pilot can prevent it from crashing is to release the bomb it is carrying over the city landscape below. The blue river runs through the heart of a densely populated city (grey). If the bomb lands in the river then it will not explode and it can be safely defused by experts. If the bomb lands in the city then it will explode, killing all the inhabitants of the city.'

Neutral outcome:

'A British bomber plane is running low on fuel. The only way the pilot can prevent it from crashing is to release the bomb it is carrying over the rural landscape below. The blue river runs through the heart of an expanse of uninhabited fields (grey). If the bomb lands in the river then it will not explode and it can be safely defused by experts. If the bomb lands in the fields then it will explode, but safely away from any humans.'

Both groups of participants read the remainder of the cover story:

'The pilot aims to drop the bomb within the circle depicted on the map. The pilot can ensure that the bomb lands within this circle, but owing to elements such as wind and thermal currents the bomb could land absolutely anywhere within this circle.'

Participants in the severe condition were then asked: 'By looking at the diagram below, what do you think is the probability that the bomb will fall on the city, killing all the inhabitants?', whilst those in the neutral condition were asked: 'By looking at the diagram below, what do you think is the probability that the bomb will fall in the fields and explode?'

All other aspects of the study were identical to the preceding studies.

Results

Prior to data analysis, we excluded participants according to the same criteria as in the previous studies. Following participant exclusions, there remained 61 participants (28 in the neutral outcome condition and 33 in the severe outcome condition).

Mean responses across all conditions were calculated and are reported in Figure 2.5. The expected effect of probability was observed, $F(2, 118) = 492.54$, $p<.001$, $MSE = 63.53$. On this occasion, however, there was no effect of severity, $F(1,$

59) = .07, p>.05, MSE = 338.73, eta_p^2 = .001, nor was there an interaction between probability and severity, $F(2, 118) = .111$, p>.05, MSE = 63.53.

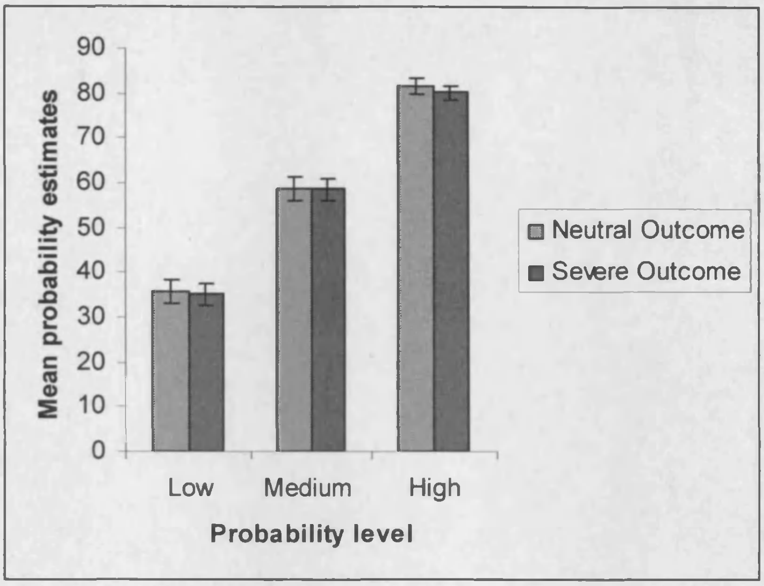


Figure 2.5. The effects of outcome utility on probability estimates. Error bars are plus and minus 1 standard error.

Discussion

Study 4 failed to replicate the effect of outcome severity observed in Studies 1 to 3. There were, however, perceptual differences between these studies that may have contributed to the different patterns of results observed. The discrete nature of the cells in the probability matrix used in the first three studies might give rise to a frequency-based representation of the information. In other contexts there have been systematic differences between reasoning using probabilities in frequency and non-frequency formats (e.g. Brase, 2008; see Gigerenzer & Hoffrage, 1995, for a review) suggesting that this might be a critical difference between the scenarios.

Study 5

Study 5 was therefore designed to reduce the perceptual differences between the two paradigms, whilst still keeping them distinct. Specifically, a grid was superimposed over the visual scene to approximate the frequency-based representation of probabilities (see Figure 2.6) in Studies 1-3. The cover story was subsequently modified accordingly.

Method

Participants

57 males and 108 females, aged between 18 and 72 (median = 35 years), completed this study, in an average time of 2.24 minutes, in return for 50 ipoints. For this study we used ipoints.co.uk™ to recruit participants, though the study continued to be run through ipsychexpts.com. Ipoints.co.uk sent an email to a subset of its members advertising the study and informing them that they would receive 50 ipoints for completing it. Ipoints can be exchanged for goods and services via ipoints.co.uk, and one ipoint has a cash value of £0.01. Thus, participants were paid the equivalent of £0.50 for participating in this study.

Design

This study employed the same 3x2 (probability x outcome severity) mixed design as the preceding studies.

Materials and procedure

The cover story from Study 4 was adapted. The cover stories stated that there was a fault with the bomb jettison equipment. Participants were informed that the display depicted the bomb sights of an ageing plane. They were then told: 'the accuracy of the sights is limited and, as such, the bomb could land in absolutely any of the grid squares within this circle.'

All other aspects of the study were identical to Study 4.

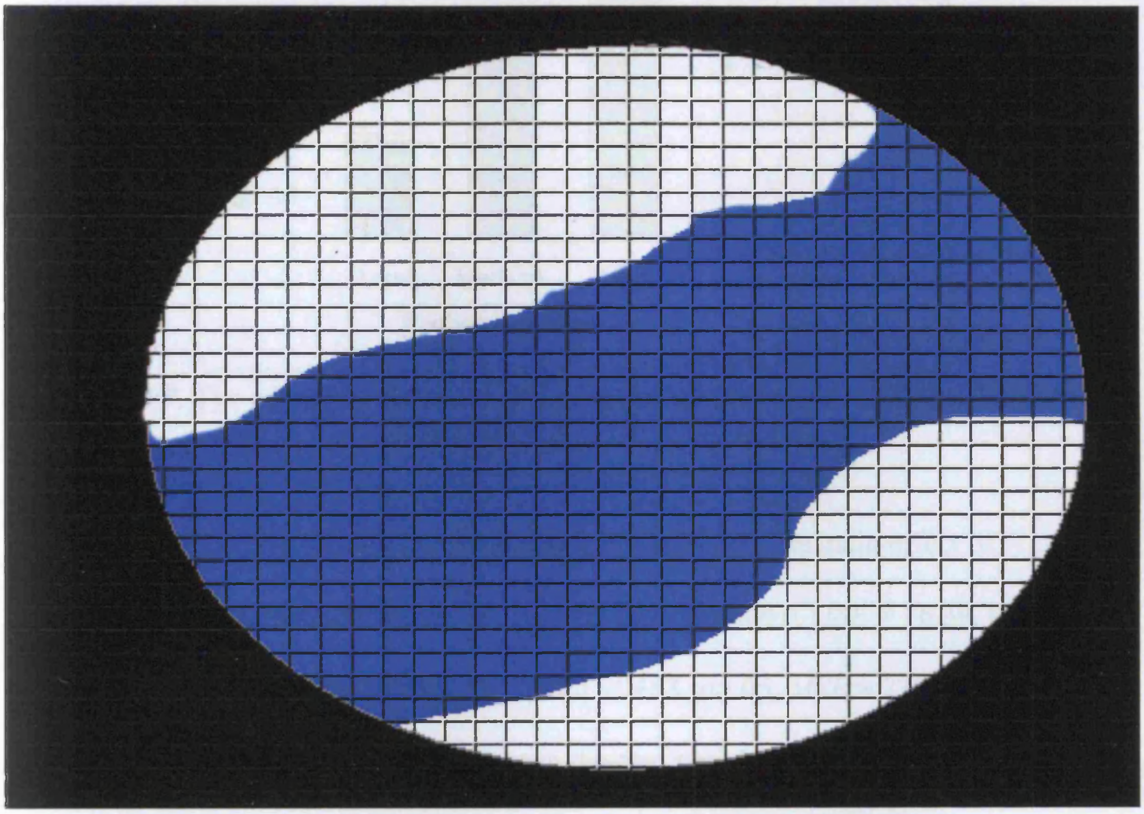


Figure 2.6. An example of the visual stimulus in Study 5 (from the medium probability level).

Results

Participants were excluded from analysis using the same criteria as before. Following exclusions, 40 males and 70 females (55 in each condition), with a median age of 35.5 years, were included in the data analysis.

Mean probability estimates were calculated for the two experimental groups and are summarised in Figure 2.7.

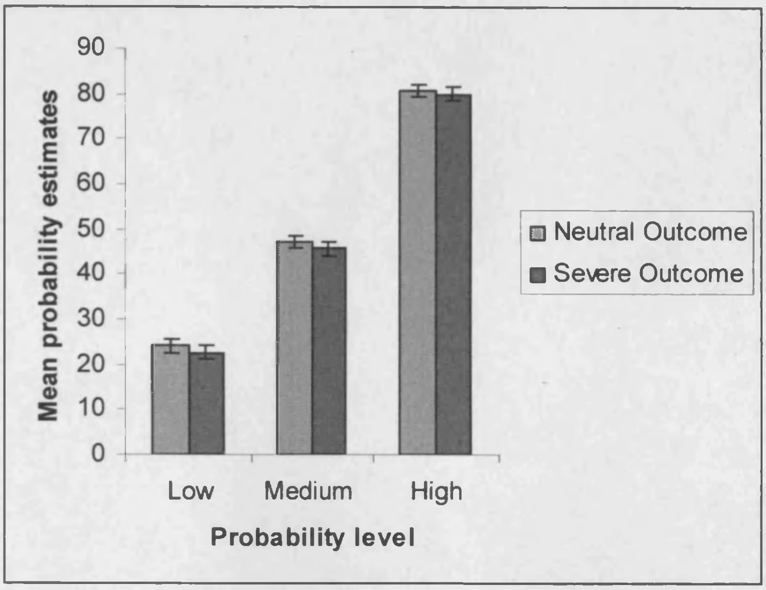


Figure 2.7. The effects of outcome utility on probability estimates in Study 5. Error bars are plus and minus 1 standard error.

Data analysis yielded the same results as Study 4. The main effect of probability was significant, $F(1.8, 193.1) = 1319.50$, $p<.001$, $MSE = 76.39$, but there was no effect of outcome severity, $F(1, 108) = .483$, $p>.05$, $MSE = 238.50$, $eta_p^2 = .004$, nor was there an interaction between the two variables, $F(1.8, 193.1) = .09$, $p>.05$, $MSE = 76.39$ (Greenhouse-Geisser corrections applied).

Discussion

Study 5 replicated the null result observed in Study 4. Hence it is clear that it was not simply the presentation of the probabilities in a format amenable to a frequency-based representation of probability that led to the effect being observed in the orchards paradigm (Studies 1-3), but not with the bomber paradigm.

What, then, could be the critical difference between the two scenarios? One possibility is that the outcome is no longer under human control in the bomber paradigm, but potentially still is in the orchards paradigm. Specifically, the critical decision point has already passed in the bomber paradigm (the plane has already flown beyond its fuel capacity and its bomb will necessarily hit the area in question, either because it is 'ditched' to save the plane or because it goes down with the plane); by contrast, participants in the orchards paradigm may believe that there is still a critical decision that could be made (e.g., the farmer could prevent his daughter from entering the orchard).

The potential relevance of such a difference is suggested by a loss asymmetry account. This account is based on the idea that there are two types of errors that can be made in estimating probabilities, overestimates and underestimates. Crucially, the costs associated with these different errors are often not equivalent (see e.g., Weber, 1994). As an example, consider the possibility of contracting meningitis if a colleague has been admitted to hospital with the disease. If a person underestimates the possibility that they will catch meningitis then the consequences are potentially very negative; that is, the individual might not be prepared when they experience the initial symptoms and therefore they might not seek medical advice immediately and the disease will not be treated early enough. The costs associated with overestimating the possibility of contracting meningitis are not as negative: increased worry and some time to have one's health checked. If, however, someone is considering the possibility that they will contract a cold, the asymmetry in the loss function is greatly reduced. The theory therefore predicts that probability estimates will be biased in order to reduce the likelihood of making the more costly error.

With reference to the present studies, the difference in the perceived human controllability of the different outcomes implies differences in the asymmetry of the loss functions associated with the severe and neutral event in each paradigm. Specifically, as there is no action that can be taken to alter the chance outcome in the bomber paradigm, there can be no asymmetry in the loss function – whether you overestimate, underestimate, or correctly estimate the probability of disaster makes no difference to whether or not the disaster will actually occur. By contrast, if participants perceived the orchards paradigm as a situation in which the farmer could potentially prevent his daughter from entering the orchard then a loss asymmetry exists in the 'severe' condition of this task. The costs associated with an underestimate of the probability of the farmer's daughter picking a fatally poisonous apple are clearly greater than those associated with an overestimate, as an underestimate might lead the farmer not to take the necessary steps to help prevent his daughter from entering the orchard and picking apples. Estimates are thus inflated in the severe outcome condition to reduce the likelihood of a costly underestimate.

Study 6

The purpose of this study was to test an asymmetric loss function based explanation for the different results observed using the two scenarios. We used the severe events from Studies 1 and 2 (where we had found an effect) and introduced a manipulation of outcome control. If the asymmetric loss function account applies, then higher probability estimates should be seen under conditions of control, than under conditions of no control.

Method

Participants

An internet sample of 81 males and 166 females, aged between 17 and 63 years (median = 25 years) completed the study, in an average time of 3.06 minutes.

Design

A 3x2 (probability x controllability) mixed design was employed with probability manipulated within participants and controllability manipulated between participants. Each participant therefore made three probability judgments (one at each probability level). The order in which participants made these three probability judgments was randomised.

Materials and procedure

This study used the same materials as in the severe outcome condition of Studies 1 and 2. The controllability manipulation was based on the following, additional, text:

No-control:

'As the safety of his daughter is of great importance to the farmer he has tried many different solutions to try and protect his daughter. He has however been unable to keep his free-spirited daughter from playing in the orchard.

There remain no feasible steps that the farmer can possibly take to remove the chance that his daughter might eat a poisonous apple. Please estimate the chance of his daughter choosing an apple from a tree that bears **fatally poisonous** fruit, if she were to randomly pick an apple from any of the trees in the orchard.'

High-control:

'The safety of his daughter is extremely important to the farmer, who is very concerned that she might eat a poisonous apple by mistake. He is therefore trying to decide whether or not to erect an electric fence that carries a small risk of harming his daughter.

In order to help him make his decision the farmer has asked you to estimate the chance of his daughter choosing an apple from a tree that bears **fatally poisonous** fruit if she were to randomly pick an apple from any of the trees in the orchard.'

There are two important pragmatic differences between the no-control and high-control conditions. Firstly, participants are informed either that 'there remain no feasible steps that the farmer can possibly take to remove the chance of his daughter…' in the *no-control* condition, whilst in the *high-control* condition they are told, 'he is therefore trying to decide whether or not to erect an electric fence…' The second difference is linked to participants' perception of their own control over the negative outcome and is conveyed in participants' instructions to estimate the probability. In the *no-control* condition, participants are simply asked to estimate the chance that the daughter will choose a fatally poisonous apple. In the *high-control* condition, participants read: 'In order to help him make his decision the farmer has asked you to estimate the chance of his daughter choosing an apple…'

In all other respects the procedure was identical to the preceding studies.

Results

Following exclusions (criteria as before), 65 males and 127 females were retained for analysis, 93 of whom were in the no-control condition and 99 in the high-control condition.

A mixed ANOVA was performed on the resulting data, summarised in Figure 2.8. The significant effect of probability was again observed, $F(1.7, 320.4) = 2949.42$, $p<.001$, $MSE = 102.86$. Crucially, there was also a main effect of the controllability manipulation on participants' probability estimates, $F(1, 190) = 6.27$, $p<.05$, $MSE =$

225.92, $eta_p^2 = .03$, such that probability estimates of the negative outcome were higher in the high-control condition. Additionally, the interaction between probability and the controllability manipulation was significant, $F(1.7, 320.4) = 4.62$, $p<.05$, $MSE = 102.86$ (Greenhouse-Geisser corrections applied). This interaction is explained by the absence of a difference between the controllability conditions at the high probability level (Figure 2.8).

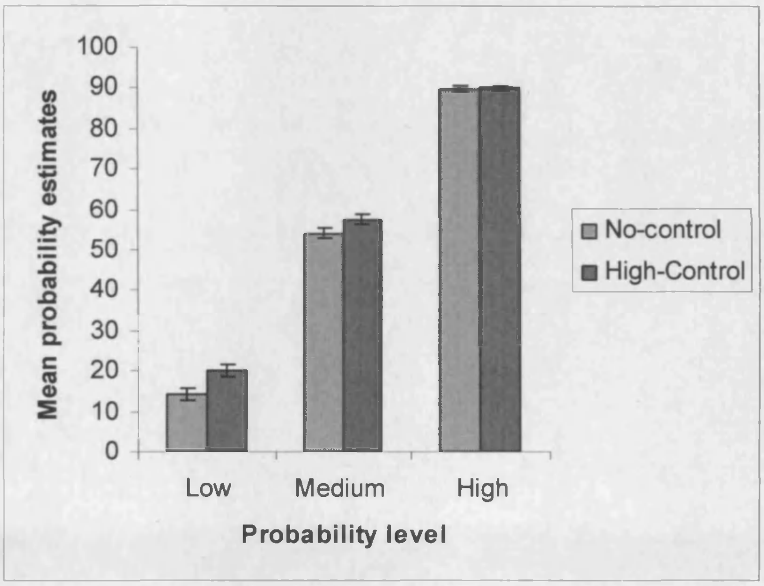


Figure 2.8. Mean probability estimates in the high-control and no-control conditions of Study 6. Error bars are plus and minus 1 standard error.

Discussion

The results of Study 6 provide support for an asymmetric loss function based explanation of the biasing impact of negative utility on probability judgments. This study made the notion of controllability explicit in its manipulations. If the loss asymmetry account is correct then there is an *implicit* sense of controllability in the version of the orchard cover story used in Studies 1 to 3. Hence it should be possible to match the data of the present study (with its explicit controllability manipulation) to the data from Studies 1 and 2 which used the same probability matrices (see Figure 2.9).

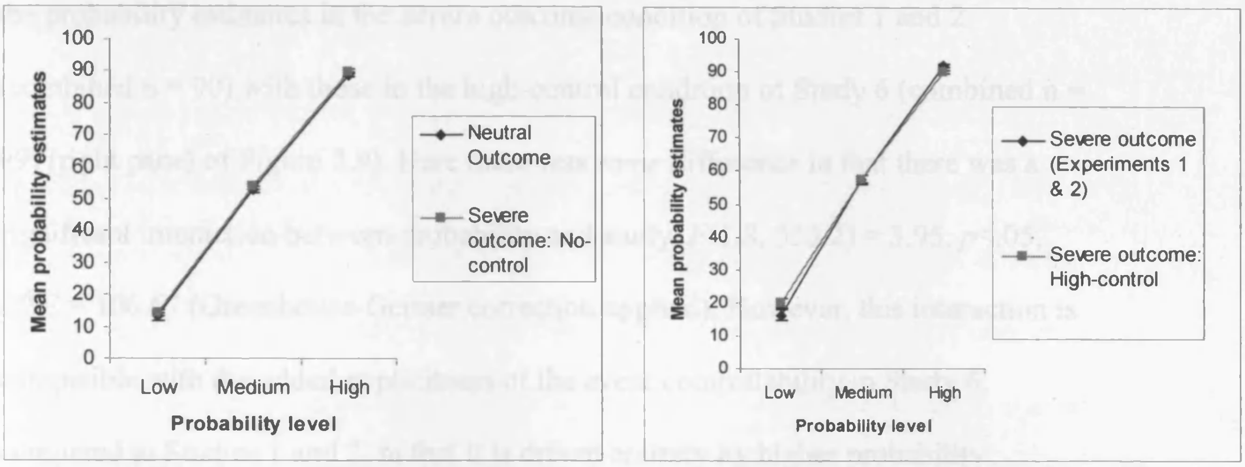


Figure 2.9. Plotted are the mean probability estimates of Studies 1 and 2 combined and the means of Study 6 in those conditions where no asymmetric loss function exists (left panel) and where an asymmetric loss function does exist (right panel). Error bars are plus and minus 1 standard error.

For there to be an asymmetry in the loss function associated with the probability estimate of an event, that event must be both controllable and severe; there is no asymmetry in the no-control condition of this study or the neutral outcome condition of Studies 1 and 2. A meta-analytic comparison of the relationships depicted in Figure 2.9 confirmed that the results for the respective conditions of Study 6 were analogous to the results of Studies 1 and 2.

Our meta-analytic procedure followed Rosenthal (1991). First, we compared the *neutral* conditions of Studies 1 and 2 (simply combining data from both, n = 83) with the no-control condition of Study 6 (n = 93).[5] As the left panel of Figure 2.9 shows, the respective means are virtually indistinguishable, and statistically there is no difference $F(1, 174) = 1.35$, $p>.05$, $MSE = 106.14$, $eta_p^2 = .01$. Second, we compared

[5] The ipsychexpts.com software prohibited participation in any experiment reported in this chapter if the IP address of the potential participant was recognised as having already completed any of the experiments reported in this chapter. Consequently, no independence assumption is violated in these meta-analyses.

the probability estimates in the *severe* outcome condition of Studies 1 and 2 (combined n = 90) with those in the high-control condition of Study 6 (combined n = 99) (right panel of Figure 2.9). Here there was *some* difference in that there was a significant interaction between probability and study, $F(1.8, 333.2) = 3.95, p<.05$, $MSE = 106.67$ (Greenhouse-Geisser correction applied). However, this interaction is compatible with the added explicitness of the event controllability in Study 6, compared to Studies 1 and 2, in that it is driven entirely by higher probability estimates at the low probability level of the high-control condition of Study 6, $F(1, 428) = 4.24, p<.05, MSE = 163.53$ (following Howell, 1997, pp. 470-471). Simple effects tests yielded no other significant differences between the points on the graphs in Figure 2.9.

One final source of support for the loss function based explanation of Studies 1 and 2 comes from a comparison of the effect sizes in those studies and the effect size of Study 6. A meta-analysis of the two manipulations comparing their effect sizes (r) (again following Rosenthal, 1991) finds no difference between the 'severity' effect of Studies 1 and 2 and the 'controllability' effect of Study 6.

In summary, the results of Study 6 coupled with the results of the meta-analytic comparison with Studies 1 and 2 provide good empirical support for a loss asymmetry account of our findings thus far. The reason severity exerts an effect in Studies 1 and 2 is because the outcomes in question are still perceived to be under human control. The final test of this explanation would be to show that adding an element of control to the bomber scenario gives rise to the same effects observed in the orchards paradigm.

Study 7

The asymmetric loss function account predicts that it should be possible to introduce an effect of outcome severity in the bomber scenario by modifying it to include an element of controllability. To test this, we factorially combined a severity and a controllability manipulation within this scenario. The loss asymmetry account predicts an interaction between controllability and severity such that severe events are assigned higher probability estimates than neutral events in the high controllability condition. However, it predicts no difference between the remaining three estimates (the two no-control conditions and the high-control/neutral outcome condition). These predictions are illustrated in Figure 2.10.

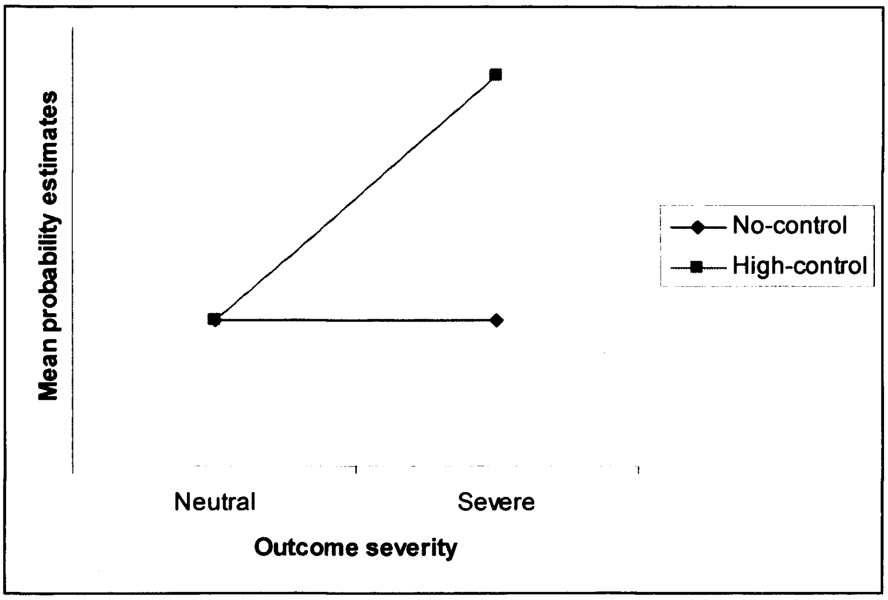


Figure 2.10. The interaction predicted in Study 7.

Method

Participants

An internet sample of 89 males and 177 females aged between 17 and 100(!) years completed the study, in an average time of 2.73 minutes. Once participants aged above 90 and under 18 were omitted, the age range was 18 to 69 years (median = 26 years).

Design

The design was a 3x2x2 (probability x outcome severity x outcome controllability) mixed design in which probability was the within-subjects variable and outcome severity and controllability were combined factorially between subjects. Participants were randomly assigned to one of the four conditions.

Materials

The visual displays were the same as those used in Study 5 (see Figure 2.6). However, in order to manipulate controllability it was necessary to use different cover stories.

The 'high-control' cover story read as below (in the severe outcome condition):

> 'The RAF are in need of a new training site for their pilots. The location currently favoured would involve flying over the area pictured below, in which the white area represents a densely populated town and the blue area represents the river that flows through that town. Crashes and falling plane debris are not uncommon occurrences in RAF training sites, and if falling debris were to land on a populous area, it would kill anybody beneath it. Any debris falling from the sky during training could land in any of the grid squares in the picture below.
>
> The RAF have asked you to use the picture below to estimate the chance that any falling debris would land on the densely populated dry land.'

The final paragraph of this cover story should be emphasised. Participants were informed that a character in the scenario had asked for their probability judgment. As such, participants could legitimately infer that their judgments might affect the final outcome through choices made by characters within the scenarios. In the 'no-control' condition, the following sentence was inserted after the first one; 'This is the only air space available to the RAF and hence must be used as the training of new pilots is essential.' In addition, in the 'no-control' condition it was not the RAF asking for the probability judgment, thus minimising the perceived influence of participants' probability judgments. The final paragraph in the 'no-control' condition therefore read:

> 'By looking at the picture below, please estimate the chance that any falling debris will land on the densely populated dry land.'

Outcome severity was manipulated within these cover stories by changing the white area from a 'densely populated town' to 'uninhabited wasteland'. If any debris was to fall in that area, participants were told it would 'litter that area'.

Procedure

The procedure was identical to that in Study 1.

Results

Following exclusions (criteria as before), 205 participants were retained for analysis. Participants' mean probability estimates in the four experimental conditions are shown in Figure 2.11. Visually, these results appear to be in line with our predictions (see Figure 2.10).

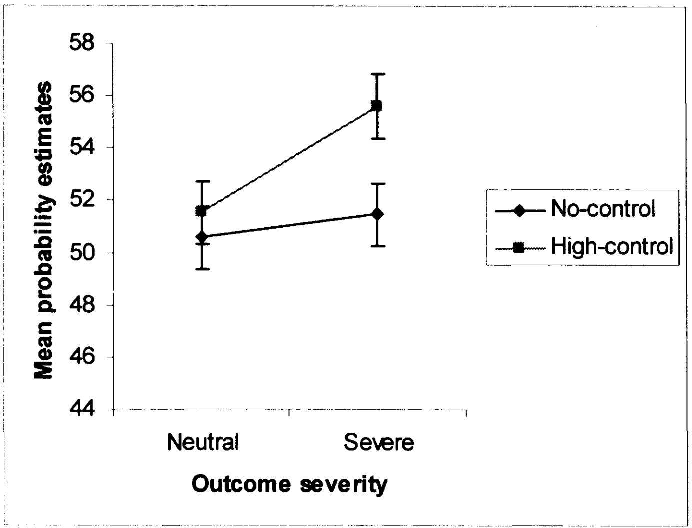


Figure 2.11. Mean probability estimates in the four experimental conditions of Study 7. Error bars are plus and minus 1 standard error.

Statistically, we observe significant effects of probability, $F(1.7, 342.9) = 2810.62, p<.001, MSE = 65.33$, severity, $F(1, 201) = 4.17, p<.05, MSE = 222.41, eta_p^2 = .02$, and controllability, $F(1, 201) = 4.48, p<.05, MSE = 222.41, eta_p^2 = .02$. In this overall ANOVA, the interaction between controllability and severity was not significant.

However, the results of planned simple effects tests (on their legitimacy in the absence of a significant overall interaction see Howell, 1997, p. 415) were in line with our predictions. A significant effect of outcome severity was observed in the high-control condition, but not in the no-control condition, $F(1, 201) = 2.51, p<.05, MSE = 222.41$; $F(1, 201) = .27, p>.05, MSE = 222.41$, as expected.

Another way to test the account is to apply Rosnow and Rosenthal's (1995) test of the predicted pattern of means. This test showed that the predicted pattern of results was supported by our data, $F(1, 201) = 10.17, p<.01, MSE = 222.41, r = .22$. This result, which takes into account our specific predictions as to the pattern of the means, adds further support to our explanation of the utility/probability interdependence observed in Studies 1 to 3.

Discussion

Study 7 provides the first demonstration of an effect of outcome severity on probability estimates with a different cover story (i.e., not involving the 'orchards' paradigm), and confirms that the critical difference between Studies 1 to 3, which found an effect of severity, and Studies 4 and 5, which failed to find this effect, lies in the respective presence or absence of an element of control. This further confirmation that an element of control is crucial for the effect of severity to obtain directly supports the asymmetric loss function account.

Chapter Discussion

We have presented seven studies investigating the effect of outcome severity on probability estimates. Our studies used a minimal paradigm, in which an objective representation of the probability to be estimated was constantly available to participants. Studies 1, 2 and 3 (using the 'orchards' scenario) showed that severe events were rated as more likely to occur than neutral events. This result was not observed within the 'bomber' scenario (Studies 4 and 5). Studies 6 and 7 demonstrated the importance of the controllability of the event in observing the effect of outcome severity on probability estimates; the effect was only observed for controllable events.

This overall pattern of results is explained through loss function asymmetry. Within the asymmetric loss function account it is assumed that people's judgments are sensitive to the 'uncertainty of the uncertainty'. For severe outcomes, it is often the case that the costs associated with underestimating their probability are greater than those associated with an overestimate. Probability judgments of such events are therefore inflated, which acts as a protective measure against the negative effects associated with an underestimate. However, there can only be costs associated with a mis-estimate of the probability of an event if a decision is subsequently based on this estimate. Thus this account is a decision-theoretic explanation and a loss asymmetry only exists if the event in question is somehow controllable (as a decision cannot make a difference if the outcome is uncontrollable). This account thus also explains why no effect was observed between the two experimental conditions in Studies 4 and 5.

Lerner and Keltner (2001, Study 4) found that a fear inducement made people more pessimistic with regard to future events than happiness or anger inducements. Some might argue that in the orchards paradigm, our severity manipulation is evoking fear in people (possibly as a result of a simulation of what may happen to the daughter [see also, Krizan & Windschitl, 2007]), which leads them to overestimate the likelihood of the negative event occurring. Such an account would be plausible if it were not for the results of Studies 6 and 7. Such an account would not be able to account for the different results obtained under conditions of low and high controllability. In fact, if controllability were to have an effect, a fear based account would make the opposite predictions as fear is typically associated with low human controllability (as controllability is defined here) (e.g., Smith & Ellsworth, 1985), and thus would predict a greater effect under conditions of low controllability than under conditions of high controllability. A fear based account does not, therefore, seem to be able to explain the data presented in this chapter.

Likewise, the data presented are not consistent with an account based on the surprisingness of the outcome. Teigen and Keren (2002) found that participants reported greater surprise for positive outcomes than negative outcomes when the events were not controllable (that is, the outcomes were determined by chance). When an outcome was perceived as controllable (an 'action outcome'), this result was reversed such that greater surprise was reported for the negative outcomes. If, as seems reasonable to assume (see e.g., Christensen, 1979; Fisk & Pidgeon, 1996, 1997, 1998), people assign lower subjective probabilities to more surprising outcomes, this result corresponds to negative controllable outcomes receiving lower probability estimates than negative uncontrollable outcomes. Such a pattern is opposite to the one observed in the present study. This inconsistency provides further evidence for the lack of a simple relationship between subjective probability and surprise (see also, e.g., Maguire & Maguire, 2009; Teigen & Keren, 2002, 2003). The present loss asymmetry based account seems to be the most parsimonious account for the data presented in this chapter.

Asymmetric loss functions have received much attention in adjacent disciplines, especially economics (e.g. Batchelor & Peel, 1998; Goodwin, 1996; Granger, 1969) and forecasting (e.g. Armstrong, 2001; Lawrence & O'Connor, 2005; Lawrence, O'Connor, & Edmundson, 2000). Within these fields, asymmetric loss functions are ubiquitous. Furthermore, in many contexts, people's sensitivity to these in their estimates has been shown to be rational (e.g., Batchelor & Peel, 1998; see also Whiteley & Sahani, 2008, and references therein). Lawrence and O'Connor (2005), for example, empirically manipulated the shape of loss functions and found that people's forecasts of business data were sensitive to these different shapes. Asymmetric loss functions have been given far less consideration in psychology, and we are aware of only a handful of studies that have investigated the concept (e.g., Birnbaum, Coffey, Mellers and Weiss, 1992; Landy, Goutcher, Trommershäuser, & Mamassian, 2007; Whiteley & Sahani, 2008) or used it to explain past results (Weber, 1994). What is novel about our present studies in this wider context is that they identify, and test experimentally, the importance of control.

Consideration of this wider literature on asymmetric loss functions also clarifies *what* needs to be controlled. Biasing influences of loss asymmetry are found in meteorological forecasting (e.g., Solow & Broadus, 1988), but clearly it is not the weather itself that is subject to control. What matters is simply the presence of further decisions on the basis of the estimated outcomes and the potential for these decisions to reduce associated costs (e.g., carrying an umbrella on a rainy day).

Identifying the impact of control also allows the resolution of inconsistencies in the literature investigating the interpretation of verbal probability expressions. As noted in the introduction, verbal probability expressions are plagued by base rate effects and, in the real world, base rate and severity are confounded. Hence, genuine tests of severity require a context in which base rates are controlled for. To date, only Weber and Hilton (1990) and Fischer and Jungermann (1996) have done this. However, in controlling for base rates, Fischer and Jungermann gave participants a rather unusual experimental question. Asked to make estimates relating to side effects of drugs they were told that "It is known that such drugs (i.e., drugs treating this disease) usually lead to headaches in 10 out of 1000 cases. The information in the leaflet says that this particular drug "rarely" leads to headaches. Which numerical interval do you think matches the word "rarely"?" (Fischer & Jungermann, 1996, p. 156). Given that participants are being given an explicit anchor for their interpretations of probability expressions, it seems unsurprising that no effect of severity was found. Hence, Weber and Hilton's studies are really the only ones to have examined a potential influence of severity while controlling for base rates in a meaningful way. However, their results were conflicting. Across two studies using regression analyses to factor out effects of base rate, and a further study in which they sought to manipulate base rates experimentally, they found higher estimates with increased severity only for some materials.

Specifically, only their first four medical scenarios, drawn from a previous study by Wallsten, Fillenbaum and Cox (1986) (see Table 2.1 [from Wallsten, Fillenbaum, & Cox, 1986, p. 574]), showed a reliable positive influence between severity and probability. However, these were also the only scenarios that involved a decision and hence an element of control. In Weber and Hilton's own scenarios, participants were asked to provide numerical probability estimates for statements by doctors given in the context of an annual medical check-up such as "your doctor tells you that there is a slight chance that you will develop an ulcer during the next year" (Weber & Hilton, 1990, p. 784) or your doctor tells you that "It is likely that you will develop a severe and common type of influenza in the next year" (Weber & Hilton, 1990, p. 787). No decision is implied in this context, so no increase with severity *should* be observed. By contrast, as can be seen from Table 2.1, the materials of Wallsten, Fillenbaum and Cox (1986) contrast a high-severity event involving a decision about a flu shot and its side effects, with low severity events involving little or no control. Consequently the strong relationship between severity and probability estimates observed for these four scenarios is consistent with our present results. Weber herself (Weber, 1994) posited that asymmetric loss functions might lead to effects of severity on the interpretation of verbal probabilities. Realising, in addition, how the presence or absence of decisions and control affects loss asymmetries allows the seemingly conflicting findings in this area to be resolved.

Table 2.1.
The scenarios used in Wallsten et al. (1986)

You normally drink about 10-12 cups of strong coffee a day. The doctor tells you that if you eliminate caffeine it is likely your gastric disturbances will stop. What is the probability that your gastric disturbances will stop?______________
You have a wart removed from your hand. The doctor tells you it is possible it will grow back again within 3 months. What is the probability it will grow back again within 3 months?____________
You severely twist your ankle in a game of soccer. The doctor tells you there is a slight chance it is badly sprained rather than broken, but that the treatment and prognosis is the same in either case. What is the probability it is sprained?___________
You are considering a flu shot to protect against Type A influenza. The doctor tells you there is a chance of severe, life-threatening side effects. What is the probability of severe, life-threatening side effects?____________

Locating the Effect

The next issue to address is where in the overall process of generating and reporting a probability estimate participants are influenced by loss asymmetries. Figure 2.12 illustrates the three major stages involved in the production of an estimate. Ascertaining the locus of the present effect with respect to this diagrammatic representation (Figure 2.12) is not a straightforward task. What seems clear is that the present effect does *not* reside in the first stage of the process. All the evidence required to make the probability estimate is available throughout the task in all our studies, and the differences we find across conditions cannot be construed as differences in the processing of this information. In particular, the differences we observe are not based on the fact that people might take more care in making their estimates and are consequently more accurate when estimates are more important (i.e., under controllable, severe conditions). Across probability levels, participants' estimates are higher when the outcome is severe. This means the estimates move *above* their objective values, in all but the high-probability condition where 'increased accuracy' and loss asymmetry based inflation happen to coincide. At the medium probability level, participants are already quite accurate in the neutral condition; the severity manipulation moves their estimates above the actual objective levels. In the low probability condition of Studies 1 and 2 the objective probability is less than 5%. However, the mean estimates in the neutral outcome condition lie at 13% and they become even higher, not lower, in the severe condition.

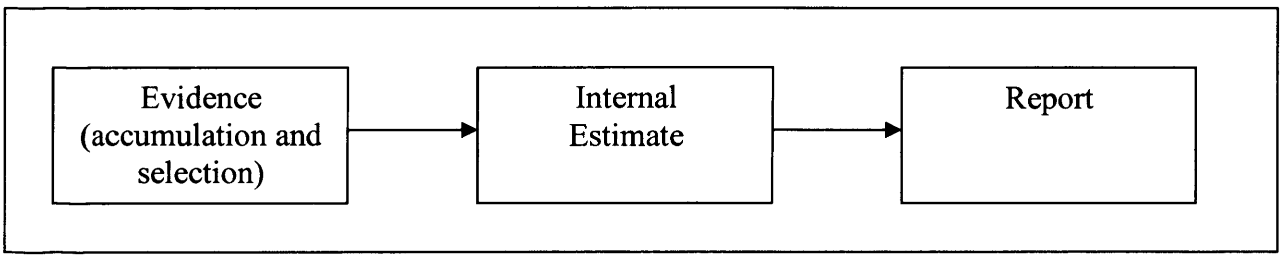


Figure 2.12. The process of making and reporting a probability estimate.

Of course, the fact that participants are given all the information they require to make their estimates does not rule out the possibility that they could be drawing on other information in addition. Specifically, participants might use information about real-world base rate (Dai et al., 2008) or about real world 'representativeness' (Mandel, 2008; Windschitl & Weber, 1999) as an additional source of knowledge in a context in which they find probability estimates difficult to make. However, effects in the orchards scenario are in the opposite direction to those predicted by an 'associative representativeness' or a 'base rate influence' account; participants will have had more experience of people picking sour rather than fatally poisonous apples, and sour apples are far more prevalent than apples sprayed with lethal pesticide.

Consequently, only the internal judgment and the report stage seem plausible sites for our loss asymmetry based effect. Do loss asymmetries bias participants' internal estimates, or their reporting of those estimates? This issue is difficult to decide conclusively, but the evidence points toward a biasing effect of which participants are unaware. For one, across Studies 1-7, those participants who did contact us with further questions following debriefing were interested exclusively in how accurate they had been. Past experimental research demonstrating effects of loss functions on estimates has been silent on the issue of whether or not participants might be aware of their bias (e.g., Birnbaum & Stegner, 1979; Birnbaum et al., 1992; Bottom & Paese, 1999; Lawrence & O'Connor, 2005; Weber, 1994), and there has been no empirical investigation of this issue. However, examination of the exact nature of loss asymmetry's influence and the mechanisms posited in this research suggests, most likely, that these biases are not conscious. Most closely related to our findings is Weber and Hilton's (1990) effect of loss asymmetry on the interpretation of verbal probability expressions. It makes little sense in this context that participants should consciously inflate an estimate of their own risk of disease in a fictitious doctor's report, and indeed Weber (1994) viewed the effect as stemming from mental simulation processes for deriving the estimates themselves (as in Einhorn & Hogarth, 1985; Hogarth & Einhorn, 1990).

Bottom and Paese (1999) had their participants engage in a bargaining task. Participants were assigned to pairs in which one of them played the role of a buyer and the other the role of a seller of a used truck. Buyers were instructed to try to get a price below a "buyer's reservation price" of $15,500, whilst sellers were instructed to attempt to agree a price above $13,500 (the seller's reservation price). During the negotiation process, participants completed a form in which they estimated their opponent's actual reservation price. In these forms, Bottom and Paese predicted and found evidence of 'wishful thinking' based on loss function asymmetry. Crucially, there was no reason to consciously bias the report of the price estimate, as this estimate was not seen by the opponent. It seems, therefore, that participants' true, subjectively held, internal estimates of the reservation price were being affected by the asymmetric loss function, rather than merely their reports. Consideration of the feedback of these participants adds further support to this conclusion. At the conclusion of the experiment both parties in the negotiation disclosed their own reservation price. A number of optimistic buyers and sellers reported that they still believed their own optimistic estimates of their opponent's price and concluded that their opponents were lying.

The final set of experimental studies involving loss asymmetries and estimates we know of are those conducted by Birnbaum and colleagues (Birnbaum et al., 1992; Birnbaum & Stegner, 1979). In their experiments, participants provided estimates of the price of a second-hand car from the point of view of both the buyer and the seller.

These two points of view induce different asymmetric concerns for over- or underestimating value: "when instructed to estimate the highest price that a buyer should pay, the judge [*participant*] considers it a costly error to set too high a price (because the buyer would suffer a loss) but a less costly error to offer too little (violating the instruction to judge the highest price)" (Birnbaum et al., 1992, p. 335). Thus their task is very similar to ours in that it requires what is ostensibly an estimate for the benefit of a third party. The bias induced by loss asymmetry figures in the best fitting models of this task in a mathematically non-trivial way that seems unlikely to be consciously accessible to participants.

In summary, related research provides some support for the contention that asymmetric loss functions are biasing participants' subjectively held, internal representations of probability, rather than merely their reports of unbiased internal probabilities.

Whether our effect involves the internal estimate or its report, it is clear from the manipulations of control that outcome severity does not *inherently* bias probability. Specifically, there is no evidence for a simple 'I fear...therefore I believe...' relationship, because the feared outcome is the same in conditions with and without control. Hence our results complement the consistent failure to find experimental evidence of an inherent bias, that is, "I wish for, therefore I believe in" (Bar-Hillel et al., 2008, p. 283), within the positive domain (Bar-Hillel & Budescu, 1995; Bar-Hillel et al., 2008; Krizan & Windschitl, 2007). In particular, our results fit with Bar-Hillel and Budescu's (1995) studies of wishful thinking in a similar paradigm in which the relevant objective probabilistic information was continuously available to participants. The apparent 'elusiveness' of the wishful thinking effect under these conditions is entirely consistent with the present findings in that there are no loss asymmetries associated with estimates of those positive events.

Chapter Summary

We found experimental evidence that outcome severity influences probability estimates via sensitivity to loss asymmetry. This is the first clear evidence of the biasing influence of utility in the negative domain. It is also, to our knowledge, the first investigation of the impact of control on loss asymmetries. Identification of the role of control allows one to make sense of related, but seemingly mixed, results in the literature on the interpretation of verbal probability expressions (Weber & Hilton, 1990). Although the evidence suggests that utility does not inherently affect probability, the prevalence of asymmetric loss functions will mean that estimates of probability are frequently biased in practice.

Chapter 3 - Estimating the Probability of Positive Events

Chapter Overview

Chapter 2 provided a systematic investigation of the potential interdependence of negative utility and estimates of probability. The majority of the literature investigating the relationship between probability and utility has, however, concerned positive events. As outlined in the introduction to Chapter 2, the degree to which this literature suggests a direct influence of positive utility on probability estimates, a 'wishful thinking' effect is unclear, since there are competing explanations in the majority of past studies. In this chapter we present four studies that use cover stories that are both affectively rich and increasingly personally relevant to test for a 'wishful thinking' effect. Despite these efforts, no effect of positive utility was observed across the four studies.

Introduction

The studies reported in the last chapter demonstrated that people's probability estimates are not routinely biased by considerations of negative utility. We did, however, report evidence suggesting an indirect effect of event utility on probability estimates through the mediation of loss function asymmetries. In the introduction to Chapter 2, we highlighted that few studies directly tested the potential biasing effect of probability estimates by utility considerations, either negative or positive. However, the majority of research on the issue has concerned positive events (e.g., Babad, 1995; Babad & Katz, 1991; Bar-Hillel & Budescu, 1995; Granberg & Brent, 1983; McGuire, 1960a, 1960b) and there are seemingly numerous indications from field study research suggesting that people are optimistic in that they overestimate the probability of positive events. Babad and Katz (1991) (see also, Babad, 1987), for example, demonstrated that wishful thinking in Israeli soccer fans persisted even when the dependent variable was sports betting, an activity which clearly rewards objectivity. Although such results suggest the presence of a wishful thinking bias in practice, the nature of the bias is not clear from the results of such field studies, which necessarily lack the control of a laboratory study. As mentioned in Chapter 2, Bar-Hillel and Budescu (1995) posit that seeming optimism observed outside controlled laboratory settings can be well explained as "an unbiased evaluation of a biased body of evidence" (Bar-Hillel & Budescu, 1995, p. 100, see also Gordon et al., 2005; Morlock, 1967). In this chapter, we will extend the direct test used in the previous chapter to test the influence of *positive* utility on probability estimates. Given that there can be no asymmetric loss function, by definition, for positive events, the results of the tests investigating the negative domain do not suggest that a wishful thinking effect should exist.[6]

It should by now be apparent that, in determining the effect of extreme utility on subjective probability, the approach taken has been to compare extreme utilities with more neutral utilities of the same valence. Such an approach is different to that in many previous studies which have been cited as evidence, or not, of an

---

[6] Although Weber (1994) proposes that underestimates of positive outcomes do carry greater costs, we do not concur with this position. We define a cost as being a negative consequence. When dealing with possible positive outcomes, there are either positive consequences (the outcome occurs) or no consequences (the outcome does not occur).

interdependence between probability and utility (e.g., Bar-Hillel & Budescu, 1995; Irwin, 1953; Marks, 1951; Morlock, 1967; Pruitt & Hoge, 1965). Irwin (1953), for example, reported a comparison of guesses made by participants when they were told that an outcome would win them a point (desirable) versus when it would lose them a point (undesirable).

There are conceptual reasons for examining positive events separately from negative events. There is a substantial body of research into affect suggesting that affect is not a unidimensional construct with negative affect at one end and positive affect at the other. Rather, negative affect and positive affect consistently emerge as two separate dimensions of self-reported mood, meaning that they are independent and, consequently (for example), a decrease in positive affect will not necessarily lead to a corresponding increase in negative affect (see e.g., Berscheid, 1983; Bradburn, 1969; Isen, 1984; Watson, Clark, & Tellegen, 1988; Watson & Tellegen, 1985; Zevon & Tellegen, 1982; for a discussion see Taylor, 1991). This result has also been replicated cross-culturally (Watson, Clark, & Tellegen, 1984), suggesting the fundamentality of the distinction between positive affect and negative affect. Furthermore, a number of researchers have found positive and negative affect to be related to different personality factors. Costa and McCrae (1980), for example, reported a strong positive relationship between neuroticism and negative affect, which was not present for positive affect, and a positive relationship between positive affect and extraversion, which was not present for negative affect (see also, Tellegen, 1984; Warr, Barter, & Brownbridge, 1983; Watson & Clark, 1984). Taylor (1991) reviews evidence supporting the insight that negative events lead to the elicitation of more physiological, affective, cognitive and social processes than positive events which, in some respects "may take care of themselves" (Taylor, 1991, p. 80). This theoretical disassociation between positive and negative affect supports the present approach of investigating the effects of positivity and negativity separately.

Many studies have already purported to find evidence of wishful thinking (the inflation of subjective probabilities concerning the occurrence of 'good 'events) in natural, real world settings (e.g., Babad, 1995; Babad & Katz, 1991; Granberg & Brent, 1983). Within a controlled laboratory setting, Price (2000) reported the existence of a wishful thinking effect in the laboratory using a competitive group paradigm, as described in the previous chapter. There, however, we argued that the effect observed by Price is well-explained by cognitive and motivational factors relating to intergroup competition (e.g., Blake & Mouton, 1961; Jourden & Heath, 1996; Sherif & Sherif, 1956). The existence of these confounding factors means that Price's study cannot be considered a satisfactory demonstration of the wishful thinking effect.

In a related study, Slovic (1966) investigated the effect of outcome utility, not on subjective probabilities about future outcomes, but on posterior conditional probabilities, referring to the contents of five bags. His experiment resembled a traditional 'conservatism' task. Participants were asked to pick one of five bags ostensibly consisting of different proportions of red and blue chips. As chips are drawn from the bag, participants are required to estimate the probability that the bag has a particular proportion of red and blue chips. Utility was manipulated in this experiment by telling participants that they would gain or lose money if the bag consisted of certain proportions of chips. Slovic's results were mixed, with some participants seemingly demonstrating optimism, and others demonstrating pessimism. Moreover, these trends in the data failed to reach statistical significance. In Experiment II, participants were offered monetary incentives for accuracy. In this experiment, there was evidence that some participants 'hedged their bets' by consciously biasing their probability estimates in favour of the bags associated with a monetary loss, in order to strategically protect themselves against losing money. This result demonstrates the pragmatic difficulties associated with using negative and positive *monetary* payoffs to manipulate outcome utility in experiments, a point also made in Chapter 2 with reference to the Pruitt and Hoge (1965) study.

The wishful thinking effect has, however, already been tested more directly within a controlled laboratory setting using a paradigm very similar to that introduced in Chapter 2. Bar-Hillel and Budescu (1995) used visual matrices (white and pink) to provide participants with an identical objective basis for their probability judgments across conditions and failed to find evidence of a wishful thinking effect. The outcome utility manipulation in this experiment was not however qualitatively equivalent to that employed in Chapter 2, and this is why the possible existence of a wishful thinking effect requires further investigation under controlled conditions.

In Bar-Hillel and Budescu's (1995) design, desirability was manipulated only through the use of monetary incentives. A white square might, for example, represent a win of 50NS (New Israeli Shekels, at the time worth approximately $20), as in their high gain condition (Study 1, Experiment 1). Their failure to find significant results could result from the affective poverty of their desirability manipulation. Rottenstreich and Hsee (2001) have demonstrated a disassociation between affect and evaluation based on monetary value. Whilst having the same monetary value, a $500 coupon 'that could be redeemed towards expenses associated with a summertime European vacation' (Rottenstreich and Hsee, 2001, p. 187) was found to carry more affective value than a $500 coupon that could be redeemed as money towards university tuition payments. Similarly, Shaffer and Arkes (in press) demonstrated that, when evaluated separately in a between-subjects design, non-cash rewards tended to be evaluated more favourably than cash rewards of an equivalent value. Given our use of between-subjects designs, necessary to avoid potential demand characteristics in these studies, a non-cash manipulation of outcome utility would seem to increase the likelihood of observing a wishful thinking effect in a laboratory study, as also posited by Rottenstreich and Hsee (2001, p. 185): "Although probability-outcome independence may hold across outcomes having different monetary values, the affective approach implies that it is unlikely to hold across outcomes having different affective values."

In addition to the lack of an affective quality to the rewards used in Bar-Hillel and Budescu (1995), the specific hypothesis tested in their data analyses may have contributed to the null results reported in their paper. Bar-Hillel and Budescu assume that if people's probability judgments are influenced by outcome (un)desirability then this is in the direction of a general optimism effect, as represented by the top panel of Figure 3.1. Although not demonstrating a direct effect of utility on probability estimates, the results in the previous chapter demonstrate an indirect effect of utility on probability estimates in the direction of pessimism (see also, Dai et al., 2008; Mandel, 2008). Consequently, a general optimism effect does not appear to exist. In all of their analyses, Bar-Hillel and Budescu treat outcome desirability as a unidimensional variable with levels from very undesirable through to very desirable, as operationalised by the monetary gain or loss associated with the outcome. If our conceptualisation (that outcome desirability is a two dimensional variable) is, in fact, more accurate, as it appears to be (see also, Taylor, 1991), then it is of little surprise that the results of their ANOVAs yielded no significant effect of wishful thinking in the majority of instances. For negative events (if a loss function asymmetry was present), the severity of the negative outcome would have been expected to exert the opposite to an optimism effect in some instances, which would consequently cancel out the desirability effect in the positive domain. Indeed, in Bar-Hillel and Budescu (1995, p. 80), when the outcome was assigned personal relevance (the condition in which all effects of outcome (un)desirability are expected to be greatest), for 5 out of the 10 target proportions in two different response formats, the judged probability of the target event was greater in the high loss condition than the low loss condition (whilst only 3 out of 10 showed the opposite effect). Averaging across the target proportions for each response format, higher probability estimates were made in the high loss condition than in the low loss condition. As their theoretical approach does not consider such a possibility, there is no further mention of this bidimensionality in their paper.

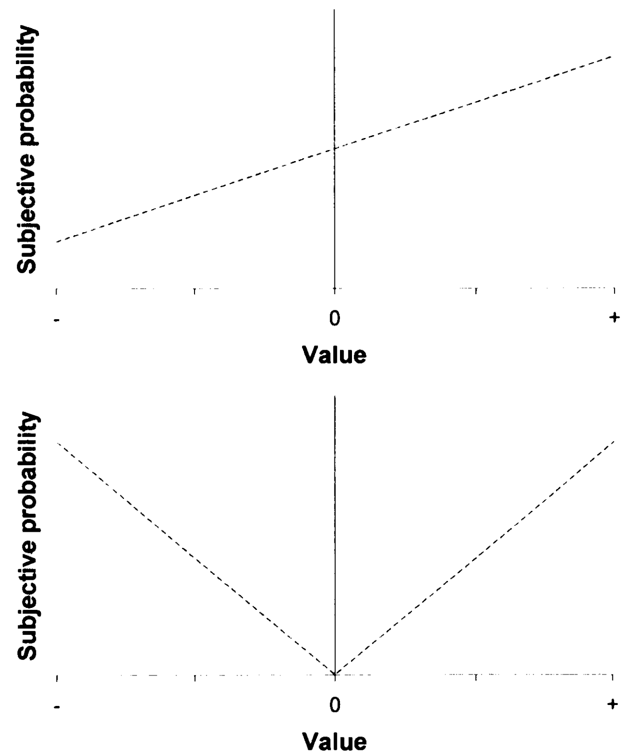


Figure 3.1. A graphical comparison of a general optimism theory (top panel) with an 'It can happen to me' (Slovic, 1966, p. 23) theory (bottom panel). Reproduced with permission from the author.

Study 8

Study 8 used the paradigms developed in the previous chapter to investigate the effect of positive outcome utility on probability estimates. In concordance with the scenarios used in the previous chapter, we sought to maintain an affective quality in the stimuli we used for this investigation, in order to enhance the possibility of finding a wishful thinking effect. In addition, and in accordance with our thinking above, Slovic (1966) postulates that it is the affect associated with extremely positive outcomes that might lead to a wishful thinking effect.

Method

Participants

Participants were recruited via the ipsychexpts.com website. The same basic checks of the data as those described in Study 1 were carried out as the data were being collected. Participants whose data did not meet the eligibility criteria were excluded from subsequent analyses. Data collection continued until there were 50 participants in each experimental condition (48 male and 52 female participants aged between 18 and 55 [median age = 23]).

Design

The hypothesis under investigation was whether probability estimates for extremely good ('positive') outcomes were different from those for more neutral ('neutral') outcomes. The central independent variable of interest was therefore the utility of the outcome, which was manipulated between participants through two different cover stories, one for each outcome utility. A second independent variable, as in the previous studies already described, was the objective probability of the outcome. This second independent variable was manipulated within participants using the same visual matrices described in the previous chapter (Studies 1 and 2; see Figure 2.1). The order in which participants saw the different probability matrices was randomised across participants. The dependent variable was the probability estimates obtained from participants.

Materials and procedure

The visual matrices used in this study were the same as those used in Studies 1 and 2 above (see Figure 2.1). The procedure was likewise the same as for those studies and it was conducted over the internet using the website ipsychexpts.com. From the perspective of the participant, the only difference between this study and Study 1 was in the cover stories: Participants read that Rolex's annual employee of the year is given the chance to take part in a lucky dip from a selection of identical boxes held in the company storeroom. Participants were informed that some of the boxes were empty whilst some contained 'a diamond encrusted Rolex watch with a retail price of £9,820' (in the positive outcome condition) or 'a Rolex paperweight with a retail price of £9.99' (in the neutral outcome condition). Participants were presented with the visual matrix and told that yellow cells represented boxes containing Rolex products, whilst the grey cells were empty.

Participants were asked to 'estimate the chance that John will pick a box containing a diamond encrusted Rolex watch' or 'estimate the chance that John will pick a box containing a Rolex paperweight'. They then gave their responses by clicking on the appropriate radio button, the radio buttons ranged from 0% (absolutely impossible) to 100% (absolutely certain [100% certain]). The scale was further anchored at 50 (a 50/50 chance). Buttons were included for every 5% interval[7].

After having made their three probability judgments and given their demographic details, participants were presented with a screen debriefing them as to the purpose of the study.

Results

Data exclusions had already been carried out before the data were analysed (see above). However, when analysing the data of the 100 eligible participants, it was found that a 17 year old male had incorrectly been included. This participant was subsequently removed from analysis, in line with departmental ethical guidelines, and analysis continued. This removal meant that there were only 49 participants in the positive outcome condition.

Figure 3.2 summarises the results and subsequent statistical analyses indicated that whilst there was a significant effect of probability, $F(2, 188) = 1434.254, p<.001$, $MSE = 88.09$, there was no effect of the outcome utility manipulation, $F(1, 94) = .278$, $p>.05$, $MSE = 165.34$, $eta_p^2 = .003$, nor was there an interaction between probability level and outcome utility, $F(2, 188) = .752, p>.05$, $MSE = 88.09$.

[7] Adam Corner assisted with the writing of the cover story for Study 8.

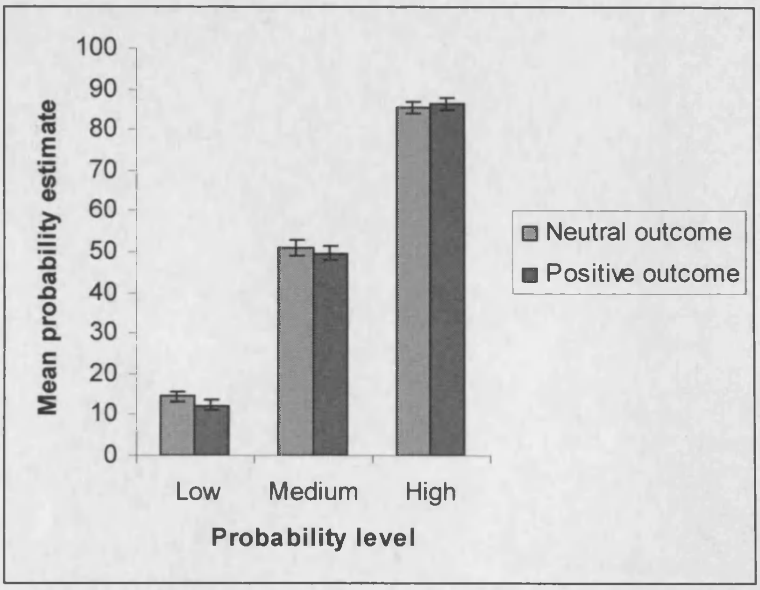


Figure 3.2. Mean probability estimates at the different probability levels in the 2 experimental groups. Error bars are plus and minus 1 standard error.

Discussion

The results of Study 8 replicate Bar-Hillel and Budescu's (1995) failure to find an effect of wishful thinking when participants are provided with an objective basis for their probability judgments. It could, however, be argued that the utility manipulation was not sufficiently strong to affect the probability estimates. A further study was designed that attempted to strengthen this manipulation to test the existence of a wishful thinking effect.

Study 9

In this study we sought to increase the strength of the manipulation of positive outcome utility, whilst still maintaining the third person nature of the scenario. Although the event included in the cover story concerned a third party, the potential eventual outcome was introduced as being beneficial to society as a whole. The cover stories were also piloted to ensure the validity of the manipulation. Six participants rated the desirability of the positive outcome and a further six rated the desirability of the neutral outcome. The positive outcome was rated as significantly more desirable than the neutral outcome, $t(10) = 3.29, p<.01$. Thus, the validity of the manipulation was confirmed.

Method

Participants

For this study we used ipoints.co.uk™ to recruit participants. The recruitment process was the same as for Study 5 and participants again received 50 ipoints for participating in this study. 58 males and 80 females, aged between 16 and 72 (median age = 34.5 years) participated in this study.

Design and materials

The design was identical to Study 1, with the same visual matrices used to provide an objective anchor for the probability estimates. A different cover story was written to manipulate outcome utility.

Neutral outcome:

'Across the planet there is great variety in plant life. One thing that has been gained from the discovery of new plant species is numerous cures for different diseases. It seems probable therefore that there is an abundance of plant life in rainforests that can yet be discovered that may hold the key for the cure of cold sores. Cold sores are itchy and unsightly and many people would love the chance to find a genuine cure for these blemishes.

A team of investigative scientists are in the Amazonian rainforest searching for plants that may form the basis of a cure for cold sores. There are indeed plants with the capacity to cure cold sores within the rainforest. However, they are only found in some of the rainforest's many research sites. The research team can only search one research site within the rainforest.

The matrix below depicts the different research sites located in the Amazonian rainforest. Plants with the capacity to cure cold sores only exist in those research sites represented by a YELLOW square.

By looking at the matrix below, what do you think is the chance that the team are searching in a site containing plants with the capacity to cure cold sores (a YELLOW area), thus one day ridding millions of cold sore sufferers of these unsightly blemishes?'

Positive outcome:

The cover story in the positive outcome condition was the same except that the plants being searched for might form the basis for the cure of cancer, rather than cold sores. Aside from changing the words cold sores to cancer, the last sentence of the first paragraph contained a description of cancer intended to highlight the severity of the disease and therefore, we hoped, the extreme positivity of finding a cure for the disease. The sentence read: 'Cancer is one of the leading causes of death in the Western World, accounting for 25% of deaths in the UK. In 2004, 153,397 people died from cancer.' In a further effort to increase the salience of the beneficial effect of finding these plants, the final clause of the final paragraph read: 'thus one day saving the lives of millions of cancer sufferers across the globe?'

Procedure

The procedure for Study 9 was identical to that for Study 5.

Results

Participants were excluded from the analysis using the same criteria as in Study 1. As was the case in Study 5 in the previous chapter, there were uneven numbers of participants in the experimental conditions. Following exclusions (criteria as before) there were 44 participants in the neutral outcome condition and 57 in the positive condition. As a result of the large discrepancy between the sizes of the two experimental groups in this study, we decided to exclude the final 13 participants to submit data to the positive condition from the analysis thus leaving 44 participants in each experimental condition.

Despite the apparent validity of the outcome utility manipulation (as suggested by the results of the pilot study), the means presented in Figure 3.3 suggest little effect of the outcome utility manipulation. This is supported by a mixed ANOVA performed on the data. No significant effect of outcome utility was observed, $F(1, 86) = 1.05$, $p>.05$, $MSE = 481.24$, $eta_p^2 = .012$, nor was there an interaction between outcome utility and probability, $F(2, 172) = .14$, $p>.05$, $MSE = 144.77$, $eta_p^2 = .002$. It should be highlighted that not only was there no significant effect of the utility manipulation, but that the non-significant trends in the data were in the opposite direction to those predicted by the wishful thinking hypothesis.

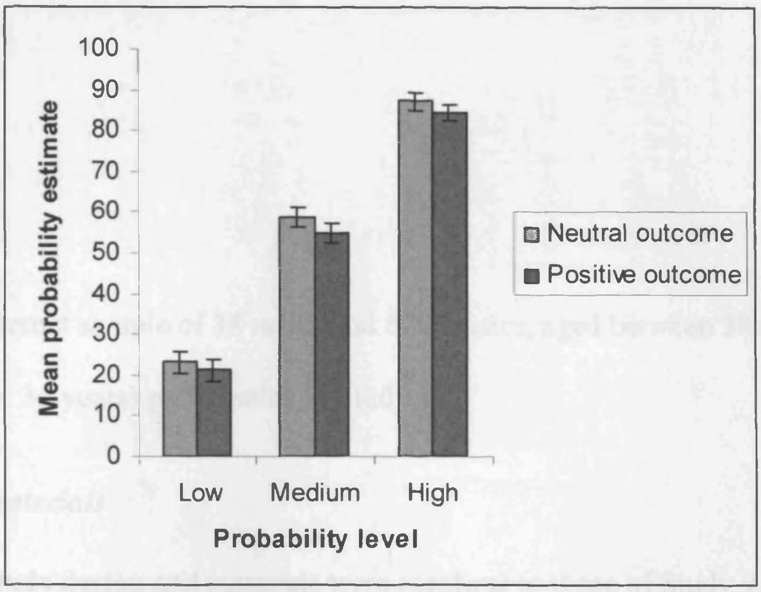


Figure 3.3. Mean probability estimates in Study 9. Error bars are plus and minus 1 standard error.

Discussion

Study 9 replicated the null result observed in Study 8, that is, no effect of outcome utility was observed on participants' probability estimates for these positively valenced events. In addition, in Study 9, across all three probability levels there was a trend for lower probability estimates to be given in the positive outcome utility condition. The results of the pilot test that tested the strength of the manipulation, coupled with the trend for the results to be in the opposite direction suggests that participants truly did not exhibit a wishful thinking effect in this study. In order to make sure that this null result in Study 9 was not, however, an artifact of the recruitment method used, we carried out a direct replication of the study that did not recruit participants via ipoints.co.uk.

Study 10

Method

Participants

An internet sample of 38 males and 62 females, aged between 38 and 62 (median age = 34 years) participated in Study 10.

Design and materials

The study design and materials were identical to those of Study 9.

Procedure

The procedure was identical to that in Study 1.

Results

Following exclusions (criteria as before), there were 48 participants in the neutral condition and 52 in the positive outcome condition. The results of the study replicated those of Study 9, showing a main effect of probability level, $F(2, 196) = 1411.92$, $p<.001$, $MSE = 92.31$, but no effect of the utility manipulation, $F(1, 98) = 0.02$, $p>.05$, $MSE = 249.85$, $eta_p^2 = .000$, or interaction between probability and utility, $F(2, 196) = 2.25$, $p>.05$, $MSE = 92.31$, $eta_p^2 = .022$. Moreover, Figure 3.4 shows that probability estimates tended to be slightly higher in the neutral condition than in the positive outcome condition (as in Study 9). In this study there was, however, a trend for the estimates at the high probability level to be higher in the positive outcome than in the neutral outcome condition.

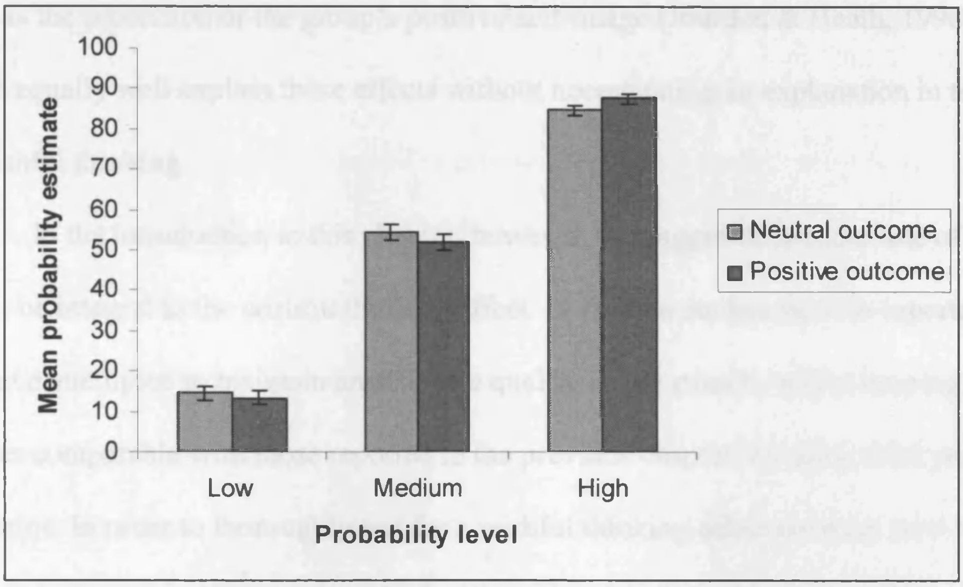


Figure 3.4. Mean probability estimates in Study 10. Error bars are plus and minus 1 standard error.

Discussion

The results of Studies, 8, 9 and 10, especially when considered in conjunction with the results of Bar-Hillel and Budescu (1995), suggest that the wishful thinking effect does not occur under controlled laboratory conditions. Such a result supports the argument that results showing that sports supporters believe their supported team is more likely to win than the opposition (e.g. Babad & Katz, 1991), and that people predict their preferred outcome of a political election (e.g. Babad, 1995; Granberg & Brent, 1983) are the result of an "unbiased evaluation of a biased body of evidence." (Bar-Hillel & Budescu, 1995, p. 100). The results of Babad (1995) and Granberg and Brent (1983), for example, could be explained by reference to the consensus bias which leads people to overestimate the number of people who share their point of view (e.g., Brown, 1982). The results of Babad and Katz (1991) similarly could be attributed to a bias in the knowledge base of sports fans who might subjectively 'know' more of their supported teams strengths than the opposition's. Alternatively, motivational and cognitive factors related to the group characteristic of a sports fan, such as the protection of the group's positive self-image (Jourden & Heath, 1996) could equally well explain these effects without necessitating an explanation in terms of wishful thinking.

In the introduction to this chapter, however, we suggested that the role of affect might be integral to the wishful thinking effect. In the two studies thus far reported, we have attempted to maintain an affective quality in our stimuli, whilst keeping the studies comparable with those reported in the previous chapter, by using third person scenarios. In order to thoroughly test for a wishful thinking effect we must now relax the latter constraint. Whilst it is undoubtedly an empirical question, it is sufficient here to assume that whilst negative events that affect unknown others can still induce significant negative affect in an individual, the same might not be true for positive events. It seems safe to assume that first-person manipulations of utility represent a stronger manipulation than do third-person manipulations. Thus, Study 11 provides a more powerful test of the wishful thinking hypothesis.

Study 11

The aim of Study 11 was to provide a controlled test of the 'wishful thinking' effect using a first person manipulation. Although the use of a first person manipulation reduces the similarity with the studies used in Chapter 2, it does increase comparability with previous experiments that have reported the existence of a 'wishful thinking' effect (e.g. Babad, 1995; Babad & Katz, 1991; Gordon et al., 2005; Irwin, 1953; Irwin & Graae, 1968; Irwin & Snodgrass, 1966; Price, 2000). As has been alluded to in Chapter 2, there are alternative explanations for the results of all the experiments in these papers. The presence of these alternative explanations means that it is not necessary to assume that there is a psychological interdependence between the constructs of probability and utility, constructs which, intuitively and normatively, should be independent (e.g., Edwards, 1962).

We have stressed the potential importance of an affective quality to the positive event being judged in observing a wishful thinking effect. Given their popularity despite their fat and sugar contents, we posited that the winning of a mars bar might provide an affordable positive event, the utility of which was primarily based on affective considerations (see e.g., Cantin & Dubé, 1999; Letarte, Dubé & Troche, 1997; Zanna & Rempel, 1988). So as to confirm this assumption, a pilot test was carried out. 13 females and 7 males first rated the degree to which five cognitive and five affective statements best described their attitude towards mars bars. All these statements were positive,[8] as Study 11 was investigating a 'wishful thinking' effect. We were therefore interested in the basis of participants' *liking* of mars bars. Having rated all ten statements, participants were required to indicate which of the ten statements best described their attitude towards mars bars. 16 participants indicated an affective statement at this stage, whilst four indicated a cognitive statement, $\chi^2(1) = 7.2, p<.01$. Following Cantin and Dubé (1999; see also, Millar & Millar, 1990), we thus concluded that positive attitudes towards mars bars are predominantly affect based. This justified our use of them in Study 11.

In line with our previous studies investigating the existence of an interdependence between utility and probability, Study 11 used the same probability matrices as already introduced. In so doing, the design of Study 11, once again,

---

[8] The cognitive statements were: Mars bars are nutritious / provide energy / are full of vitamins / satisfy hunger / are low in calories. The affective statements were: Mars bars give a pleasant mouthfeel / are palatable / are tasty / are gratifying / are not bland.

closely resembled the experiments reported in Bar-Hillel and Budescu (1995). In 6 out of 7 experiments reported in that paper, Bar-Hillel and Budescu (1995, Study 1) failed to observe any evidence for a wishful thinking effect using visual matrices similar to ours, and a cash based manipulation of desirability. We argue that a cash based manipulation is affectively poor and it is therefore necessary to repeat their study using a more affectively rich reward in the desirable outcome condition.

Method

Participants

94 female and 6 male psychology undergraduates (aged 18-38 years; median = 19 years) at Cardiff University participated in this study in return for course credit.

Design

Study 11 employed a slight change in design from our previous studies investigating the relationship between probability and utility. Outcome utility was again manipulated between participants over two levels, positive and neutral, by means of a cover story. The desirability of an outcome was manipulated by giving participants in the positive condition a mars bar if the critical event occurred (they drew a black straw). Participants in the positive condition therefore read the following text:

> 'We are interested in how likely you believe outcomes to be. In the tin in front of you, the buried end of some of the straws is black. In a minute you will be allowed to pick ONE straw from the tray. If you pick a black straw from the tray you will win the mars bar. The black and white cells in the picture below are distributed in the same proportions as the black and plain straws in the tray. The location of black cells in the picture below does not

however correspond, in any way, with the location of the black straws. Based on the proportion of black cells in this matrix, please estimate the chance that you will pick a black straw from the tray in front of you, and therefore win a mars bar.

Please report this chance as a number between 0 (it is impossible that you will pick a black straw) and 100 (it is an absolute certainty that you will pick a black straw) _______________'

The text that participants read in the neutral condition was identical except that all reference to mars bars was removed.

We recognised that our utility manipulation was subjective and dependent upon our participants liking mars bars. For this reason, participants in the positive condition completed an extra section to the questionnaire, on a separate page, after having made their probability estimate. Participants were asked whether they liked mars bars and how much they liked / disliked them (a bit, moderately, a lot).

The inclusion of different levels of objective probability was important to maintain the generalisability of the study, but the experimental set-up necessitated its inclusion as a between participants, rather than within participants manipulation. Each participant therefore saw one of five probability matrices beneath the cover story.

Materials

Participants were handed a typed version of one of the cover stories with one of the black and white visual probability matrices below it. In front of the participants, as they read the cover story and filled in their probability rating was a circular metal tin with a depth of 10.5 cm and a diameter of 23.5 cm. The tin was a former 'Roses' chocolate tin and still had the corresponding artwork on the side. The tin was filled with wild bird seed to a level 2 cm below the top of the tin. Bedded in this bird seed were 120 plastic drinking straws, 114 of which had black tape around the end which had been buried in the bird seed (a proportion which corresponded to the proportion of black cells to white cells in the most black-dominant matrix). Without extracting them from the bird seed, the straws that had been blackened were indistinguishable from those that had not been.

In the positive outcome condition, a regular sized mars bar was placed in front of the tin, clearly visible to participants throughout the duration of the study. The purpose of this was two-fold. Firstly, participants could see that a mars bar was available to be won and thus participants were more likely to view the information provided as accurate. Secondly, ratings of the desirability of a liked food stuff have been shown to be higher in the presence of the food stuff than in its absence (Cornell, Rodin, & Weingarten, 1989). Thus, having a mars bar in sight throughout the study should increase the subjective value of the mars bar (for those participants who like mars bars), and therefore increase the power of the experimental manipulation.

Procedure

Participants were asked to sit in a chair at a table. Upon the table was the experimental questionnaire with a consent form on top of it, the tin with the straws protruding from it, and if participants were in the positive outcome condition then the mars bar was also visible on the table. Participants were asked to complete the consent form and then continue with the written part of the study in their own time. The experimenter asked each participant to inform him when they had completed the written part. Upon completion of the questionnaire, participants were allowed to pick a straw from the tin. If they were in the positive outcome condition and picked a black straw (100% of participants in this condition) then they were given the mars bar.

Participants who subsequently said that they did not want the mars bar as they did not like them were asked if they would like to take it anyway to give to a friend. Upon completion of the study, participants were thanked and debriefed.

Results

At the first stage of analysis, the estimates of all 100 participants were included in a 5x2 (probability x outcome utility) factorial ANOVA. Participants who responded that they did not like mars bars were not excluded from the first analysis as it was felt that regardless of whether or not an individual liked mars bars, the presence of a reward should still enhance the desirability of the outcome with which that reward is associated. The mean probability estimates of all 100 participants are therefore plotted in Figure 3.5. The probability manipulation was successful, $F(4, 90) = 59.73, p<.001$, $MSE = 182.02$. The utility manipulation did not affect probability estimates, $F(1, 90) = .29, p>.05, MSE = 182.02, eta_p^2 = .003$, nor was there an interaction between desirability and probability, $F(4, 90) = .06, p>.05, MSE = 182.02, eta_p^2 = .003$.

We collected data to ascertain whether participants viewed a mars bar as desirable. Subsequently, participants in the positive condition who reported that they did not like mars bars were excluded from the analysis. The recalculated means are displayed alongside the other means in Figure 3.5. By comparing the means of the positive condition from all participants ('all positive') with those of only those participants who reported that they liked mars bars ('positive with exclusions'), it is clear that there is little difference between these means. A factorial ANOVA comparing the estimates of participants in the neutral condition with those of participants in the positive condition who liked mars bars again showed no effect of outcome utility, $F(1, 81) = .12, p>.05, MSE = 194.11, eta_p^2 = .001$, and no interaction between outcome utility and probability, $F(4, 81) = .12$, $p > .05$, $MSE = 194.11$, $eta_p^2 = .006$.

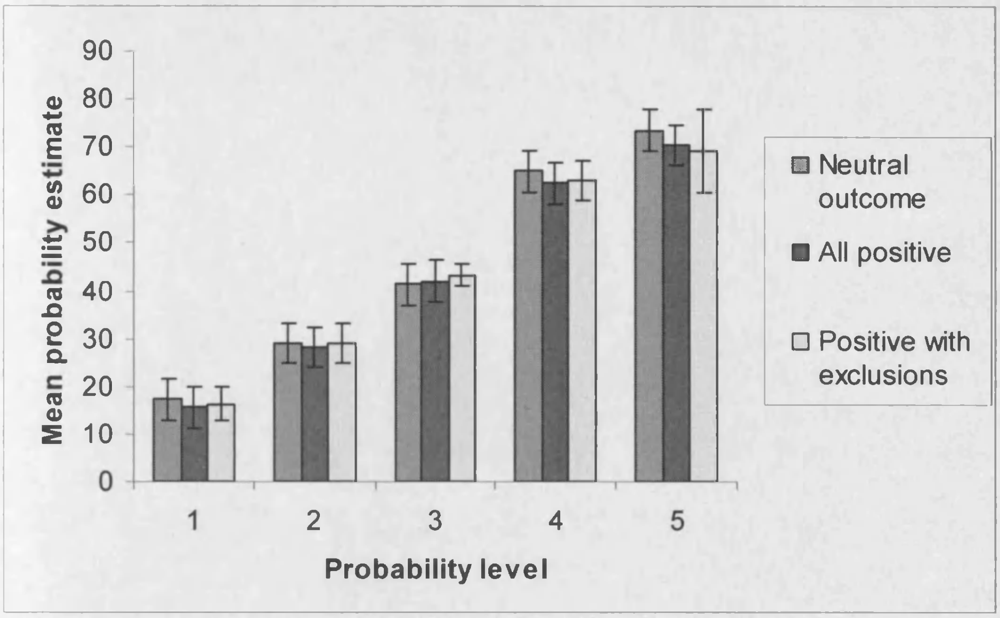


Figure 3.5. Mean probability estimates from participants in the two outcome utility conditions, both with and without exclusions, and the five objective probability levels. Error bars are plus and minus 1 standard error.

Discussion

The null result in Study 11 provides further evidence that people are able to maintain independence between their probability estimates and the utility of the outcomes whose likelihood of occurrence is being estimated. This conclusion is further strengthened by conducting a meta-analysis across Studies 8-11. Following Rosenthal (1991, pp. 73-74), we first computed the effect sizes for the manipulation of outcome utility across all three studies and concluded that there was no significant difference between the studies, $\chi^2(3) = .43$, $p > .50$. Having observed this similarity across the studies, we were able to determine whether there was a significant effect of utility on probability estimates if the data from all four studies were combined (see Rosenthal, 1991, p. 85). In this meta-analysis, no significant effect of utility was observed ($z = 1.02$, $p = .15$). To further test the reliability of the null result observed across all three studies, we used the same meta-analytic procedure to conduct an analysis of the probability estimates only at the low probability level. The rationale for this analysis was that in the real-world, positive outcomes associated with luck are comparatively rare (see Dai et al., 2008; Mandel, 2008). Consequently, these are the probability levels at which participants would most likely experience potential positive outcomes and so might best mirror real-world probability estimates. Once again however, the combination of probability values across all four studies still did not yield a significant effect of outcome utility ($z = 1.24$, $p = .11$). Although this result seems to approach significance, it is worth stressing that the direction of a possible effect at the low probability level across all four studies was in the opposite direction to that predicted by a wishful thinking hypothesis. Given the number of datapoints contributing to these meta-analyses, we are confident in our conclusion that we have observed no evidence of an effect of positive outcome utility on probability estimates in this chapter.

It is well known that conventional hypothesis testing cannot provide support in favour of a null hypothesis, it can only make the alternative hypothesis less likely. Researchers have recently argued for the application of Bayesian statistical tests that can provide support for the null hypothesis over and above an alternative (e.g., Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009). In order to provide further support for our contention that positive utility does not influence estimates of probability, we therefore used the Bayes Factor calculator at pcl.missouri.edu (Rouder et al., 2009) to test the viability of this null hypothesis. The functionality of this calculator is limited to *t*-tests. Given that we have already demonstrated that it is acceptable to combine the data from these studies, we carried out three large *t*-tests on each probability level, combining data from Studies 8, 9 and 10[9]. The results of the *t*-tests were then entered into the Bayes' Factor calculator. It is important to first note that across the low and medium probability conditions the trend in the data suggested neutral outcomes to be estimated as more likely than very good outcomes, whilst in the high probability condition this trend was reversed. As Chapter 2 demonstrated only small effect sizes of utility on probability estimates, we used the unit information prior for the Bayes Factor calculations, as it is more suitable for such situations (see Rouder et al., 2009). The resulting Bayes factors for the low, medium and high probability levels were, respectively: 4.57, 3.10, and 7.74. These figures indicate the degree to which the data were more likely to have been sampled from a single source (i.e., no effect) than from two sources with different means (the alternative hypothesis). All these odds can be thought to correspond to 'substantial' evidence in favour of the null hypothesis (Kass & Raftery, 1995, p. 777). It is also possible to reduce the scale of the prior on effect size. Smaller values of this scaling factor, *r*, correspond to smaller priors as to the size of the effect predicted by the alternative hypothesis (Rouder et al., 2009). By varying the size of *r*, it is possible to provide yet another test of the null hypothesis:

> "When the odds favoring the null approach one from above as the upper limit on the possible size of the effect aproaches zero, the null is unbeatable. When the odds never favor the alternative by more than a small amount for any

[9] The use of different probability levels in Study 11 negated its inclusion in this analysis.

assumed upper limit on the possible size of an effect, considerations of precision and parsimony favor the null" (Gallistel, 2009, p. 441).

The latter result is exactly that observed for all the Bayes Factors reported here, suggesting that the null hypothesis, that 'there is no effect of positive outcome utility on probability estimates in these studies', should be preferred over the alternative hypothesis. The analyses presented thus show no evidence for an effect of positive outcome utility on probability estimates within a controlled laboratory procedure, mirroring the results of Bar-Hillel and Budescu (1995).

It should be noted that the cover stories of Studies 1, 2, 3, 6 and 7 in Chapter 2, where a significant result was observed, all featured a request from a character in the scenario for a probability estimate. In the present chapter, participants were simply asked to estimate a probability (by the experimenter). This raises the possibility that probability estimates were inflated in the severe and high control conditions of Chapter 2 because participants were (implicitly or explicitly) attempting to persuade the character to make the decision that they favoured, by making the negative consequences associated with not making that decision seem more likely. Such an explanation is entirely consistent with an asymmetric loss function approach. Future research should investigate such a possibility, and determine the degree to which a persuasive context is important, and whether such persuasion effects might be observed for outcomes with positive utilities. For example, were the leader of a team of investigative scientists to request a probability judgment when deciding whether or not to undertake the expedition described in Study 9, might a similar effect be observed as in Studies 1, 2, 3, 6 and 7, because participants believe that even a small chance of finding a cure for cancer is worth pursuing. What the results from both

Chapters 2 and 3 do demonstrate, however, is a lack of evidence for a direct biasing effect of utility on estimates of probability.

Chapter Summary

Across three different studies (and one replication) increasing in personal relevance, no effect of positive outcome utility on probability estimates was observed. The failure to observe such an effect is consistent with the results presented in the previous chapter. Although the previous chapter demonstrated that severe outcomes were sometimes rated as more likely to occur than neutral outcomes, this effect was attributed to the presence of an asymmetric loss function in the severe condition rather than being related to the utility manipulation per se. Therefore, our conclusion thus far is that there is no evidence that people's probability estimates are generally biased by considerations of outcome utility. There is, however, one paradigm in which participants have consistently been reported to display optimism, and that is in comparative ratings of their likelihood of experiencing valenced future events (e.g., Weinstein, 1980). Participants' ratings in this paradigm have consistently been interpreted as demonstrating 'unrealistic optimism'. The next chapter addresses this phenomenon.

Chapter 4 - Investigating the True Status of 'Unrealistic Optimism'

Chapter Overview

A robust finding in social psychology is that people judge negative events as less likely to happen to themselves than to the average person, a behaviour interpreted as showing that people are 'unrealistically optimistic'. We propose that the data fail to clearly establish that people are indeed unrealistic. We demonstrate how unbiased responses can result in data patterns commonly interpreted as indicative of optimism for purely statistical reasons. Specifically, we show how extant data from unrealistic optimism studies are plagued by the statistical consequences of sampling constraints and the response scales used, in combination with the comparative rarity of truly negative events. We further describe two new studies supporting these claims. The results of these studies lead us to conclude that the presence of such statistical artifacts means that, despite decades of research, there exists little compelling empirical evidence for the assertion that people are unrealistically optimistic about future life events.

Introduction

Chapters 2 and 3 have been concerned with the issue of whether outcome utility biases estimates of the probability of that outcome. Despite previous claims that utility does bias estimates of probability (e.g. Crandall et al., 1955; Edwards, 1953, 1962; Irwin, 1953; Marks, 1951; Morlock & Hertz, 1964; Pruitt & Hoge, 1965), the present research echoes other recent findings in the literature suggesting that, once all other variables are controlled for, there is no evidence for a direct effect of outcome utility on probability estimates (see also, e.g., Bar-Hillel & Budescu, 1995; Bar-Hillel et al., 2008; Dai et al., 2008; Krizan & Windschitl, 2007; Mandel, 2008). In the real-world, however, a variety of potential mediators, including asymmetric loss functions (see Chapter 2) can lead to an *indirect* biasing effect of utility on probability estimates (e.g., Bar-Hillel et al., 2008; for a review see Krizan & Windschitl, 2007).

A ubiquitous finding related to the present research theme is that (seemingly) people "are often overoptimistic about the future" (Chambers, Windschitl, & Suls, 2003, p. 1343). The underlying phenomenon, referred to in the literature as unrealistic optimism, is that people perceive their own future as being better than the average person's. That is, they rate positive future events as more likely to happen to themselves than the average person and negative events as less likely to happen to themselves than the average person (e.g., Burger & Burns, 1988; Campbell, Greenauer, Macaluso, & End, 2007; Harris & Guten, 1979; Harris & Middleton, 1994; Kirscht et al., 1966; Lek & Bishop, 1995; Perloff & Fetzer, 1986; Weinstein, 1980, 1982, 1984, 1987, 1989a; Weinstein & Klein, 1995; Whitley & Hern, 1991).

Not only is unrealistic optimism seemingly a firmly established empirical phenomenon, it is also deeply embedded in applied work pertaining to risk perception and risk behaviour, as documented by the substantial body of literature relating to the phenomenon within health psychology (e.g., Cohn, Macfarlane, Yanez, & Imai, 1995; Gerrard, Gibbons, Benthin, & Hessling, 1996; Gerrard, Gibbons, & Bushman, 1996; Hampson, Andrews, Barckley, Lichtenstein, & Lee, 2000; Lek & Bishop, 1995; Rothman & Kiviniemi, 1999; van der Velde & van der Pligt, 1991; van der Velde et al., 1992, 1994; Weinstein, 1999, 2000; Weinstein & Klein, 1996; Welkenhuysen, Evers-Kiebooms, Decruyenaere, & van den Berghe, 1996). Here, researchers and practitioners are concerned that people will not take the necessary preventative steps to protect themselves because they underestimate their chances of contracting disease.

It suffices to say, a clear understanding of the psychology of risk perception is essential for effective communication of health information. Consequently, the seeming discrepancy from the results observed using this paradigm and the conclusions from Chapters 2 and 3 needs reconciling. How might such reconciliation come about?

The first possibility is that unrealistic optimism occurs through a mediator, such that it is another example of the indirect biasing of probabilities by utilities (for a review see Krizan & Windschitl, 2007). Indeed, there are a number of candidates for potential mediators of this effect in the literature (for a review see Helweg-Larsen & Shepperd, 2001). Taylor and Brown (1988; see also, Kirscht et al., 1966; Lund, 1925; Zakay, 1996) view the effect as resulting from a self-serving motivational bias designed to protect self-esteem and guard against depression. A second school of thought highlights the importance of cognitive factors as mediators of the effect (e.g., Chambers et al., 2003; Kruger, 1999; Kruger & Burrus, 2004; Price, Pentecost & Voth, 2002; Weinstein, 1980; for a review see Chambers & Windschitl, 2004).

It is, however, our contention that none of the above accounts are necessary to account for the unrealistic optimism phenomenon, as it can readily be accounted for as being a statistical artifact of the methods used in studies demonstrating the phenomenon. In this chapter, we therefore re-examine the assumption that people's optimism is unrealistic. Specifically, we demonstrate that the results of studies demonstrating unrealistic optimism can parsimoniously be viewed as statistical artifacts, as opposed to demonstrations of a genuine human bias, and that responses made by participants in these studies could result from accurate probabilistic knowledge. In short, we argue that there presently exists no satisfactory evidence for the assertion that people genuinely exhibit the so-called unrealistic optimism bias. While future evidence *could*, of course, provide firm support for a genuine optimistic bias, the evidence thus far does not justify the widely held view that people are unrealistically optimistic.

The Statistical Artifacts

> "It is usually impossible to demonstrate that an individual's optimistic expectations about the future are unrealistic. An individual might be quite correct in asserting that his or her chances of experiencing a negative event are less than average" (Weinstein, 1980, p. 806).

Without detailed individual knowledge about our participants, coupled (ideally) with an ability to see into the future, it is impossible to determine whether a specific individual is accurate in stating that they are less likely to experience a given event than the average person. However, it has been assumed that the realism of people's expectations can readily be assessed at a group level. Campbell et al. (2007, p. 1275; see also Bauman & Siegel, 1987, p. 331; Taylor & Brown, 1988, p. 194), for example, state that:

> "on a group level unrealistic optimism is evident when the majority of respondents feel that negative events are less likely to happen to them than the average person."

Both the terms 'majority' and 'average' are ambiguous. The former can refer simply to the largest group (a 'simple majority') or to a group that constitutes more than 50% (an 'absolute majority'); 'average', of course, can refer to a mean, median, or mode. In terms of detecting 'unrealism', simple majorities are useless, as it is readily possible to be the largest group, but nevertheless be below average, whether the average is assessed as the mean, median or mode (see Figure 4.1a). Absolute majorities, however, are only somewhat more constraining: while it is, by definition, impossible for more than 50% of observations to lie below the median, it is easily possible for the average construed as the mean or mode. Only a moderate degree of skew and a limited range of values are enough to give rise to distributions where the absolute majority is above or below the mean (Figure 4.1b), as also noted in Weinstein (1980, p. 809). Moreover, many real-world distributions have this property. In particular, it arises readily for binomial distributions associated with binary outcomes, for example, whether or not a person will experience a given negative life event. Hence, the presence or absence of a majority, at least below the average construed as the mean, will not allow reliable inference about whether or not expectations are unrealistic.

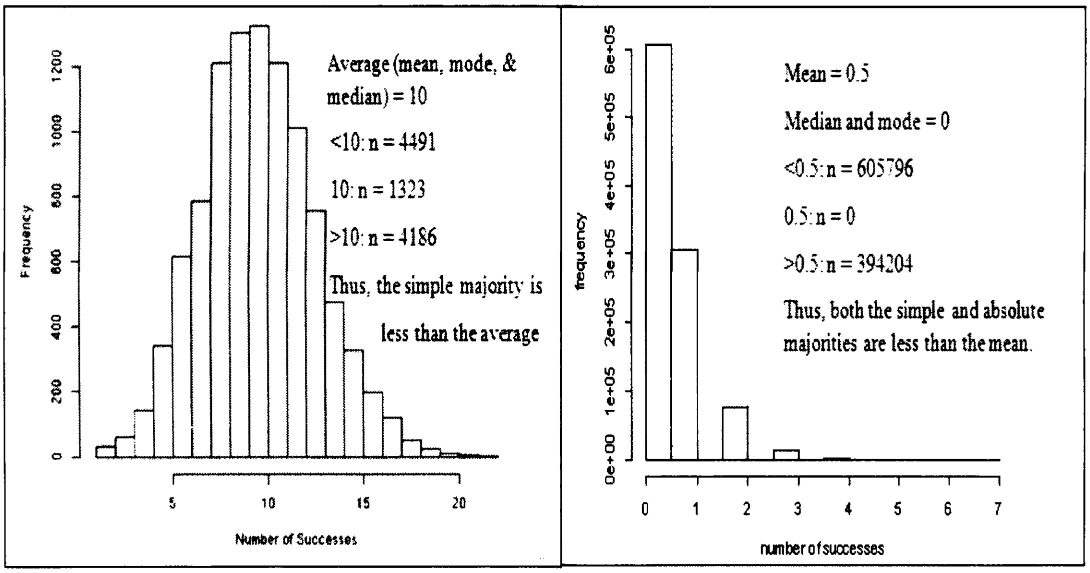


Figure 4.1. Sample distributions in which the majority outcome is less than the average. The histograms represent the outcome distributions of simple simulations of binomial numbers. Figure 4.1a (left panel) shows the results of 10,000 samples of 100 trials of a binomial process in which the probability of success is 0.1. Figure 4.1b (right panel) presents a corresponding simulation result for a binomial process in which the probability of a success is 0.005.

Nevertheless, it has been suggested that further constraints on the size of the majority allow the desired inference. For example, McKenna (1993, p. 39) proposed that:

"When as a group, the *vast* [italics added] majority perceive their chances of a negative event as being less than average then clearly this is not just optimistic but also unrealistic".

The validity of such a conclusion depends on the frequency of the event being judged. The negative events used in unrealistic optimism studies are typically rare. To illustrate, cancer, which is a disease generally considered to be quite prevalent will affect less than half the population, approximately 40% (Office for National Statistics, 2000)[10]. Moreover, contracting cancer is one of the least rare of the negative events typically used in unrealistic optimism studies. In these studies, it is more usual that the type of cancer will be specified (e.g., lung cancer). In Weinstein (1982), there were three cancer items: Lung cancer, skin cancer, and cancer. Lung cancer is the most common form of cancer in men, and third most common in women. However, across both sexes, it is predicted that only 6% of people will contract lung cancer (ONS, 2000).

In order to clarify the impact of event rarity and to aid our subsequent exposition of the statistical problems associated with standard tests of unrealistic optimism, we introduce a simple thought experiment that we will refer back to throughout this chapter.

A Thought Experiment: 'Unrealistic optimism' in Perfect Predictors

In this thought experiment, we assume that people have perfect knowledge. Hence they know (for certain) whether or not they will eventually contract Disease X, which has a lifetime prevalence of 5% (that is, in the course of their lifetime, 5% of

---

[10] All statistics reported are for England and Wales.

people will contract Disease X). As they are perfect, our perfect predictors also know that the prevalence (base rate) of Disease X is 5%.

Experiment

Participants are asked whether they have a chance of contracting Disease X that is smaller than, greater than, or the same as the average person.

Thought process

Participants *know* whether or not they will contract the disease and thus assign a percentage of either 0 or 100% to *their* chance. Participants also *know* the base rate of the disease which is the best answer they can give to the question "What is the chance of the average person…" (see also, Klar, Medding, & Sarel, 1996). Consequently, they assign the average person a chance of 5%. To answer the question posed, participants would compare their chance (0 or 100%) with the average person's chance (5%) to report whether their chance is greater or less than the average person's.

Results

95% of these participants (a percentage presumably sufficiently high to be classified as a vast majority), knowing that their chance of catching Disease X is 0%, will *accurately* report that they are less likely to catch the disease than the average person, whilst just 5% of participants will report that they are more likely to catch the disease than the average person. Crucially, the reports of these perfect predictors cannot (by definition) be unrealistic.

In other words, even a 'vast majority' of people indicating that their chance of experiencing the event in question is less than the average person's cannot guarantee that this group of people has anything other than entirely realistic expectations. For sufficiently rare events, not just the majority, but also the 'vast majority', of people can genuinely have a less than average (mean) chance of experiencing those events, as demonstrated in Figure 4.1b.

In summary, *any* evaluation based merely on the number of people providing an 'optimistic response' relative to the number providing a non-optimistic response is insufficient evidence that a group of people are unrealistic in their reports.

A Different Methodology

The most popular measure for assessing unrealistic optimism was first used by Weinstein (1980). This measure does not ask participants simply to indicate whether their chance of experiencing a given event is greater or less than the average person's. Rather, participants are required to make a comparative, quantitative, response indicating the *degree* to which they are more or less likely to experience an event than the average person. The logic of this approach is simple: Given a sufficiently large sample, the mean value of participants' expressions of optimism and pessimism (with reference to the average person) should be realism. For example, with our perfect predictors, 95% of them have a 5% less than average chance (-5%) of contracting Disease X, whilst 5% have a 95% greater than average chance (+95%). If the members of this population accurately report these percentage differences, the mean of their responses will equal: $(95\% \times -5) + (5\% \times 95) = 0$. Consequently, a population average less than zero demonstrates that at least some of the population are unrealistically optimistic regarding this negative event. Weinstein (1980) found mean responses less than zero for negative events and greater than zero for positive events. Thus, Weinstein's participants seemingly displayed a kind of 'invulnerability bias', or unrealistic optimism.

However, this methodology is vulnerable to statistical artifacts. Specifically, studies using Weinstein's method of comparative responses generally do not use a continuous -100% to +100% response scale. Rather, the response scale typically used in this paradigm is a seven point scale from -3 (chances much less than the average person) to +3 (chances much greater than the average person) (e.g., Covey & Davies, 2004; Klar et al., 1996; Price et al., 2002; Weinstein, 1982, 1984, 1987; Weinstein & Klein, 1995). It is the nature of this attenuated response scale that could be producing the results most commonly interpreted as demonstrating unrealistic optimism, as we shall now demonstrate with further reference to the thought experiment outlined above.

Returning to perfect predictors

In this version of our thought experiment, our perfect predictors are required to make a response on a -3 to +3 response scale regarding their relative chance of catching Disease X, which, once again, has a base rate of 5%. Thus 95% of these participants *know* that they have a slightly lower chance than the average person of catching Disease X (because 0% is 5% less than the 5% average) and hence circle -1 on the response scale.[11] 5% of these participants *know* that they have a much greater chance than the average person of catching this disease (because 100% is 95% greater than 5%) and therefore circle +3 on the response scale. The mean response of our population of perfect predictors is therefore -0.8 and not 0. Indeed, even for a representative sample of just 20 participants, such data would resemble significant 'unrealistic' optimism, $t(19) = 4$, $p = .001$.

---

[11] Alternatively, participants might consider the relative difference, not the absolute difference in risk, in which case they evaluate the ratio of the difference between their risk and the average person's risk and thus consider the distinction between 5% and 0% to be maximal. Consequently, assuming that participants consider their relative risk as a difference score constitutes a conservative assumption.

Scale Attenuation

At the heart of this seeming paradox, where individually unbiased responses lead to a seemingly biased group level response, is the constrained nature of the response scale. The choice of the -3 to +3 response scale was justified in the original unrealistic optimism studies with the following considerations:

> "It emphasizes the comparative aspect of the risk judgments, does not demand unnatural numerical estimates (such as percentile rankings), and, unlike a scale used previously (Weinstein, 1980), *is not vulnerable to a few extreme responses* [italics added]" (Weinstein, 1982, p. 486).

While this might often be a desirable property of scales, the problematic result of the thought experiment above stems directly *from* the scale's invulnerability to a few extreme responses in conjunction with the rarity of Disease X. Only a small proportion of the population will contract the disease, and their responses are necessarily 'extreme'. Had this representative sample of participants been able to use the whole percentage range to indicate their relative chances, the mean response would have been zero. Figure 4.2, however, illustrates that even if the 'worse off' (affected) were always to rate themselves as being maximally more likely to experience an event than the average person (i.e., +5 on a less attenuated -5 to +5 scale; +3 on the -3 to +3 scale), for rare events the average response will be negative. The truncated scale simply does not allow the responses of the two groups to be far enough apart that they can numerically balance out.

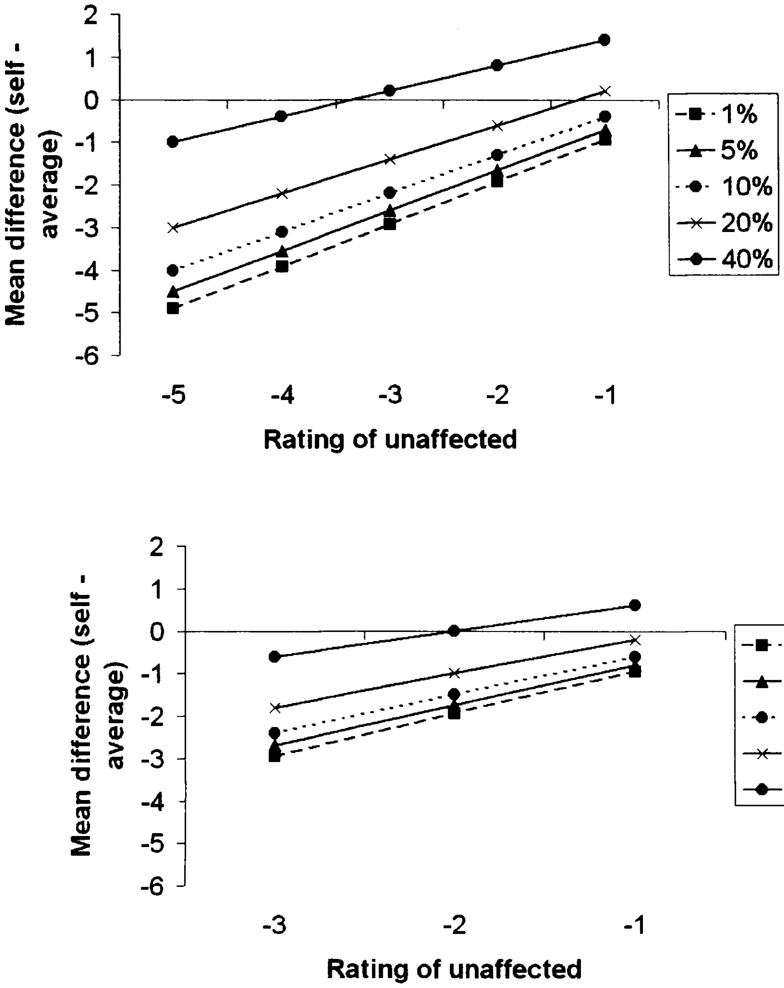


Figure 4.2. The effect of scale truncation on the mean difference score (y-axis) provided by perfect predictors for events of different base rates and for different negative ratings provided by those who will not experience the event (x-axis). Calculations assume that those people who will get the disease report the maximum value on the scale. The top panel shows the effect for a -5 to +5 response scale. The bottom panel shows the effect for a -3 to +3 scale.

Not all studies have used the -3 to +3 response scale. For example, Weinstein (1980) gave participants a 15 point scale with the values "100% less (no chance), 80% less, 60% less, 40% less, 20% less, 10% less, average, 10% more, 20% more, 40% more, 60% more, 80% more, 100% more, 3 times average, and 5 times average" (Weinstein, 1980, p. 809-811). Clearly this scale enables more extreme responses than the typical -3 to +3 scale. However, this is still not enough for our example of people with perfect knowledge about their susceptibility to a disease with a base rate of 5%. The 'worse off' minority who have a 100% chance of contracting the disease would want to state that they are 20 times (i.e., 100/5) more likely than the average person to contract the disease. The 15 point scale still does not, however, allow for such a response. Consequently, it can still give rise to an artifactual effect of seeming optimism, even though the effect will be less pronounced.

That greater scale attenuation leads to (seemingly) greater unrealistic optimism is demonstrated by comparing the two panels of Figure 4.2. Figure 4.2 displays *statistical* optimism (mean difference less than zero) and pessimism (mean difference greater than zero) in samples of perfect predictors for diseases with different base rates and in situations where the 'better off' majority (unaffected by the disease) report different 'less than average' chances. For example, in situations where the unaffected report -1, greater 'optimism' will be observed using the -3 to +3 response scale (bottom panel) than the less attenuated -5 to +5 scale (top panel). The seeming unrealistic pessimism shown for the more common events in Figure 4.2 results from the pessimistic nature of the response of +5 or +3 from the worse off minority for these events.

Directly in line with this is the empirical finding that greater optimism is observed when participants are given an attenuated (-4 to +4) scale than when they are given a larger (-100 to +100) scale (Otten & van der Pligt, 1996). Hence, this finding can be taken as evidence that scale attenuation plays a genuine role in unrealistic optimism results.

However, Otten and van der Pligt (1996) still observed significant optimism using the -100 to +100 scale (with fixed increments of 1). Can this result be reconciled with the statistical artifact hypothesis? For one, extremely rare events will have base rates of less than 0.5%, that is, less than the smallest increment. This will make even this -100 to +100 response scale an *attenuated* response scale. However, there is also a further statistical mechanism through which accurate individual responses might appear optimistic at the group level.

Undersampling of the Minority

Unrealistic optimism studies typically obtain responses from a sample of the population, and not the population as a whole. It is a statistical consequence of binomial distributions that minorities in the population are more likely to be underrepresented than overrepresented in a sample of that population (see e.g., Fox & Hadar, 2006; Hertwig, Barron, Weber, & Erev, 2004; Ungemach, Chater, & Stewart, 2009). Consequently, regardless of the response scale chosen, or the methodology used, the 'worse off' minority (those more likely than the average person to contract the disease) are more likely to be underrepresented in the sample than are the 'better off' majority. If underrepresented, the crucial influence of the positive responses from the worse off minority on the group average will be attenuated, leading to an overall appearance of optimism in the group data.

If we return to our hypothetical example in which all members of the population have perfect knowledge as to whether they will contract a given disease, the mean of these responses (given an unattenuated response scale) will be zero, assuming that responses are obtained from the whole population. If responses are only obtained from a *sample* of that population, then the responses will mean to zero only in the event that the characteristics of the sample match the characteristics of the population. The recognition that the minority are more likely to be undersampled than oversampled from the population makes it more likely that the mean will be less than zero as opposed to greater than zero, again giving the statistical illusion of an optimistic bias. The magnitude of this undersampling can be estimated from distributions such as those shown in Figure 4.1 above: displayed are the results of samples drawn from a population in which a binary outcome (e.g., success/no success reflecting disease/no disease) occurs with a base rate corresponding to the respective probability of success. As the base rate becomes lower (right panel vs. left) the proportion of samples that contain fewer than average 'successes' becomes more and more extreme. Figure 4.3 graphs this excess proportion for different base rates and sample sizes in order to give a further indication of the effect of event rarity and study sample size on the extent to which the less likely outcome (the minority) is undersampled rather than oversampled relative to the population distribution.

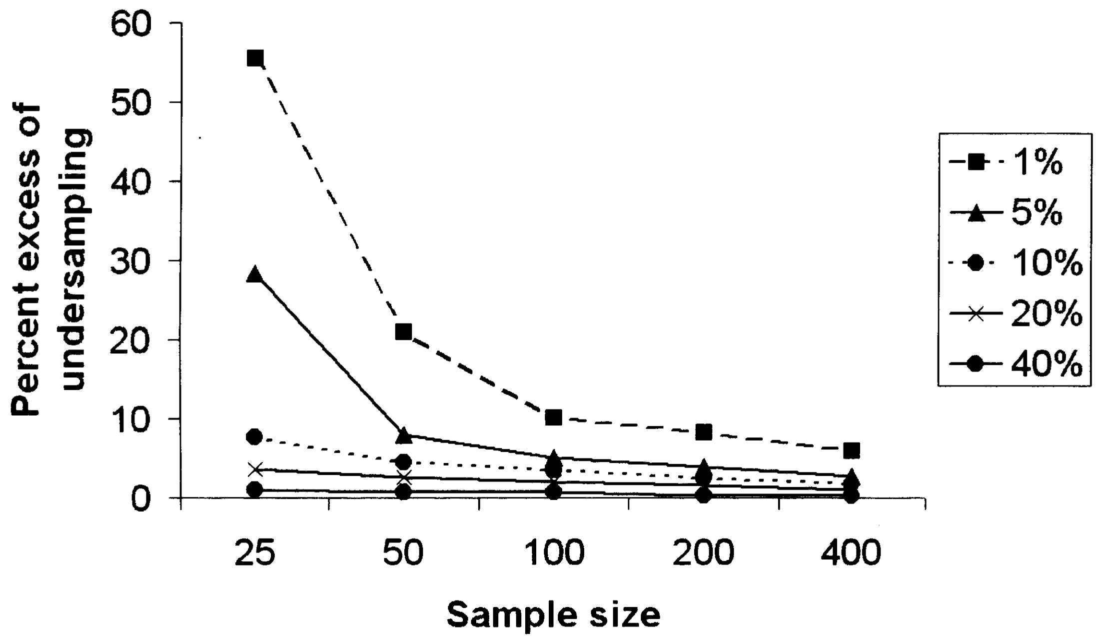


Figure 4.3. The excess of instances in which the minority was undersampled relative to the majority. Graphed are the results for 1 million simulated samples of size 25-400, for 5 different base rates.

As can be seen in Figure 4.3, the effect of minority undersampling decreases as sample size increases. This reflects the law of large numbers which states that as sample size increases, the more representative the sample will be of the total population (Bernoulli, 1713). This gives rise to a testable empirical prediction. If minority undersampling is indeed contributing to the effects observed in unrealistic optimism studies, greater effect sizes should be observed in smaller samples, in which the minority is more likely to be undersampled. It should be noted that *in general* the same relationship between sample size and effect size would be predicted by the publishing bias towards significant results (the 'file drawer problem' [e.g., Rosenthal, 1979; Sterling, 1959]): Studies investigating small effect sizes require large samples in order to attain statistical significance. Thus, phenomena with smaller effect sizes will be associated with studies containing larger samples. However, for research on a specific, single phenomenon such as unrealistic optimism, publishing bias against null results will not exert the same pressure. Studies on a single phenomenon (employing similar methodologies) *should* find similar effect sizes, simply because they are measuring the 'same thing', and, conceptually, measures of effect size seek to provide a measure of efficacy that is (largely) independent of sample size (e.g., Keppel, Saufley, & Tokunaga, 1992). Hence there is no general reason to expect a clear correlation between sample size and effect size across studies that all investigate unrealistic optimism. Therefore, the presence of any such correlation would be indicative of some other, underlying effect.

We found nine studies in which all the information was readily available to compute an estimate of the effect size (r), calculated from the t statistic (see Rosenthal, 1991, p. 19) [12]. The effect size was calculated for each study included in this meta-analysis and correlated with study sample size. A significant negative correlation was observed, $r(9) = -.69, p<.05$ (2-tailed) ($r_{\text{adj}} = -.64$) (see Figure 4.4). Of

---

[12] These studies were: Burger & Burns (1988); Campbell et al. (2007); Lek & Bishop (1995); Otten & van der Pligt (1996, Study 1 and Study 2); Pyszczynski, Holt, & Greenberg (1987); Regan, Snyder, & Kassin (1995); Weinstein (1980); Weinstein & Klein (1995). Where appropriate, the analyses used data only from the control group.

course, this meta-analytic correlation is based only on a small number of studies. This correlation, which accounts for 41% (r^2 value adjusted for the small sample size [see Howell, 1997, p. 240]) of the variance in effect sizes across the nine studies, does however lend some support for the potential biasing effect of undersampling the minority in unrealistic optimism studies.

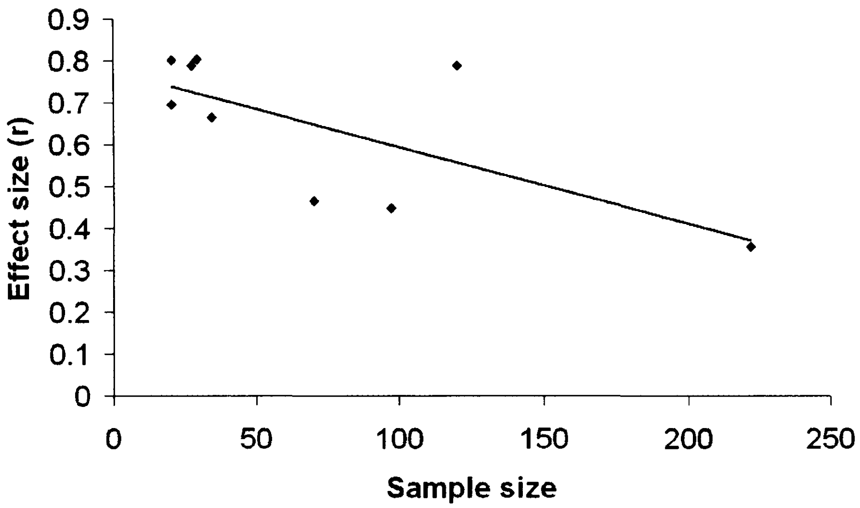


Figure 4.4. The relationship between study sample size and the effect size of unrealistic optimism, based on a meta-analysis of nine published studies.

Real World Feasibility of the Statistical Artifacts

Thus far, we have demonstrated the conceptual frailty of the standard comparative test by demonstrating how seeming unrealistic optimism could be observed for perfect predictors. Furthermore, we have also presented initial empirical evidence for the statistical mechanisms proposed by showing that the effect size in published studies using this technique is moderated by both scale size (Otten & van der Pligt, 1996) and sample size (meta-analysis above). Of course, the participants in these real studies are not perfect predictors. However, complete knowledge of the future is not required, and we demonstrate in the following how the same statistical effects can arise given only partial (but again, intuitively, realistic) knowledge of the future. People's knowledge of both their personal risk and the general, average risk (base rate) stems from a variety of sources: In addition to accessible healthcare (where a doctor might, for example, inform an individual of their general level of health), public information campaigns, personal experience and family history all provide informative cues regarding the chance of contracting a certain disease. In fact, regarding personal risk, the strongest predictor is typically also the most accessible, given that "for many common diseases, having an affected close relative is the strongest predictor of an individual's lifetime risk of developing the disease" (Walter & Emery, 2006, p. 472). For example, a longitudinal study investigating the relationship between family history and risk of coronary heart disease illustrates that relative risk is substantially different for men with and without a family history of heart disease. Hawe, Talmud, Miller and Humphries (2003) reported that men with a family history of heart disease were 1.73 times more likely to suffer a 'coronary heart disease event' than men without a family history. Added to this, smokers with a family history were 3.01 times more likely to suffer a 'coronary heart disease event' than non smokers with no family history. Were an unrealistic optimism study conducted with a sample in possession of this knowledge (because they had heard about this study, for example), what would the results look like?

Of the 2,827 males included in Hawe et al.'s (2003), 6.6% had a coronary event during the follow-up period. Consequently, the average risk with which people *should* compare themselves in a classic unrealistic optimism study would be 6.6% (see also Klar et al., 1996). 1,000 males answered "yes" to the question: "Has any person in your family ever had a heart attack" (Hawe et al., 2003, p. 99) whilst 1,827 answered "no." In the following, we will simulate the responses of participants from these two different groups on a -3 to +3 unrealistic optimism scale under the simplification that this is the only evidence these individuals have for their chances of suffering a heart attack. Given that 5.3% of males who did not have a family history of heart disease suffered a heart attack, it would be reasonable (and realistic) for all 1,827 to answer "-1" (i.e., 'I am slightly less likely than the average person to suffer a heart attack'). 9% of the males who did have a family history of heart disease suffered a heart attack. It therefore seems reasonable (and realistic) for all 1,000 of these males to answer "+1" (i.e., 'I am slightly more likely than the average person to suffer a heart attack'). An increase in risk from 6.6% to 9.0% does not seem to merit a response higher than this on the scale (though clearly how people believe they should convert such a relative risk onto such a response scale is an empirical question, see also Footnote 10). However, assuming that these males did respond in this way, what would be the result of the statistical analysis of this hypothetical unrealistic optimism study? The mean response equals -0.29 and a single sample *t*-test confirms that this is significantly less than zero, $t(2826) = 16.26, p<.001$. Thus, such responses would be interpreted by an unrealistic optimism researcher as demonstrating unrealistic optimism. Moreover, a representative sample of only 60 people from these 2,827 would give rise to a significant effect of seeming unrealistic optimism, with 39 reporting "-1" and 21 reporting "+1", $t(59) = 2.42, p<.02$. Once again, however, the responses made by each *individual* seem perfectly realistic. The observed effect stems entirely from the rarity of the event and the low discriminability of an attenuated response scale. This again demonstrates the unsatisfactory nature of using group data to infer a bias at the level of the individual. In the above example it would be hard to argue that any of the individuals' responses were biased, but the group level results suggest exactly that.

More generally, any individual can have some knowledge reflecting a disease's base rate and, typically, some personal information that increases or decreases their own likelihood of contracting the disease. The rational person should combine these two pieces of knowledge to update their degree of belief in their chance of contracting the disease. We shall demonstrate that rational individual responses on a -3 to +3 response scale can easily result in seemingly unrealistic optimism at a group level, even on the basis of a test result with extremely limited diagnosticity. Diagnosticity is captured in the ratio between a true positive 'test' result, $P(e|h)$, and a false positive 'test' result, $P(e|\neg h)$ (where e is a positive test result, h is contracting the disease, and the negation symbol '$\neg$' denotes the complement, that is, 'not e' or 'not h'). In this scenario, we will assume that this ratio is only 1.5:1 (and the same ratio is assumed to hold for negative 'test' results). Thus, $P(e|h) = .6$ and $P(e|\neg h) = .4$. Consider again the case of lung cancer which has a base rate of 6% (ONS, 2000). Equations 4.1 and 4.2 (Bayes' Theorem) show the normative updating of belief in contracting a disease, given a positive 'test' result, $P(h|e)$, and a negative 'test' result, $P(h|\neg e)$, respectively. In these equations, $P(h)$ represents the prior degree of belief that Disease X will be contracted, which, if people are rational and possess accurate knowledge, would equal the disease base rate (see also, Klar et al., 1996).

$$P(h \mid e) = \frac{P(h)P(e \mid h)}{P(h)P(e \mid h) + P(\neg h)P(e \mid \neg h)} \qquad \text{(Equation 4.1)}$$

$$P(h \mid \neg e) = \frac{P(h)P(\neg e \mid h)}{P(\neg h)P(\neg e \mid \neg h) + P(h)P(\neg e \mid h)} \qquad \text{(Equation 4.2)}$$

The proportion of people in the population who will receive a positive or negative test result is given by Equations 4.3 and 4.4 respectively.

$$P(h)P(e \mid h) + P(\neg h)P(e \mid \neg h) \qquad \text{(Equation 4.3)}$$

$$P(h)P(\neg e \mid h) + P(\neg h)P(\neg e \mid \neg h) \qquad \text{(Equation 4.4)}$$

For the ratio of true results to false results outlined above, and the base rate of 6%, these equations mean that, overall, 41% of people should rate their chance of contracting lung cancer as 8.7% and 59% of people should rate their chance as 4.1%. As with the previous examples above, it is not clear how participants should translate these figures onto a -3 to +3 response scale. However, the deviations from the base rate seem comparable for both those receiving a positive test result and those receiving a negative result: 8.7% vs. 6% for the 'worse off' and 4.1% vs. 6% for the 'better off'. Consequently, 41% of responses of +1 for the 'worse off' might rationally be combined with 59% of responses of -1 for the 'better off', resulting in an average response of -0.18. Even on the basis of such a non-diagnostic test result, significant 'optimism' would be observed in a representative sample of 115 participants, $t(114) = 1.98$, $p=.05$. Once again, seemingly rational responses at the individual level resemble a bias at the group level on such an attenuated scale. Moreover, minority undersampling would only serve to further exaggerate this effect.

As potential evidence against our claim that people have access to and might make some use of knowledge about risk factors stands Weinstein's (1984) study. Weinstein collected data on actual risk factors in addition to people's relative risk judgments. Weinstein reported that "associations between risk judgments and relevant risk factors were often weak or nonexistent" (p. 446). However, this summary assessment seems difficult to sustain: For seven out of the ten events he considered, at least one of the events' risk factors correlated significantly with absolute risk judgments for that event. For one of the three events where no relationship was observed (automobile accident injury), no comparative optimism was observed, whilst the other two events were suicide and mugging, events for which a significant correlation with a risk factor *was* observed subsequently in Weinstein (1989b). Hence there seems ample evidence of sensitivity to risk factors in Weinstein's data, despite the fact that the strength of the risk factor-risk judgment relationship might have been attenuated through two mechanisms. Firstly, Weinstein's analysis could not take into account any potential interplay between different risk factors in influencing personal risk; reported are only pairwise linear correlations for all risk factors, with no consideration of potential interactions between different factors. Secondly, pairwise correlations between risk factors and risk judgments could be attenuated by the rarity of negative events. For rare events, only a minority of people will have strong risk factors, by definition. Thus, a lack of variability in the risk factor data will reduce the likelihood of observing strong correlations between risk factors and risk judgments (a similar observation is made in van der Velde et al., 1992, p. 24).

Furthermore, general studies on risk perception have reported results suggesting that people's estimates of personal risk *are* grounded in an objective reality. These studies report that the more risk behaviours people engage in, the more vulnerable they rate themselves to negative consequences resulting from those behaviours (Cohn et al., 1995; Gerrard, Gibbons, & Bushman, 1996; Martha, Sanchez, & Gomà-i-Freixanet, 2009). Moreover, Gerrard, Gibbons, Benthin, and Hessling (1996) used a longitudinal design to demonstrate that change in risk behaviours predicted the corresponding change in ratings of vulnerability. This research suggests that people can and do recognise those factors that place them more at risk than others for certain negative events.[13]

---

[13] The only other exception came from Bauman and Siegel (1987) who reported that 83% of the gay men in their sample who engaged in sexual practices that put them at high risk for contracting AIDS (66 men in total) rated the risk of their sexual practices (with regards to contracting AIDS) as 5 or less on a 10 point scale (on which 10 indicated most risky). 85% of these men reported engaging in at least one practice which they believe reduced the risk of AIDS, but which objectively made no difference. Consequently, the underestimation of risk reported in this study might be a result of an accurate risk assessment based on inaccurate knowledge, rather than reflecting systematic optimism. In addition, Bauman and Siegel did not include any questions relating to participants' knowledge about their sexual partners. Any men who engaged in high risk sexual practices with a partner who they *knew* to be HIV

In summary, there is evidence to suggest both that people have access to sufficient knowledge and are sufficiently sensitive to it in their judgments of risk for the statistical mechanisms identified to 'bite' in practice.

Moderators of Unrealistic Optimism

We have argued that previous evidence of unrealistic optimism might merely be a statistical artifact. Such a contention might, however, seem difficult to uphold in light of detailed understanding of the phenomenon in terms of its moderators[14]. It is known that event frequency, specificity of the comparison target, experience with the event, event controllability, and mood/anxiety of the participant all affect the degree of unrealistic optimism. In this section, we will argue that none of these known moderators conflict with the statistical artifact hypothesis.

Event Frequency

It is well-established that unrealistic optimism decreases as the frequency of the event increases (e.g., Chambers et al., 2003; Harris, Griffin, & Murray, 2008; Kruger & Burrus, 2004; Weinstein, 1980, 1982, 1987). For example, Weinstein (1982, 1984, 1987) found seemingly unrealistic optimism in participants' judgments of their likelihood of contracting both lung and skin cancer, but no optimism for cancer in

---

negative would be quite accurate in reporting the riskiness of these activities as low. In the absence of such a question, it is difficult to interpret Bauman and Siegel's results.

[14] We follow Helweg-Larsen and Shepperd (2001) in our use of the term 'moderator' by using it to refer to variables that have been shown to produce "differences…in people's optimistic bias reports" (Helweg-Larsen & Shepperd, 2001, p. 75).

general.[15] The dependence on event frequency is itself at the heart of the statistical artifact hypothesis, as is evident from Figures 4.2 and 4.3. Given its prevalence of 40% (ONS, 2000), cancer is common enough that a statistical 'unrealistic optimism' effect will rarely occur (see Figures 4.2 & 4.3). Hence, that unrealistic optimism is moderated by event frequency provides direct support for the statistical artifact hypothesis.

Specificity of the Comparison Target

The degree of unrealistic optimism decreases as the target with whom participants are comparing themselves becomes more specific (Burger & Burns, 1988; Harris & Middleton, 1994; Klar et al., 1996; Perloff & Fetzer, 1986, Regan et al., 1995; Whitley & Hern, 1991; Zakay, 1984, 1996; see also Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995).[16] We have assumed in the preceding that the judgments people make of their own risks are qualitatively different from those they make for the average other. When assessing their own chances of experiencing a negative event, people are estimating a probability about a singular event (an epistemic probability). However, when assessing the chances of the average person experiencing the event, they estimate a frequentist probability which relies on distributional statistics, namely, the base rate, that is, what percentage of people contract cancer (see also Klar et al., 1996). As the comparison target is made more specific, the judgments between self and the target become more consistent, for they

---

[15] An exception is Price et al. (2002) who reported evidence for an optimistic bias about the chance of contracting cancer in half of the conditions in their two studies.

[16] The only exception was a study by Hoorens and Buunk (1993) that found little effect of the specificity of the comparison target.

are now able to estimate an epistemic probability for the unique event of this other, single, person contracting cancer. Given the assumption that people are estimating their risks of experiencing *rare* events in these studies, once again it is probable that the likelihood of an individuated comparison target experiencing a negative event will be less than 'the average person'; so less relative optimism *should* be observed. This proposal is further supported by Helweg-Larsen and Shepperd's (2001) extensive review, which established that the moderating effect of the comparison target affected unrealistic optimism through the risk associated with the target comparison rather than the risk associated with the self.

The same conceptual difference between judgments about the self and judgments about the average person can also explain another finding, namely that, overall, comparative judgments are better predicted by judgments of self-risk than judgments of the average peer's risk (e.g., Chambers et al., 2003; Kruger & Burrus, 2004; Price et al., 2002; Rose, Endo, Windschitl, & Suls, 2008). Together with the moderating effect of event frequency, this finding has been used to support the egocentrism account of unrealistic optimism, which posits that people's comparative judgments are predominantly based on their own absolute chances of experiencing an event with an insufficient consideration of the chances of others (Chambers et al., 2003; Klar & Giladi, 1999; Kruger, 1999; Kruger & Burrus, 2004; Weinstein & Lachendro, 1982).[17]

---

[17] One difficulty for this account lies in recent evidence by Price, Smith and Lench (2006) who found that comparative ratings between the self and the average member of a group were reduced when perceptions of the average member's chance of experiencing an event increased, under circumstances where individual risk was held constant. This clearly indicates sensitivity to the average at least in some circumstances.

However, there are, once again, purely statistical reasons why group risk might be a less powerful predictor. If indeed the average person's risk is assessed by using raw statistical knowledge to make a distributional judgment, whereas a singular risk judgment is made for an individuated person such as the self then it is unsurprising that people's comparative judgments are better predicted by self risk judgments than judgments of the average person's risk: there is likely to be greater variability in singular judgments than distributional judgments, which usually increases predictive power (Howell, 1997, p. 266).

Experience with the Event and Event Controllability

Unrealistic optimism has been shown to decrease both as people's experience with the event increases and as the perceived controllability of the event decreases (e.g., Harris et al., 2008; Helweg-Larsen, 1999; van der Velde et al., 1992; Weinstein, 1980, 1982, 1987, 1989b; Zakay, 1984, 1996). Helweg-Larsen and Shepperd (2001) demonstrated that both these factors influenced estimates of personal risk, rather than estimates of the comparison target's risk. Such a finding makes sense as both experience and controllability can be considered 'sources of knowledge'. Consequently, given that family history will increase experience with a disease as well as increasing susceptibility to it (Walter & Emery, 2006), it can be expected to increase ratings of personal susceptibility whilst not changing perceptions of the target person's susceptibility, thus making a relative response appear less optimistic.[18] Controllability also affects knowledge to the extent that people make efforts to avoid

---

[18] Of course, were you to have knowledge that the target person had a family history of the disease (for example), then this knowledge would affect ratings of their susceptibility and thus increase the optimism of your response.

undesirable but controllable events. Self-knowledge of one's endeavour to avoid the event will increase the seeming optimism of one's relative responses.

Typically, people have been faulted for not sufficiently taking into account protective measures made by the 'average' person (e.g., Chambers et al., 2003; Kruger & Burrus, 2004; Weinstein, 1980; Weinstein & Lachendro, 1982). However, it is rational to assume that the event base rate (average person) both includes people who do and people who do *not* take protective measures. To see this one need only consider the way that the base rate in the 'Real world feasibility of the statistical artifacts' section above includes both people who receive a positive test result and people who receive a negative test result because the test evidence is only probabilistically related to the disease. Given that the base rate is comprised of both people who do *and* people who do not take protective measures, those who do take protective measures *are* in actual fact necessarily less at risk than the 'average person' (base rate).

For both experience and controllability, their impact is based on the fact that they provide sources of knowledge with which individuals can update their estimates of personal risk in the way outlined above; that is, they are analogous to test results. Experience of a disease means that people will know more about its causes (including, frequently, knowledge of family history), whilst event controllability provides people with knowledge of whether they do or do not take protective measures. Both allow people to differentiate their personal susceptibility from the average (base rate). Consequently people move further towards being perfect predictors, which amplifies the effects of scale attenuation. This increases the likelihood of a statistical illusion of unrealistic optimism.

Finally, in addition to being a source of knowledge, controllability will have a separate impact through its influence on event frequency. The controllability of a negative event is likely to reduce its base rate because people will tend to take protective measures to avoid it. Hence, the moderating effect of controllability is only interpretable once event frequency has been controlled for. Zakay (1984), for example, observed significant interactions between event valence and controllability in comparative responses. It is clear from his data, however, that these effects are readily explained with reference to the event's perceived frequency, which is lower for controllable negative events and higher for controllable positive events, than it is for their uncontrollable counterparts.

Mood and Anxiety

Responses are typically less optimistic when a negative mood is induced in individuals (e.g., Abele & Hermer, 1993; Drake, 1984, 1987; Drake & Ulrich, 1992; Salovey & Birnbaum, 1989) and unrealistic optimism is frequently *not* observed in dysphoric individuals (so-called 'depressive realism') (e.g., Alloy & Ahrens, 1987; Pietromonaco & Markus, 1985; Pyszczynski et al., 1987). This has led to the speculation that optimistic illusions "may be adaptive for mental health and well-being" (Taylor & Brown, 1988, p. 193). Importantly, Helweg-Larsen and Shepperd (2001) demonstrated that mood and anxiety influence the degree of optimism in people's judgments predominantly via personal risk estimates: people are less optimistic about their own future when they are in a negative mood. This seems unsurprising (see e.g., Wisco, 2009) and, crucially, does not imply that people are more *realistic* on these occasions. Such a conclusion is warranted only if 'normal' levels of optimism are indeed unrealistic, which is the very assumption challenged in this chapter. Unless this assumption can be independently supported, the moderating effects of mood demonstrate only that people in a negative mood are more negative about future life events; they do not identify who is more or less realistic, and it seems equally possible that dysphoric individuals are overly pessimistic. The same argument, finally, applies to other individual difference moderators of the effect, such as anxiety and defensiveness (see e.g., Harris et al., 2008), as well as cross-cultural results, which have generally found less 'optimism' in Eastern cultures (e.g., Chang, Asakawa, & Sanna, 2001; Heine & Lehman, 1995).

Summary

What is known about the moderators of unrealistic optimism either directly supports, or is entirely compatible with, the possibility that 'unrealistic optimism' is solely a statistical artifact. Thus, the identification of such moderators in the unrealistic optimism literature does not nullify the arguments presented here, which question the true status of unrealistic optimism. Rather, the moderators themselves might be bi-products of the statistical artifacts outlined.

A Critical Test of the Statistical Artifact Hypothesis

As we have seen so far, the rare nature of negative events plays a critical role in producing what is potentially an illusion of unrealistic optimism at a group level. Furthermore, under the statistical artifact hypothesis, the rarer a negative event the greater is the degree of seeming optimism that should be seen, and, as just noted, this relationship has been observed in past studies (e.g., Chambers et al., 2003; Kruger & Burrus, 2004; Weinstein, 1980, 1982, 1987).

Our argument thus far has focussed on people's estimates of negative events, as these constitute the bulk of the unrealistic optimism literature. However, the same statistical mechanisms should apply to judgments of the chance of experiencing *positive* events, on the reasonable assumption that very positive events, like very negative events, are rare. Again, the low base rate of extremely positive events implies that most people will not experience the event in question. For positive events, however, this failure constitutes a bad thing, not a good thing. Hence, the statistical mechanisms introduced above that push the group response towards the 'majority' outcome will result in seeming *pessimism* for positive events. By definition, this is the opposite of what should be found if people were genuinely over-optimistic about their futures. Consequently, while the unrealistic optimism and statistical artifact hypotheses make the same predictions for negative events, they make opposite predictions for positive events.

Unfortunately, studies investigating the possibility of unrealistic optimism for people's estimates of positive events are far fewer than those investigating negative events. The evidence from those that have included positive events is also much more equivocal than it is for negative events (e.g., Chang et al., 2001). Whilst some studies report pessimism (e.g., Moore & Small, 2008), a number of others have reported optimism, such that people view themselves as more likely than the average person to experience positive events (e.g., Campbell et al., 2007; Regan et al., 1995; Weinstein, 1980). However, the statistical artifact hypothesis only predicts unrealistic pessimism for rare events. For positive events that are relatively common, the reverse logic applies. For common events, the chance of *not* experiencing them constitutes the rare outcome. Scale attenuation and minority undersampling thus make it more likely that the average comparative response for common events will be positive, a result interpreted as pessimism for negative events and optimism for positive events. Thus, studies that have observed pessimism for rare positive events but optimism for common positive events (Chambers et al., 2003; Kruger & Burrus, 2004) provide direct support for the statistical artifact hypothesis.

Moreover, the positive events in those studies which have largely found optimism are arguably not rare. Weinstein's seminal (1980) paper, for example, used positive events (p. 810), such as "Owning your own home" and "Living past eighty", which were far less rare than the negative events, and the statistical artifact hypothesis would not necessarily have predicted pessimism for them. This is supported further by Weinstein's finding that the perceived probability of the event was the single biggest predictor of participants' comparative judgments for positive events such that greater comparative responses (interpreted as greater 'optimism') were displayed the more prevalent the positive event was perceived to be.

Perhaps as a result of the practical implications of the unrealistic optimism phenomenon for negative events, particularly in health psychology, very few subsequent studies have further addressed unrealistic optimism in positive events. Some have used very similar materials to Weinstein (1980) (Pyszczynski et al., 1987; Regan et al., 1995) and, consequently, the same argument is levelled against them. Although frequency information was not directly collected in the study of Zakay (1984), if it is inferred from ratings of self and other's chances, then his results suggest optimism for frequent positive events and pessimism for rare positive events, precisely the pattern predicted by the statistical artifact hypothesis. Zakay (1996) also reports results in which the most prevalent positive events yield unrealistic optimism, whilst the least prevalent demonstrate pessimism.[19] More recently, Campbell et al. (2007, p. 1277) used positive events such as "keeping in touch with family" and "downloading music" when using the internet (events that are anything but rare for people who use the internet and consider these events to be positive) and found optimism. This preponderance of common positive events in unrealistic optimism research was also noted by Hoorens, Smits and Shepperd (2008) who concluded that "researchers have

[19] The response scale used in Zakay (1996) (-100 to +100 for both self and others' chances) makes it difficult to determine the actual perceived frequency of the events.

particularly sampled common desirable events and rare undesirable events, the very kinds of events that are likely to produce comparative optimism" (p. 442). Their own study sought to overcome this limitation by having participants self-generate events; however, the most frequently generated event types in their study were again "variations on themes that typically appear in studies involving experimenter-generated lists of events" (Hoorens et al., 2008, pp. 445-446).

In summary, within the unrealistic optimism literature there is far less evidence with positive events, and it is unclear that the sometimes observed optimistic responses for positive events resulted from anything other than their statistical properties – namely that they were much more prevalent than the negative events studied. Because positive events can distinguish directly between a genuine optimistic bias and our statistical artifact hypothesis, further tests are essential. Hence we conducted a replication of the 'classic' unrealistic optimism study, but maintained equal focus on positive and negative events. The key question was whether we would observe optimism or pessimism for rare positive events.

Study 12

Study 12 was a replication of a 'classic' unrealistic optimism study, with an equal focus on positive and negative events.

Method

Participants

102 female undergraduates, aged 18-24 years (median age = 19), from Cardiff University participated in this study in return for course credit or monetary payment. Only females were used in order to reduce unnecessary variability resulting from gender differences in the desirability of, and susceptibility to, different events.

Materials

A 282 item questionnaire was developed. The questionnaire asked seven types of question about 40 life events. The remaining two questions were open-ended qualitative questions asking: "Please list any bad things that you think are more likely to happen to you than the average female student in your year" and "Please list any good things that you think are less likely to happen to you than the average female student in your year." The purpose of these questions was to determine whether people are optimistic about all aspects of their life.

Table 4.1 (see results section) lists the future life events that were chosen for inclusion in this study. The majority of these items were taken from Weinstein (1980). 26 items were taken directly from those described in Weinstein (1980, p. 810), and a further 12 were adapted from the original 42 items in Weinstein (1980)[20] in order to update them, remove any ambiguity, ensure their relevance for UK undergraduate students in the year 2008,[21] and, most importantly, create rarer positive events (for example, 'living past 80' was replaced with 'living past 90'). We also added two, putatively rare, positive events not included in Weinstein (1980): 'Marry a film star' and an extra level of starting salary such that participants were asked about three levels of starting salary, as opposed to two in the original study.

---

[20] Four of Weinstein's original items were left out. These were: "Dropping out of college", this was to reduce any extra variance introduced as a result of participants being both first and second year students, "Decayed tooth extracted" and "Having gum problems", as such events may not be *future* events for some of the sample, and "attempting suicide", which was removed for ethical reasons.

[21] Note that this study was completed in February and March (2008), before the onset of the economic crisis.

The questionnaire was ordered into seven blocks with each block containing 40 questions, such that each block asked a specific question about each life event. Four of these question blocks, concerning relative chance, event controllability, event frequency and event desirability were theoretically motivated by either the unrealistic optimism or statistical artifact hypothesis. Three additional question blocks, concerning event importance, event desirability to the average person, and number of steps taken to approach/avoid the event relative to the average person were included for exploratory reasons. As the theoretically motivated questions were able to sufficiently answer the research question, the three 'exploratory' blocks will not be discussed further. At the beginning of the first block participants read:

> 'In this experiment you will be asked to estimate your chance of experiencing different events in your life. For each event, please judge your chance of experiencing it, RELATIVE TO THE CHANCE OF THE AVERAGE FEMALE STUDENT IN YOUR YEAR.
>
> Please answer using the numerical scale, where a response of 0 means that you think your chances of experiencing the event AT SOME STAGE IN YOUR LIFE are the same as the average female student's AT SOME STAGE IN THEIR LIFE, a response of -5 means that you think your chances are much less than the average student, and +5 means you think your chances are much more than the average student.'

The 40 questions in this block were then phrased and responses scales designed as in the example below:

'Compared with the average female student in your year, how likely do you think you are to like the job you do after you graduate from university? (Please circle)[AA]

-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5

←------------Less likely---←------Same chance----→---- More likely---------------→'

This 11-point scale is less attenuated than the -3 to +3 response scale most typically used in unrealistic optimism studies. Consequently, our replication is a conservative test of the statistical account for previous unrealistic optimism.

Block 2 required participants to 'indicate how desirable different life events are to you.' An example question, and related response scale is:

'How desirable is it to you to have a starting salary of more than £20,000? (Please circle) [ID]'

-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5

←-------Undesirable-------------←--------------→------------ Desirable------------→

The perceived controllability of events was collected in Block 3, which required participants 'to estimate the degree to which certain life events are under your control.' An example question and its response scale are:

'How much control do you think you have over whether you will visit the Amazonian rainforest? (Please circle) [NF]

0 1 2 3 4 5 6 7 8 9 10

←---------Uncontrollable-----←---------------------→-----Controllable--------------→'

Block 4 required participants to 'estimate the number of your peers who will experience different events in their life. For each event, please estimate how many women out of 100 average female students in your year will experience it.' This response comprised the subjective frequency of the event. An example question is:

> 'Out of 100 female students in your year, how many do you think will have a heart attack before they're forty? [GG] _______________'

Design

A within-participants design was employed. Within each question block there were four potential orderings of the life events, and in each ordering participants rated positive and negative events alternately and similar questions (e.g. different starting salaries) were not located in close proximity to each other. Participants always completed Block 1 first as it comprised the main dependent variable of interest in the study. Six orders of the remaining six blocks (six because they included the three exploratory blocks which were not included in subsequent analyses) were devised such that each block occurred in a different position in each of these six orders and each block was not always adjacent to the same blocks.

Procedure

Participants completed the study in a large laboratory in groups of up to nine participants simultaneously. Participants were asked to complete the questionnaires in their own time and to ask the experimenter if they had any questions. Upon completion, participants were thanked, paid as necessary, and debriefed as to the purpose of the study.

Results

The first step in the analysis was to classify events as positive or negative. This classification was based on the mean ratings of desirability collected from participants for each event. 21 events were classified as negative ($p<.05$) and 19 were classified as positive ($p<.05$) on an 11-point -5 (undesirable) to +5 (desirable) scale. The subjective ratings were as had been expected with the exception of the event 'marry a film star' which was rated as a slightly negative event by our participants.

Table 4.1 shows the results for both positive and negative events, arranged in order of decreasing optimism, as indicated by the mean comparative judgment. A positive value in the mean comparative judgment column indicates that participants tended to rate their own chances of experiencing the event as greater than average, whilst a negative value indicates that participants rated their chances as less than average. In addition to ratings of comparative risk, Table 4.1 also shows the ratio of optimistic to pessimistic individual responses, as in Weinstein (1980). Consistent with Weinstein's finding, the two measures are highly correlated with $r=.77$ for positive events and -.93 for negative events, and the general pattern of results in the study is the same across the two measures.[22] Consequently, we limit our analysis to the comparative responses, as in Weinstein (1980).

---

[22] Given Cohen and Cohen's (1983, p. 75) critique of correlations with ratios, we also conducted a correlation between the comparative responses and the number of optimistic responses given for each event. The resulting correlations (0.91 for positive events and -0.91 for negative events) make us confident that the results of both modes of analysis are comparable.

Table 4.1

'Unrealistic optimism' for future life events.

Event	Mean comparative judgment of own chances vs others' chances	No. of optimistic responses divided by no. of pessimistic responses	Mean perceived frequency
Positive events			
Own own home	1.28 ***	11.17 ***	72.35
Like job after university	0.65 ***	2.39 ***	52.78
Starting salary >£20,000	0.4 **	2.93 ***	53.16
Not spend a night in hospital in 5 years	0.25 ns.	1.39 ns.	53.38
Have a mentally gifted child	0.16 ns.	1.67 ns.	19.62
Visit Amazonian rainforest	-0.12 ns.	0.98 ns.	10.31
Home's value doubles in 5 years	-0.19 ns.	0.80 ns.	25.50
Live past 90 years old	***-0.45*** **	0.73 ns.	22.58
Maintain a constant weight for 10 years	***-0.67*** **	0.73 ns.	32.86
Graduate with a first	***-0.69*** **	***0.60*** *	25.71
Work recognised with an award	***-0.74*** ***	***0.43*** ***	11.39
Last whole winter without being ill	***-0.74*** ***	***0.48*** ***	28.91
Receive good job offer before graduating	***-0.83*** ***	***0.24*** ***	27.05
Starting salary >£30,000	***-0.84*** ***	***0.31*** ***	26.20
Achievements acknowledged in national press	***-0.97*** ***	***0.27*** ***	7.53
Earn >£80,000 in 10 years time	***-1.08*** ***	***0.22*** ***	16.24
Nationwide recognition within profession	***-1.26*** ***	***0.23*** ***	7.11
Starting salary >£40,000	***-1.38*** ***	***0.16*** ***	13.28
Marry a millionaire	***-1.52*** ***	***0.19*** ***	4.01
Negative events			
Marry a film star	-1.84 ***	9.43 ***	1.10
Contract AIDS	-1.75 ***	8.86 ***	3.33
Divorced within 5 years of marriage	-1.25 ***	3.93 ***	32.56
Lung cancer	-1.21 ***	4.18 ***	12.57
Have a drinking problem	-0.88 ***	2.15 ***	13.35
Be sued	-0.82 ***	3.06 ***	10.51
Be fired from a job	-0.66 ***	2.83 ***	22.28
Heart attack before 40	-0.65 ***	2.27 ***	6.09
Be unable to have children	-0.07 ns.	0.96 ns.	11.20
Heart attack	0.03 ns.	1.03 ns.	17.48
Have car stolen	0.03 ns.	0.79 ns.	20.19
Out of work for 6 months	0.04 ns.	1.00 ns.	37.27
Be the victim of a mugging	0.09 ns.	0.71 ns.	24.25
Buy a car that turns out to be terrible	0.19 ns.	0.63 ns.	39.35
Realise chose the wrong career	0.20 ns.	***0.58*** *	34.66
Be the victim of burglary	0.22 ns.	***0.39*** ***	40.74
Be in bed ill for 2 or more days in a year	0.29 ns.	0.77 ns.	67.06
Forced to take an unattractive job	***0.32*** *	***0.43*** ***	49.78
Cancer	***0.34*** *	0.62 ns.	32.31
Break a bone	***0.38*** *	0.70 ns.	48.39
Injured in a road accident	***0.41*** ***	***0.26*** ***	26.92

Note. ns. = nonsignificant. Means in bold and italic font represent significant pessimism.

*$p < .05$. **$p < .01$. ***$p < .001$.

Optimism for negative events, pessimism for positive events

As a first test of the general unrealistic optimism effect, participants' comparative judgments of their own chances versus others' chances were averaged across all negative events. The mean response was -0.32, a result which was significantly below the neutral point (zero), $t(101) = 4.52$, $p<.001$ (2-tailed). This replicates the traditional 'unrealistic optimism' effect. We next employed the same analysis using responses to the positive events. The results for the positive events matched those for negative events: Participants rated the positive events as *less likely* to occur to themselves than the average person (mean = -0.46), $t(101) = 5.46$, $p<.001$ (2-tailed), thus displaying significant 'pessimism' at the group level, in line with the statistical artifact hypothesis. Our study was primarily based on Weinstein (1980) and yet he found optimism for positive events while we find pessimism. Our results do not constitute a failed replication, however, for the positive events in the present study were deliberately modified to make them rarer. Indeed, when comparing the results reported in Weinstein (1980, Table 1) with those in our study (Table 4.1), only two directly comparable events show opposite results (significant optimism in Weinstein's study and significant pessimism in the current study). The first of these, 'receiving a good job offer before graduation,' might be explained by the increase in the number of university graduates between 1980 and 2008, thus making such an event rarer in 2008 than it was in 1980. The contrasting results for, 'your work recognized with an award' might, speculatively, be related to cross-cultural differences. Otherwise, the results of our study and of Weinstein (1980) match.

In conclusion, (rare) positive events overall elicited pessimism, in line with the statistical artifact hypothesis and in opposition to the hypothesis of a genuine, optimistic bias.

Comparing the effects of perceived frequency and event valence

Looking more closely at Table 4.1, it is clear that although the overall analyses clearly replicate Weinstein's (1980) result of seeming unrealistic optimism for negative events, the individual events present a much more equivocal pattern. The mean responses for 12 of the 21 negative events are in a *pessimistic* rather than optimistic direction (although only 4 are significantly so). Interestingly, the data for positive events seem more clear cut than the data for negative events. The mean responses for 14 of the 19 positive events are in a *pessimistic* direction, 12 of these significantly so. By contrast, the means for only 3 of the positive events were significantly optimistic. Across all events, therefore, the means were in an optimistic direction for 14 events, whilst they were in a pessimistic direction for 26 events (p=.08 by the binomial test). To what extent is this variability across events explained by the statistical artifact hypothesis?

As a first test, events were divided into four categories on the basis of participants' ratings of desirability and frequency: Positive – rare; positive – common; negative – rare; negative – common. Figure 4.5 shows the mean comparative probability judgments made for these events. Common events were viewed as comparatively more likely to occur to the self than the average person than rare events were, $F(1, 101) = 146.50, p<.001, MSE = 0.43, eta_p^2 = .59$. Notably, no other significant effects were observed in the analysis of variance (ANOVA). In particular, there was no effect of event valence on comparative ratings, $F(1, 101) = 1.32, p > .05$, $MSE = 1.52$, nor was there any interaction between frequency and valence, $F(1, 101) = 3.60, p>.05, MSE = 0.30$. As evident from Figure 4.5, the non-significant trend in comparative ratings for positive and negative events was actually in the direction of pessimism (negative events were rated as comparatively more likely than positive events).

That differences in comparative ratings are driven exclusively by event frequency and not by event valence is further suggested by the fact the two most 'biased' sets of comparative responses were for the most neutral items in our data set, marry a millionaire and marry a film star, both of which had mean desirability ratings that deviated from zero by less than one scale value. The large 'bias' here is, however, predicted by the statistical artifact hypothesis, precisely because these events were perceived to be the rarest events of their respective valences.

This further confirms that the data in Study 12 provided no evidence for a general unrealistic optimism effect.

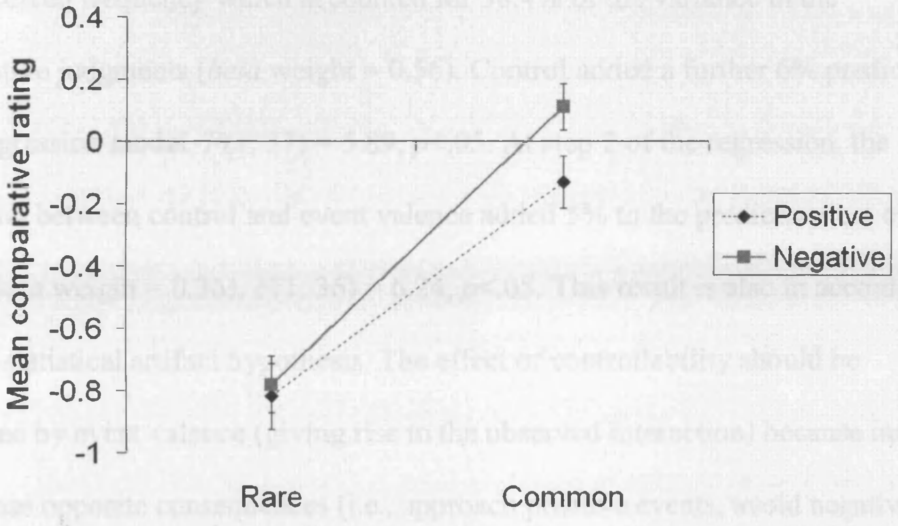


Figure 4.5. Mean comparative ratings for events according to a 4 way classification on the basis of perceived prevalence and desirability. Error bars are plus and minus 1 standard error.

Regression analyses

The preceding analysis provided strong support for the contention that the comparative probability judgments in this study are affected by perceived frequency rather than event valence. Whilst perceived frequency appears to be the best predictor of comparative responses, we also wanted to assess this quantitatively. Furthermore, we had collected data from participants on event controllability, a known moderator, and on the desirability of the events in question. If ratings reflect a genuine optimistic bias, which represents a kind of 'wishful thinking', then one might expect this bias to increase with the perceived desirability of the event in question. We performed a regression analysis, which also included event valence (coded as a dummy variable), to determine the relative contributions of these variables in predicting the comparative judgments.

After transforming the predictor variables to *z* scores (see Howell, 1997, p. 517), we performed a forwards regression. Main effects were added at the first step of the regression, with n-way interactions added at the n^{th} steps. At step 1, two significant predictors emerged in the regression model. As expected, the most powerful predictor was perceived frequency which accounted for 58.4% of the variance in the comparative judgments (*beta* weight = 0.56). Control added a further 6% predictivity to the regression model, $F(1, 37) = 5.89$, $p<.05$. At step 2 of the regression, the interaction between control and event valence added 5% to the predictiveness of the model (*beta* weight = 0.36), $F(1, 36) = 6.24$, $p<.05$. This result is also in accordance with the statistical artifact hypothesis. The effect of controllability should be moderated by event valence (giving rise to the observed interaction) because increased control has opposite consequences (i.e., approach positive events, avoid negative events) for events of different valence. This conclusion was supported by an inspection of the residuals from step 1 of the regression, with controllability dichotomised (via a median split) for the graphical illustration presented in Figure 4.6. Figure 4.6 also shows that deviations from the best fit regression line were, once again, in the direction of pessimism, not optimism (i.e., positive for negative events and negative for positive events when perceived control was low).

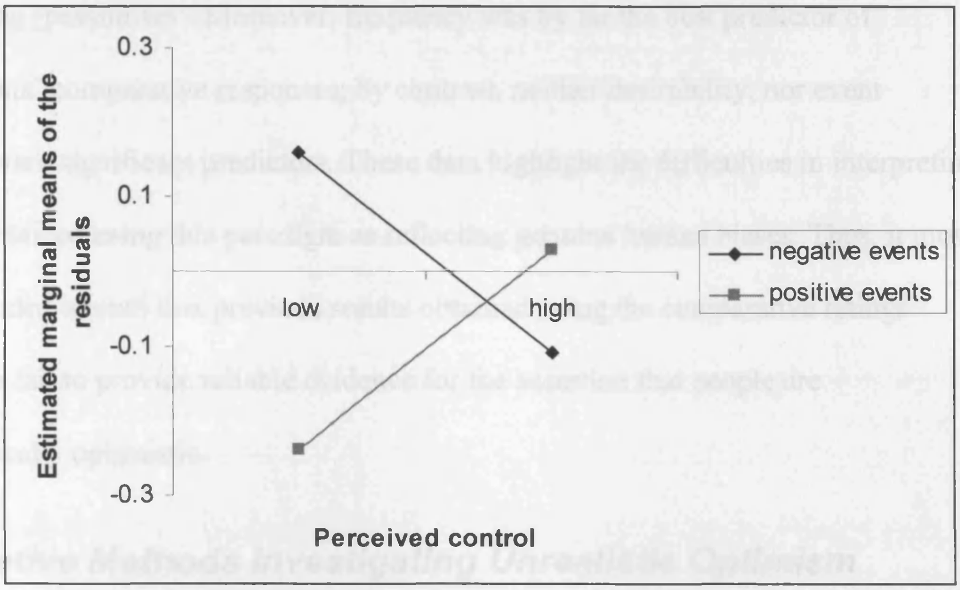


Figure 4.6. The interaction between perceived control and event desirability after controlling for the significant effects observed at step 1 of the regression (main effects of perceived frequency and control)

No other significant predictors emerged from the regression model. From these statistics, it is clear that perceived frequency is the best predictor of participants' comparative responses (see also, Chambers et al., 2003; Kruger & Burrus, 2004; Moore & Small, 2008; Rose et al., 2008), with desirability unable to significantly add any predictive power to the regression model. An additional, by-subjects, analysis of the relationship between frequency and comparative responses suggested that this result generalises not only across all events, but across the whole population (mean coefficient = .28; $t[101] = 14.69, p<.001$) (Lorch & Myers, 1990).

Summary of Study 12

The aim of Study 12 was to distinguish directly between two explanations for the extant data in the literature, a genuine optimistic bias and statistical artifacts. The primary test was whether rare positive events were rated as more likely to occur to the self than to the average person or vice versa. In line with the statistical artifact hypothesis, and in contrast to the predictions of a genuine bias account, rare positive events were rated as less likely to occur to the self than to the average person, thus resembling 'pessimism'. Moreover, frequency was by far the best predictor of participants' comparative responses; by contrast, neither desirability, nor event valence were significant predictors. These data highlight the difficulties in interpreting results obtained using this paradigm as reflecting genuine human biases. Thus, it must be concluded overall that previous results obtained using the comparative ratings paradigm fail to provide reliable evidence for the assertion that people are unrealistically optimistic.

Alternative Methods Investigating Unrealistic Optimism

Thus far, we have concentrated our discussion on the 'direct' method of eliciting comparative ratings as a test for unrealistic optimism, in which participants rate their own chances of experiencing an event relative to the average person. We have argued that this paradigm does not provide a sufficient test of the unrealistic nature of the 'optimism' observed. However, there are a number of further methods that have been used in unrealistic optimism research. Do these offer more robust support for the phenomenon?

The 'Indirect' Method for Examining Unrealistic Optimism

The first, and main, alternative to the 'direct' method is the so-called 'indirect' method. Though less prevalent than the 'direct' method (Weinstein & Klein, 1996), the 'indirect' method has been used by a number of studies in the literature (e.g., Dewbery, Ing, James, Nixon, & Richardson, 1990; Dewbery & Richardson, 1990; Eysenck & Derakshan, 1997; Hoorens & Buunk, 1993; Miller, Ashton, & McHoskey, 1990; Pietromonaco & Markus, 1985; Pyszczynski et al., 1987; Salovey & Birnbaum, 1989; van der Velde & van der Pligt, 1991; van der Velde et al., 1992, 1994; for a review see Helweg-Larsen & Shepperd, 2001). Within this paradigm, participants separately rate their own chance of experiencing an event and the average person's chance of experiencing the same event (most typically on a seven-point scale from 1 [extremely low chance] to 7 [extremely high chance]). The experimenter then calculates the relative judgment by subtracting the participant's judgment of the average person's chance of experiencing the event from the participant's judgment of their own chance of experiencing the event. This procedure yields a difference score, which (for negative events) is taken as evidence for relative optimism if negative and for relative pessimism if positive.[23]

Minority undersampling holds for the indirect scale to the same degree as it does for the direct scale. Given, therefore, that for rare negative events the 'worse off' minority are more likely to be undersampled than oversampled, the average response using the indirect scale may again resemble optimism for purely statistical reasons.

The relationship between optimism and perceived frequency is, however, not as straightforward as it is for the direct method. For one, the indirect method is subject to scale latitude effects (Klar & Ayal, 2004). As the base rate, and hence average risk decreases, so does the opportunity for the 'self' responses to be lower than the 'average person' responses. This should give rise to an inverse relationship between optimism and frequency.

Furthermore, the effects of scale attenuation are more complex for the indirect method. In order to demonstrate this we will once again use perfect predictors: Given that responses must be translated onto a 1-7 response scale, it is first necessary to ascertain how probability ratings should be translated onto such a scale. Table 4.2

---

[23] For positive events, the reverse relationship holds. Typically, however, as with the majority of unrealistic optimism studies, the 'indirect' paradigm has assessed people's risk ratings for negative events.

shows two candidates for the rational translation of percentage estimates onto a 1-7 scale. Using the same assumptions for our perfect predictors as before, we calculated the mean difference score by subtracting the rated risk of the average person from the mean rated self risk. Again, negative scores would typically be interpreted as optimism and positive as pessimism for negative events. Figures 4.7 and 4.8 show the mean difference scores obtained for events of different base rates using the two different translation criteria illustrated in Table 4.2. It can be seen that whilst the attenuated response scale makes it impossible for the perfect predictors to appear perfect at a group level, thus questioning the validity of this scale, there is no systematic relationship between the direction of their bias and perceived frequency. This is consistent with the weaker correlation between optimism and event frequency observed using the indirect method as opposed to the direct method (e.g., Klar & Ayal, 2004; Price et al., 2002; Rose et al., 2008; see also, Chambers et al., 2003).

Table 4.2
Two possible translations (a) and (b) of percentage risks onto a 7-point scale

Scale value	Value meaning	Percentage risk (a)	Percentage risk (b)
1	extremely unlikely	0-9	0-14
2	very unlikely	10-29	15-29
3	unlikely	30-49	30-44
4	50/50	50	45-55
5	likely	51-70	56-70
6	very unlikely	71-90	71-85
7	extremely likely	91-100	86-100

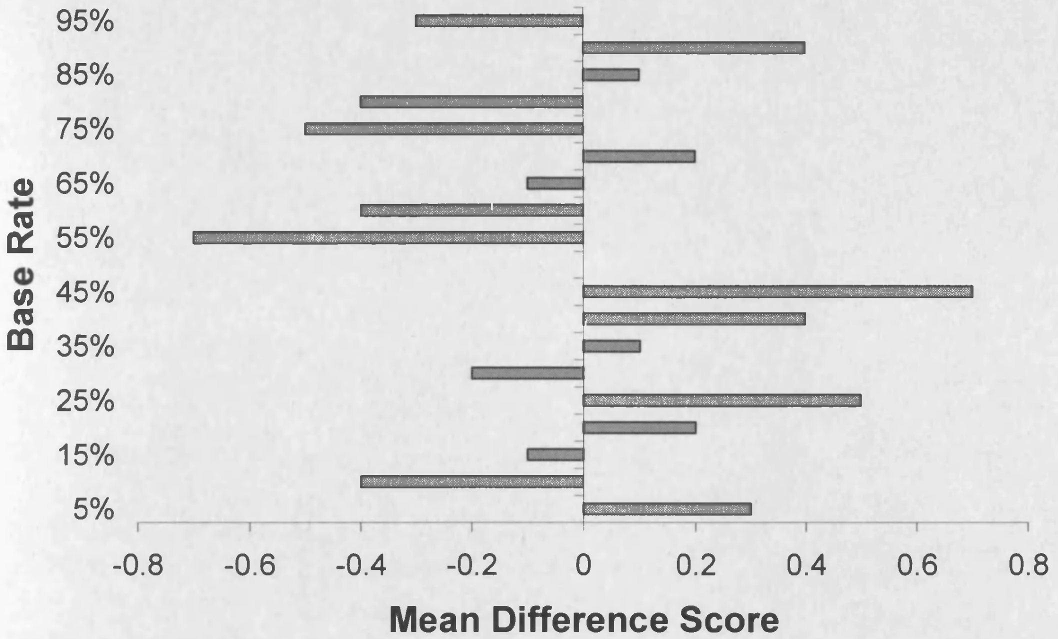


Figure 4.7. Predicted mean difference scores of perfect predictors reporting their (and the average person's) chances of experiencing events of different base rates on a 1-7 scale (using translation a).

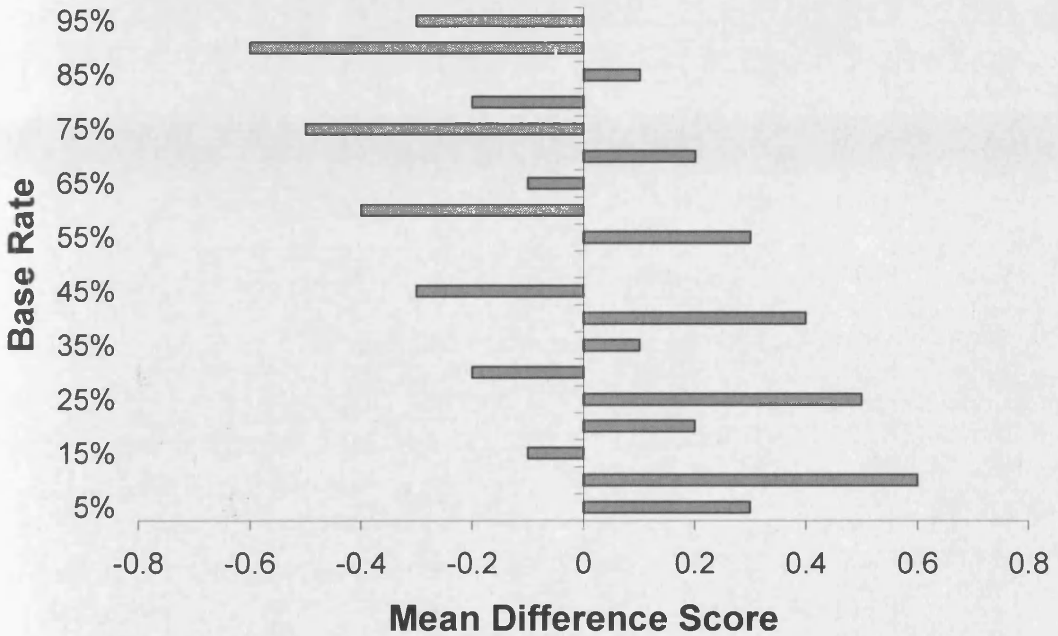


Figure 4.8. Predicted mean difference scores of perfect predictors reporting their (and the average person's) chances of experiencing events of different base rates on a 1-7 scale (using translation b).

Given that the relationship between event frequency and predicted optimism is not systematic, can the unrealistic optimism generally observed using the indirect method be explained purely by minority undersampling? In the real world, people are not perfect predictors. For non-perfect, but rational, predictors there is another statistical mechanism which will result in seeming optimism being observed at the group level. We shall refer to this mechanism as *base rate regression.* It is well documented that people overestimate the frequency of rare events and underestimate the frequency of common events (e.g., Attneave, 1953; Lichtenstein et al., 1978), a phenomenon which can be explained in terms of statistical regression to the mean (see Figure 4.9, top panel) (e.g., Erev et al., 1994; Hertwig, Pachur, & Kurzenhäuser, 2005; Moore & Healy, 2008; Moore & Small, 2007, 2008). Evidence for the direct relevance of this to studies of unrealistic optimism comes from the finding that people generally *overestimate*, and are therefore pessimistic about, their absolute risk for negative events (e.g., Causse, Delhomme, & Kouabenan, 2005; van der Velde et al., 1992, 1994; see also, Moore & Small, 2007).

Regressive probability estimates can be simulated using the formula $y = mx + c$, where x is the objective probability, m is less than 1 and c is solved for the condition where both objective (x) and subjective (y) probability estimates equal 0.5.[24] From this, we can simulate the responses of a population of rational Bayesians who have regressive estimates of the base rate and who have received test results relating to their likelihood of contracting a disease. For this test, people are four times more likely to contract the disease if they receive a positive test score than if they receive a negative

---

[24] This regression equation is a psychological simplification at the extreme ends of the probability scale. Probabilities of 0 and 1 will generally be estimated accurately by participants. We do not consider impossible or certain events in this chapter, nor does the following hinge in any way on extremely low or extremely high probabilities.

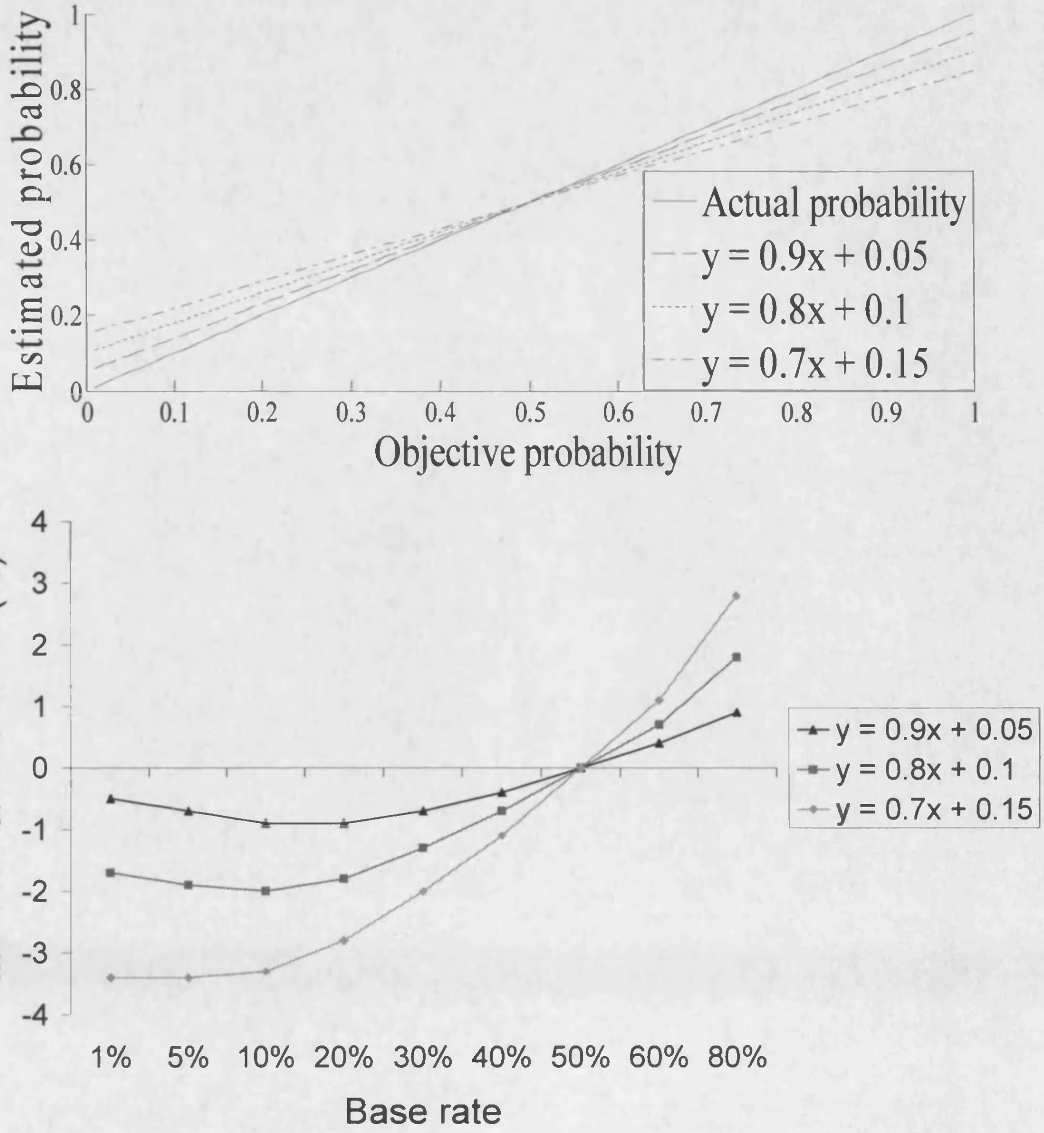


Figure 4.9. The top panel illustrates the effect of regression to the mean on probability estimates. The bottom panel demonstrates the effect of such base rate regression on mean difference scores for events of different base rates. Responses are made by predictors who have a result of a test with a likelihood ratio of 4:1. Responses are made on an unattenuated, indirect scale.

test score. Consequently, the test's 'hit' rate, $P(e|h)$, is .8 (the chance of a positive result given that they will contract the disease) and its 'false positive' rate, $P(e|\neg h)$, is .2 (the chance of a positive result given that they will not contract the disease). Equations 4.1 and 4.2 (Bayes' Theorem) illustrate how a rational person should update their degree of belief on receipt of evidence (e.g., a test result). As rational Bayesians, our hypothetical participants should update their degrees of belief based on their estimates of the base rate and the characteristics of the test (which we shall assume they have been told). To demonstrate the effect of base rate regression in isolation, our simulations will involve an indirect method whereby participants rate both their own and the average person's risk on an unattenuated, continuous, 0-100 scale. Figure 4.9 (bottom panel) shows the resulting mean difference scores that are obtained from differently regressive estimates of the base rate, across the range of base rates for rare (and not so rare) events.

Figure 4.9 (bottom panel) demonstrates that the average response will be optimistic for rare negative events, given rational updating from a regressed estimate of the base rate. Unpublished data provide a first estimate of the scale of base rate regression in the context of unrealistic optimism studies: Clutterbuck (2008) presented participants with 10 standard negative events such as contracting particular cancers, diabetes, coronary heart disease, or being in a road traffic accident. Participants indicated the expected incidence within a sample of 1,000 people. Their estimates were compared to actual figures published by the UK government and relevant health related charities (e.g., the British Heart Foundation). The actual mean rate for the events was approximately 50 per 1000; participants' estimates, by contrast, were approximately 200 per 1000. This corresponds to an objective estimate of 5% and a subjective estimate of 20%. Thus, the regressive estimates assumed in Figure 4.9 seem inherently psychologically plausible and, in fact, might even be considered conservative (however, see also, Windschitl, 2002, on potential difficulties associated with the interpretation of such data).

The small but consistent effect shown in Figure 4.9 (bottom panel) arises on an unattenuated scale. Consequently, it applies to the direct method as well as the indirect method. The slightly curvilinear relationship with event frequency observed in Figure 4.9 (bottom panel) is a consequence of the scale latitude effect that emerges with the indirect method. To the extent that the direct method is less susceptible to scale latitude (Klar & Ayal, 2004), a somewhat stronger, linear relationship with event frequency will be observed.

In the context of the indirect method, our next question is how do scale attenuation effects interact with the effects of base rate regression? Figures 4.7 and 4.8 demonstrated that the most widely used scale has peculiar properties in that events of certain base rates will provide negative average difference scores and events of other base rates will result in positive average difference scores, resulting purely from the translation of a frequency onto this 1-7 scale. However, as Figures 4.7 and 4.8 demonstrate, it is not possible to make predictions as to whether participants' difference scores should mean to zero, or some other number, without knowledge of both the precise base rate and participants' translation strategies. Figure 4.10 shows the effect of base rate regression ($y = 0.7x + 0.15$, where y = estimated base rate and x = true base rate) on average difference scores obtained for rare events (base rates less than 0.5), assuming that participants' translate their chance estimates onto a seven point response scale as prescribed in the right hand column (b) of Table 4.2. The non-systematic effects of scale attenuation mean that for some base rates the average difference score is predicted to be positive. Overall, however, the effect of base rate regression makes comparative responses more negative even with attenuated response scales. This is illustrated in Figure 4.10 which, for example, shows that for base rates below 0.35 there appear to be considerably more negative difference scores than positive difference scores predicted by the base rate regression mechanism (represented by the solid line). Moreover, given that there is no predefined, 'obvious', way to translate real-world knowledge of risk onto a 1-7 response scale, participants have the option to translate it in such a way as makes their future look rosier. Such a strategy would not imply that people are genuinely unrealistically optimistic. Rather, when a crude response scale forces on them the choice of seeming either pessimistic or optimistic, they choose the latter.

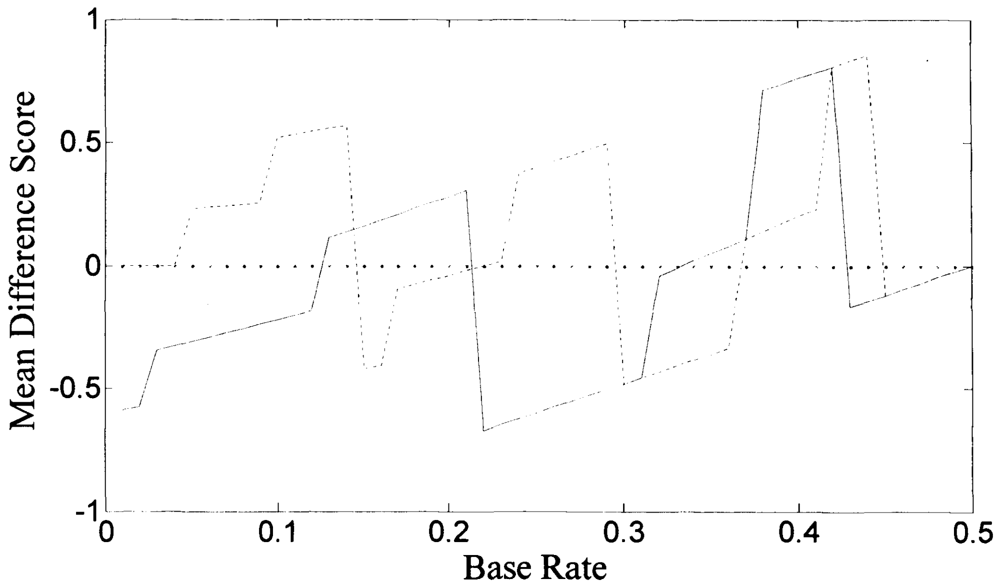


Figure 4.10. The effect of base rate regression on responses translated onto a 1-7 scale, using translation b. In the simulation, participants have received the result of a test with a likelihood ratio of 4:1. The dotted line indicates the effect of scale attenuation alone, whilst the solid line shows the mean difference score for events of different base rates, for individuals whose base rate estimates are regressive (y = 0.7x + 0.15).

As noted, the predictions depicted in Figure 4.10 are dependent on the translation strategy that participants use. Without knowledge of this strategy it is not possible for any theory to make detailed predictions. The statistical artifact hypothesis is unable to make strong predictions about the relationship between seeming optimism and frequency. At the same time, unrealistic optimism is unable to predict that rational responses should mean to zero, for this is not true for the majority of base rates (see Figures 4.7, 4.8, 4.10). One potential avenue for future research is therefore to determine the translation strategies participants use. Only with knowledge of this, and the base rate of the events under investigation, can precise predictions be made concerning what constitutes rational responses using the indirect response scale.

In the meantime, however, minority undersampling and base rate regression will always lead mean difference scores to appear more negative (for rare events). Moreover, the vagaries of the 'indirect' scale, as highlighted above, make the data obtained from it even more difficult to interpret than those obtained using the 'direct' method.

Base rate regression versus differential regression.

Moore and colleagues (Moore & Healy, 2008; Moore & Small, 2008) suggest that unrealistic optimism can be explained in terms of differential regression. We show here how differential regression is different from our base rate regression mechanism, and why differential regression cannot explain unrealistic optimism. The differential regression hypothesis assumes that people possess incomplete knowledge, but that their knowledge of themselves is less incomplete than their knowledge of other people. Incomplete knowledge results in estimates that regress towards the mean (e.g., Figure 4.9, top panel). Given that people's knowledge of themselves is less incomplete than their knowledge of others, estimates of self risk will be less regressive than estimates of others' risk. Consequently, for rare negative events (those for which unrealistic optimism is most prevalent, e.g., Study 12), participants' estimates of their own risk will be greater than the base rate (as their estimates regress towards the midpoint of the scale). However, participants' estimates of the average person's risk are predicted to be greater than their estimates of self risk (resembling unrealistic optimism for negative events), as the more uncertain knowledge leads to a more regressive estimate.

There is, however, a crucial difference between unrealistic optimism and the contexts for which the differential regression account was originally devised, such as performance on a quiz. In the quiz case, both group and individual estimates are derived from a common prior. In such situations, people can have better knowledge about their own performance than about other people's. However, in the context of unrealistic optimism studies, where participants provide estimates of the likelihood of experiencing binary future events, the estimate of the average person's chance *is* the prior and participants derive estimates of their own risk from this prior (if updating their belief rationally, in accordance with the prescriptions of Bayes' theorem, as assumed in Moore & colleagues' account). By this process, estimates of self risk are not more complete than estimates of the average person's risk, as any incompleteness in knowledge of the average person's risk will be propagated in derivations of self risk estimates. Moreover, for binary future events, whilst people may know the base rate perfectly, this is typically not possible for their own risk. Indeed, the incompleteness of this latter knowledge is not measured by deviation from the average, but from the extremes of 0% and 100%.

As a consequence of these conceptual differences, the Bayesian process assumed by Moore and colleagues (Moore & Healy, 2008; Moore & Small, 2008) does not in fact give rise to unrealistic optimism effects. This process is the very same, standard, process of rational Bayesian belief updating we have invoked throughout: "people begin with some prior expectation and then update their belief when they get new evidence" (Moore & Small, 2008, p. 164). The prior expectation is derived from the base rate, which also constitutes the average risk estimate: "If one were asked to estimate another person's outcome yet knew nothing about that person, the group's average would be a good opening assumption (or what statisticians call a *prior*)" (Moore & Small, 2007, p. 973). Given evidence about their own susceptibility to future events, people update their chances of experiencing the event from this prior, as prescribed by Bayes' Theorem (and outlined above). This does not, however, result in data suggesting unrealistic optimism at the group level. Consider our previous example concerning a population of people who each receive a test result indicating their susceptibility to lung cancer. Without the mechanisms of scale attenuation or minority undersampling, the estimates of self risk in this example (8.7% for 41% of the population and 4.1% for 59% of the population) average 6% (bar rounding error), which is the base rate. Furthermore, the result is not peculiar to this specific example. It arises for any test characteristics or base rate values (see Appendix for proof).

In short, a rational Bayesian process will not inherently give rise to unrealistic optimism. Our base rate regression mechanism requires an error (albeit itself a both understandable and unbiased one): a misestimate of the base rate. Its effect rests on the discrepancy between the actual base rate and the perceived base rate: the absolute value of individual's estimates is driven by the perceived base rate (because they are derived from it via Bayesian updating); the actual number of individuals receiving particular test outcomes (or any other individuating information such as family history), however, is driven by the actual base rate. In other words, people's estimates are based on how they think the world is. The proportions of people receiving the different test results, however, depend on the way the world actually is. Consequently, given a discrepancy between the actual and perceived base rate, the average self estimates will no longer equal either of these base rates (see Figure 4.9, bottom panel).

In contrast to the differential regression hypothesis, our 'base rate regression' mechanism will lead to more optimistic seeming responses, but this results from the discrepancy between the real base rate and the perceived base rate, not the discrepancy between estimates of self chance and the average person's chance (as assumed by Moore & colleagues).

Assignment to Percentiles

A further, possible test for unrealistic optimism is based on assignment to percentiles. Participants are required to estimate the percentile rank of their chances of experiencing an event relative to a specified sample (e.g., Moore & Small, 2008; Weinstein & Lachendro, 1982). For example, Moore and Small asked participants to estimate their percentile rankings relative to all other participants in the experiment:

> "If you think you are more likely than anyone else in this experiment to experience the event, enter "100" as your percentile. If you think that you are the least likely person to experience the event, enter "0" as your percentile. If you think your chances are exactly in the middle, enter "50" as your percentile. All numbers between 0 and 100 are acceptable responses" (Moore & Small, 2008, p. 147).

A bias was inferred in this experiment if the average percentile rank differed from 50 (see also, Weinstein & Lachendro, 1982).

When percentile ranking scales are used in experiments investigating relative chances of experiencing binary events in the real-world (e.g., Moore & Small, 2008; Weinstein & Lachendro, 1982), it is unlikely that even rational percentile rankings will mean to 50. Returning once more to perfect predictors, a percentile ranking is not appropriate, as it is meaningless to provide percentile rankings for essentially categorical data (either a person will experience the event or they will not). Given that the events being used in these studies are binary events, this in itself is a problem for this measure. However, percentile rankings seem inappropriate even for non-perfect predictors. Although people do have access to a variety of sources of information by which they may differentiate their chances of experiencing an event from those of the average person, these sources are limited. Consequently, it seems likely that the finest comparison people would be able to make on a percentile ranking scale would be approximately 20% (i.e., they may be able to divide the comparison group into fifths). Consider therefore a sample of 150 people (Moore & Small used 158 in their study). The 30 individuals who know that they are in the least likely fifth of individuals to experience an event would seem to be able to rationally respond (on this scale) with any number between 0 and 20% (as 20% of people have the same chance as them of experiencing the event). This is the same range as the other groups of people who could respond between 20 and 40%, 40 and 60% and so on. This example highlights the major problem with percentile rankings. As soon as there are people with equal chances to other people (i.e., ties in the data), the task: (a) becomes confusing for the participant, and (b) it can no longer be assumed that the mean rankings will be 50%. These points are further illustrated with a numerical example. If the 30 individuals in each fifth of the population above respond with 10%, 30%, 50%, 70%, 90%, then the mean percentile ranking is 50%. However, as 30 of the participants *are* the least likely people to experience the event, it also seems rational for them to report a percentile ranking of 0%, and, in fact, this seems to be what is asked of them in the experimental instructions. This response strategy would give rise to percentile rankings of 0%, 20%, 40%, 60%, 80% with a mean percentile ranking of 40% (i.e., less than 50%). It would seem unsatisfactory to interpret this result as showing these participants to be unrealistically optimistic.

Longitudinal Studies

> "It is usually impossible to demonstrate that an individual's optimistic expectations about the future are unrealistic" (Weinstein, 1980, p. 806).

If researchers could see into the future, it would be possible to compare individuals' expectations with the outcomes that they will actually experience. Longitudinal studies essentially allow researchers to see into the future by comparing outcomes at Time 2 with expectations at Time 1. In the context of unrealistic optimism studies though, this is not as simple as might be thought. Typically, these studies involve estimates about binary events: For example, a person will either have a heart attack or not have a heart attack; they cannot have .7 of a heart attack. Consequently, these events are not amenable to a longitudinal design, as participants' probabilistic expectations about experiencing these events are not directly comparable with the binary outcome values, at least at the level of the individual event.

In order to make meaningful comparisons between binary outcome events and probabilistic estimates of the likelihood of occurrence of those events, the events must be aggregated in some way. One potential method is to ask an individual to provide binary ratings of a number of events, providing either a 'yes' or 'no' response to the question of whether they will experience each one within a particular time frame (e.g., ten years). Ten years later, the researcher would check the number of those events that the participant experienced. The total number of 'yes' responses both for expectation and outcome would then be compared. For negative events, if the number of 'yes' expectations is less than the number of 'yes' outcomes then that individual might be considered unrealistically optimistic. Prima facie, this appears to be a reasonable strategy. However, the nature of a participant's task in this study seems problematic. When completing a questionnaire whilst healthy, and without the ability to forecast the future, it would seem bizarre for a participant to circle 'yes' to any life threatening events (e.g., road accident, cancer, kidney failure etc.). Yet it would seem inappropriate to attribute such a reluctance to unrealistic optimism.

At first consideration, a better method would be to adopt the approach used widely in the judgment literature in order to study the extent to which people's probability assessments are 'calibrated'. This requires a very large selection of potential events for which participants provide probability estimates. Events are then combined by 'binning' all events that the participant assigned a particular range of probabilities to (e.g., 10-20% chance; 21-30% chance etc.). The 'calibration' of the participant's responses is then subsequently evaluated by calculating the proportion of events in each 'bin' that actually occurred. To the degree that the participant's responses were well calibrated, between 10 and 20% of events that they assign a 10-20% chance of occurring should occur (see e.g., Keren, 1991; Lichtenstein et al., 1982; Wallsten & Budescu, 1983; Yates, 1990). Unfortunately, this method is itself subject to statistical artifacts (e.g., Erev et al., 1994), as discussed in Chapter 1.

In light of the problems highlighted above, it is no coincidence that longitudinal studies typically use non-binary dependent variables, for which a direct comparison can be made between prediction and outcome. Although therefore addressing slightly different questions, two such longitudinal studies have purportedly found evidence of optimism effects. Unfortunately, these studies suffer from other methodological difficulties.

Shepperd, Ouellette, and Fernandez (1996, Study 1) found that liberal arts university students estimated their starting salary (four months prior to graduation) as higher than that actually received by liberal arts graduates. However, the "average starting salary of graduating seniors across the university was noticeably higher" than that for liberal arts graduates specifically (Shepperd et al., 1996, p. 847). Four months prior to graduation it is conceivable that these students would not have known that they were likely to earn less than the average graduate of this university. Hence, even if they were to have perfect knowledge of the mean starting salary for graduating seniors, and all considered themselves as completely 'average', given this statistic a significant optimistic bias could still have emerged from the data. Shepperd et al. (Study 2) found that students were optimistic in their predictions of their exam grade when estimating it one month prior to the exam, although performance estimates were very well calibrated at the group level upon completing the exam.[25] However, without knowledge of the content of the exam, it might not be surprising that students were optimistic prior to the exam. As exams tend to become more difficult as one progresses through the education system, an estimate based on past experience of exams is likely to appear optimistic relative to their actual performance. The most striking finding from Shepperd et al.'s study appears to be the accuracy of participants' predictions having taken the exam.

Lachman, Röcke, Rosnick, and Ryff (2008) reported that individuals aged 32-64 (the youngest respondents in this survey were 32 years old) predicted greater life satisfaction in ten years time than they subsequently reported experiencing at Time 2 (between eight and ten years later). It is not, however, clear that these respondents were unrealistically optimistic. This is evidenced by another observed effect, namely that, in retrospect, participants at Time 2 rated their life satisfaction at Time 1 as lower than they did when they rated it at Time 1. Thus, it is possible that participants' levels of life satisfaction had genuinely increased in this time, consistent with their predictions. Given this interpretation, the explanation of the lack of a difference between present ratings of life satisfaction (mean ratings of 7.71 and 7.77 [on a 0-10 scale]) at the two time points and the observed difference between future (at Time 1) and present (at Time 2) is that participants change their usage of the response scale as their degree of life satisfaction changes. Presumably, life satisfaction perceptions are most intuitive on a relative (or ordinal) scale (see e.g., Stewart et al., 2006).

[25] That is until immediately prior to feedback when they displayed pessimism (a result not of interest to the present discussion).

Consequently, in the knowledge that they are reasonably happy with their lives, people's present life ratings reflect this whilst allowing for them to become even more satisfied (for who knows how satisfied it is possible to be?!), and these ratings serve as an anchor for their ratings of past and future life satisfaction. Given this interpretation, these results demonstrate realism and not bias in these participants' predictions of the future.

In summary, longitudinal designs enable the comparison of a predictive estimate with actual outcomes. However, such studies are difficult to conduct with the kinds of binary outcomes that have formed the focus of unrealistic optimism research to date. Shepperd et al.'s (1996) and Lachman et al.'s (2008) non-binary events do enable meaningful comparisons to be made at the individual level. In the absence of confounding factors and results, such studies could provide good measures of potential optimism in people's expectations. However, by moving away from well-specified binary events to general constructs such as life satisfaction, the study results are not as clear in their interpretation. For example, when investigating well-being there is little knowledge of the factors that participants themselves consider important for well-being, with different people citing different factors as important (Ryff, 1989).

What is Needed for a Direct Test of Unrealistic Optimism?

> "The most obvious way in which unrealistic optimism would present itself would be as an underestimation of the actual likelihood of experiencing a negative event...One major problem faced by such studies is the difficulty of determining the actual risk, the statistic that is accurate for the particular individual under investigation" (Weinstein & Klein, 1996, p. 2).

All of the methodological concerns affecting unrealistic optimism studies would be removed if it were possible to compare participants' judgments with an objective probability. Longitudinal studies provide one possibility by combining events into sets and considering outcome proportions ('calibration'). Unfortunately, as discussed in the introduction to this thesis, this runs into well-known problems of its own (e.g., Erev et al., 1994). A final possibility, however, is to make use of known, objective, probabilities in the context of an experimental design.

We therefore chose to conduct an experimental test of unrealistic optimism within the same paradigm as used in Chapters 2 and 3, using a visual representation of probability and a 'self' versus 'average person' manipulation. Given the previous success of this paradigm, it seems a good candidate paradigm for a direct test of the unrealistic optimism phenomenon. Given an objective probability that is constant across participants, it would not be possible to explain away any observed effect through statistical artifacts.

Study 13

Study 13 aimed to provide a direct, experimental, test of the unrealistic optimism phenomenon. Crucial to this design was the fact that participants were supplied with an objective basis for their subjective estimates and that this objective basis was *identical* across the experimental manipulations.

Method

Participants

96 Cardiff University undergraduate students participated in the study in return for either course credit or cash payment.

Design

In order to test the hypothesis that people believe that their chance of contracting a disease is lower than the chance of other people like them (unrealistic optimism), the between participants independent variable was the potential victim; namely whether participants were judging their own ('your') chance of being exposed to an MRSA like disease, or 'Sarah's' (who was also a Cardiff University student) chance. The full design was a 2 x 3 mixed design, as participants based their judgments on three different probability matrices, and therefore the within subjects variable was the three probability levels, high, medium and low. The dependent variable was the probability estimates, which participants made by writing a number between 0 (it is impossible that *you* [*Sarah*] will be put in a bed infected by the virus) and 100 (it is a certainty that you [*Sarah*] will be put in a bed infected by the virus). The order in which participants made their judgments using the high, low and medium probabilities was counterbalanced across participants in each condition.

Materials and procedure

In order to completely counterbalance the presentation order of the three probability matrices, six booklet orders were prepared for each condition. A booklet consisted of three pages. Each page repeated the same cover story. The cover stories, which contained the person manipulation, are reproduced below (the words used in the Sarah condition are included in italics):

> 'Drug resistant viruses are becoming more and more prevalent in British hospitals. Many of these viruses are potentially deadly and MRSA is a well known example. At some stage in your/*her* life you/*Cardiff University student Sarah* will be admitted to hospital and unfortunately the prevalence of these

drug resistant viruses is showing no signs of decreasing. In the future therefore you/*she* might well find yourself/*herself* facing the following situation. Please read the situation carefully and imagine that it is reality.

You have/*Sarah has* been admitted to a South Wales hospital for a routine procedure. However, an often fatal drug resistant virus is thriving in this hospital. 75% of people who become infected with this virus die from it. This virus contaminates a number of the hospital's beds. The matrix below represents the distribution of hospital beds infected by the virus (BLACK squares). White squares represent those beds not infected by the virus.

By looking at the matrix below please estimate the chance that you/*Sarah* will be put in a bed infected by the virus (BLACK) thus exposing you/*her* to it.'

The matrix referred to in the text was one of three probability matrices (black and white versions of those used in Study 1).

Having completed a consent form and made their way through the experimental booklet, participants were thanked, debriefed as to the purpose of the study and paid (where appropriate).

Results

One participant was excluded from the analyses as their three probability estimates did not correspond to the basic rank order of the probability levels. After this exclusion there were 95 participants included in the data analysis, 47 in the 'you' condition and 48 in the 'Sarah' condition.

The probability variable was the only variable to have a significant effect on participants' probability estimates, $F(2, 186) = 1151.81, p<.001, MSE = 101.80$.

Neither the person manipulation, $F(1, 93) = 1.958$, $p>.05$, $MSE = 206.02$, $eta_p^2 = .02$, nor the interaction between the two variables, $F(2, 186) = .959$, $p>.05$, $MSE = 101.80$, reached significance. Examining the pattern of the results (Figure 4.11), it seemed possible that there was a significant difference in probabilities relating to the self and to Sarah when the probabilities were low. As the unrealistic optimism effect is typically reported for events with low base rates, there was also a theoretical rationale to analyse the judgments of the low probabilities independently. A two-tailed t-test indicated that the difference between the two experimental groups in their probability judgments at the low probability level was not significant, $t(93) = 1.66$, $p=.10$. Despite the hypothesis for the present study being one-tailed, the two-tailed test was performed because the direction of the hypothesis was in the opposite direction to the pattern of the results. As such, not only did the results of Study 13 show no significant difference between the two groups' risk ratings, but the (weak) trend in the data was in the opposite direction to that predicted by unrealistic optimism. Of course, this is simply a null effect, and one based on a hypothetical scenario. However, in Chapter 2 this paradigm was found to be powerful enough to observe significant effects, as well as to demonstrate the importance of a moderating variable (event controllability), thus strengthening the interpretation of this null result.

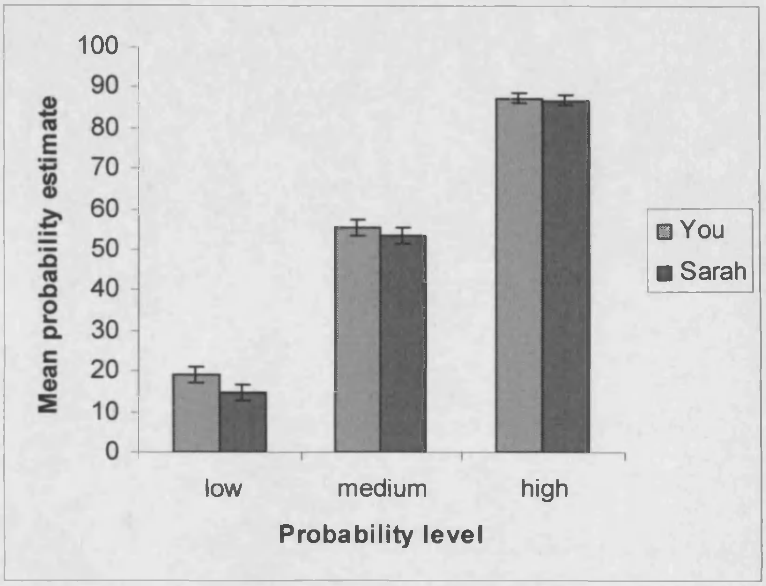


Figure 4.11. Mean probability estimates made across probability levels by participants in both groups. Error bars and plus and minus 1 standard error.

What is the Future for Unrealistic Optimism?

We have introduced a range of statistical concerns that plague studies seeking to demonstrate that people are unrealistically optimistic about their futures (e.g., Weinstein, 1980, 1982, 1984, 1987). The statistical distribution of the events in the real world (specifically, the rarity of these events) coupled with the fact that minorities are more likely to be undersampled than oversampled, and the nature of the response scales used, mean that aggregate data resembling optimism can be found readily even with entirely unbiased predictors. In addition to this stands the likely effect of the fact that people's estimates of event base rates are regressive.

In two studies we tested whether there was any evidence for genuine optimism, beyond the statistical artifacts these mechanisms will give rise to; neither of these studies found any such evidence. Study 12 replicated the traditional unrealistic optimism effect for negative events; however, this effect was driven by frequency as expected from a purely statistical perspective, and event desirability was not able to account for any variance in comparative judgments once perceived frequency was controlled for. Moreover, Study 12 did not find any evidence of unrealistic optimism in comparative responses to rare positive events, contrary to the predictions of unrealistic optimism, but as predicted by the Statistical Artifact Hypothesis. Study 13 provided an experimental test of the unrealistic optimism phenomenon within a fictional scenario in which each participant's true, objective probability of contracting a disease was independently specified. Again, no evidence for an unrealistically optimistic bias was observed under these controlled conditions. Consequently, the statistical artifacts highlighted in this chapter seriously question the validity of the conclusion that people are unrealistically optimistic.

The statistical artifacts introduced here can also explain a number of previously identified moderators of the effect, including: Specificity of the comparison target, event controllability, experience and event frequency. Moreover, statistical artifacts readily explain the weaker correlation typically observed between optimism and event frequency when measured using the indirect method as opposed to the direct method (e.g., Klar & Ayal, 2004; Price et al., 2002; Rose et al., 2008; see also, Chambers et al., 2003). Hence the artifacts identified would seem to cloud our understanding of the optimism phenomena, even if new evidence that provides robust support for unrealistic optimism were found.

How likely is such evidence to be found? In the final section of this chapter we examine the basic plausibility of the claim that people might be unrealistically optimistic in light of other research. For example, unrealistic optimism, though not clearly established empirically, might seem theoretically plausible in light of the fact that people are known to be subject to a range of other 'positive' biases such as a tendency to overconfidence, to overestimating their abilities and to an exaggeration of their degree of control.

Related Phenomena

Unrealistic optimism may be perceived as similar to, and hence made plausible by, a number of other phenomena.

The planning fallacy (e.g., Buehler & Griffin, 2003; Buehler, Griffin, & MacDonald, 1997; Buehler, Griffin, & Ross, 1994; Kahneman & Tversky, 1979b; see also, Arnold, 1986; Kahneman & Lovallo, 1993) is the phenomenon whereby people predict that tasks will take less time to complete than they do and predict that they will complete more work in a given time period than they subsequently do. Thus, people could be considered overly optimistic in their planning predictions. This phenomenon appears robust. However, in none of these tasks are people required to give probability judgments; hence the planning fallacy seems conceptually rather distinct. Consequently, one would want to infer little from the existence of this fallacy on its own.

There are, however, a number of other phenomena that seem conceptually closer to unrealistic optimism such as the so-called illusion of control (e.g., Langer, 1975; Langer & Roth, 1975), people's seeming belief that they are better than average in terms of ability (e.g., Svenson, 1981), or the finding of overconfidence in calibration studies of probability judgment (e.g., Kahneman & Tversky, 1973). However, these biases have also come under scrutiny on statistical grounds, as we will discuss next. This has questioned the traditional, motivational based explanations for these phenomena and, in these cases too, has raised the possibility that people are not as biased or irrational as previously believed.

We have already summarised Erev et al.'s (1994) statistical account of over- and underconfidence in Chapter 1. Other related phenomena for which statistical accounts have been proposed include: The Hard/Easy effect (Juslin, Winman, & Olsson, 2000), the False Consensus effect (Dawes & Mulford, 1996), the Illusion of

Control (Matute, Vadillo, Blanco, & Musca, 2007) and the Better-than-Average/Worse-than-Average effect (Moore & Healy, 2008; Moore & Small, 2007).

Moore and colleagues' account of the Better-than-Average/Worse-than-Average effect (Moore & Healy, 2008; Moore & Small, 2007, 2008; see also, Kruger, Windschitl, Burrus, Fessel, & Chambers, 2008) is based on the underlying assumption that people have imperfect knowledge about their performance on a task, but their knowledge of other people's performance is more imperfect. Although not a tenable assumption in unrealistic optimism studies (as discussed above), this assumption seems very plausible when considering the traditional Better-than-Average/Worse-than-Average effects. Moore and colleagues' theory "is based on the Bayesian notion that people begin with some prior expectation and then update their belief when they get new evidence" (Moore & Small, 2008, p. 164). Imperfection of knowledge thus results in estimates that are regressive towards the prior expectation. Estimates are more regressive for others than for the self as people's knowledge of others is worse than their knowledge of themselves. To illustrate this point, imagine the following example: Before taking a test, you expect to answer approximately 50% of questions correctly. Upon completion of the test, you realise that it was easier than you had expected. Consequently, you know that you have answered more than 50% of questions correctly. When asked to estimate your own performance, your estimate will be based on evidence (i.e., your experience of taking the test) which is imperfect. This imperfection will lead to the regression of your estimate towards your prior (as prescribed by Bayes' Theorem), which was 50% (Moore & Healy, 2008; Moore & Small, 2007). Such a process will ensure these absolute estimates resemble underconfidence in such easy tasks. By contrast, if asked to estimate the test performance of another person, you do not have possession of the same evidence relating to their performance on the test. Consequently, these estimates will regress towards 50% to a greater degree than those for your self. Thus, for easy tasks, absolute underconfidence should be observed, but comparative overconfidence (relative to estimates of the average) should be observed, giving rise to the "Better-than-Average effect". For difficult tasks, the reverse result is obtained: absolute overconfidence, but comparative underconfidence; the so-called "Worse-than-Average effect".

A Summary of the Future for Unrealistic Optimism

It seems then that rather less support exists for the theoretical possibility of an optimistic bias regarding future life events than might be assumed. Key phenomena that could be linked to such a bias have themselves come under attack on the basis of statistical considerations.

Chapter Conclusions

We have introduced a range of methodological concerns that plague traditional studies 'demonstrating' that people are unrealistically optimistic. Typically, these demonstrations are based on people rating their chances of experiencing negative events as being less than the average person's. At the root of these methodological concerns lies the fact that the negative events that form the focus of these studies are generally rare events. This gives rise to three statistical problems: the effects of scale attenuation, minority under-sampling, and base rate regression. All three are independent statistical mechanisms by which seeming optimism may emerge from entirely unbiased predictors.

Indeed, we demonstrated that the response scales used in optimism research give rise to seeming bias with predictors that are not only rational, but perfect, that is, in possession of fully accurate knowledge about the state of the world. This is true of the scales used in both the so-called 'direct' and 'indirect' methods. It would seem a minimum requirement for the validity of a response scale, that genuinely accurate responses do, in fact, appear accurate using that scale. Consequently, the response scales on which unrealistic optimism research is based seem to fail this most basic requirement of validity.

We then demonstrated how scale attenuation would generate seeming optimism with the direct method in the absence of perfect knowledge. Assuming only very weak, but unbiased, diagnostic information, 'unrealistic optimism' emerges readily, as we showed both analytically and with reference to specific examples modelled on extant research. Moreover, empirical evidence for the contention that scale attenuation contributes to seeming unrealistic optimism was cited. This research demonstrated that greater optimism was observed using a more attenuated response scale than a less attenuated response scale (Otten & Van der Pligt, 1996).

Minority under-sampling also affects both the direct and the indirect method. We provided empirical evidence for the role of minority under-sampling through a meta-analytic correlation between the sample sizes and effect sizes of previous studies.

Base rate regression can also affect both the direct and the indirect method, and there exists independent empirical support for the critical assumption that people's estimates of probabilities are frequently regressive.

To probe further whether robust evidence for unrealistic optimism can, in fact, be found, we also conducted two studies of our own. These were designed such that the statistical problems identified could be ruled out as alternative explanations. Neither of these found any evidence of unrealistic optimism. Study 12 examined the conflicting predictions of the optimism and the statistical artifact hypothesis concerning low probability, positive events. The results obtained were in direct opposition to optimism, but in line with the statistical hypothesis. Study 13 failed to find any evidence of optimism in an experimental paradigm that has successfully detected differences in closely related circumstances (see Chapter 2).

Finally, we considered other potential measures of unrealistic optimism, and summarised research relating to other phenomena that might make unrealistic optimism at least seem plausible. We concluded that unrealistic optimism is extremely hard to test. Similarly, many other, potentially related, phenomena have themselves been questioned in light of realistic statistical explanations for data that have been offered in support of them.

In summary, there seems, presently, to be no compelling evidence for a general, unrealistically optimistic bias. Furthermore, the outlook regarding such a bias, if anything, seems rather bleak, particularly when the results of this chapter are considered in conjunction with the results of Chapter 3, which found no evidence of a simple wishful thinking bias.

Some of us might sometimes be overoptimistic. Certain subgroups of the population might demonstrate a bias, for example, entrepeneurs, gamblers and smokers (e.g., Cooper, Woo, & Dunkelberg, 1988; Coventry & Norman, 1998; Delfabbro & Winefield, 2000; Griffiths, 1994, 1995; Hansen & Malotte, 1986; Ladouceur, Gaboury, Dumont, & Rochette, 1988; Lee, 1989; McKenna, Warburton, & Winwood, 1993; Rogers & Webley, 2001; Wagenaar, 1988; Walker, 1992; Weinstein, Marcus, & Moser, 2005; but see also Delfabbro, 2004; Rise, Strype, & Sutton, 2002; Sutton, 1999, 2002). Similarly, almost all of us might be optimistic about some very particular things, for example, the planning fallacy seems a near universal phenomenon, both empirically and anecdotally. The existence of a general optimistic bias cannot, however, be inferred from these more specific ones.

By questioning the status of unrealistic optimism, this chapter adds to the literature already suggesting that other human biases may simply be statistical artifacts (e.g., Dawes & Mulford, 1996; Erev et al., 1994; Juslin et al., 2000; Moore & Healy, 2008; Pfeifer, 1994; Soll, 1996). A considerable, and growing, body of research thus suggests that people's probability estimates may be more rational than often assumed. Moreover, the statistical account advanced means that it is not clear that there is any inconsistency between the lack of a wishful thinking effect, as observed in Chapter 3, and previous research findings.

Chapter 5 - General Discussion

In the introduction to this thesis, we highlighted the importance of understanding probability judgments situated in the presence of utility-laden situations. In Chapters 2 and 3, we subsequently presented a systematic experimental investigation of the potential biasing effect of utility on probability estimates, whilst providing participants with an objective anchor for their estimates.

Chapter 2 first demonstrated an effect of outcome severity on probability estimates by which very negative events were judged more likely to occur than more neutral events. Subsequent investigation demonstrated that this effect was moderated by the controllability of the negative event, such that only controllable negative events were rated as more likely to occur, whilst uncontrollable negative events were assigned the same probability rating as neutral events. This led to the development of a decision-theoretic explanation for the effect in terms of asymmetric loss functions (e.g., Weber, 1994). The nature of this explanation means that the effect can be considered a rational response to a recognition that human judgment of uncertainty is, in itself, uncertain; it does not seem to be an irrational bias resulting solely from the utilities of the events considered.

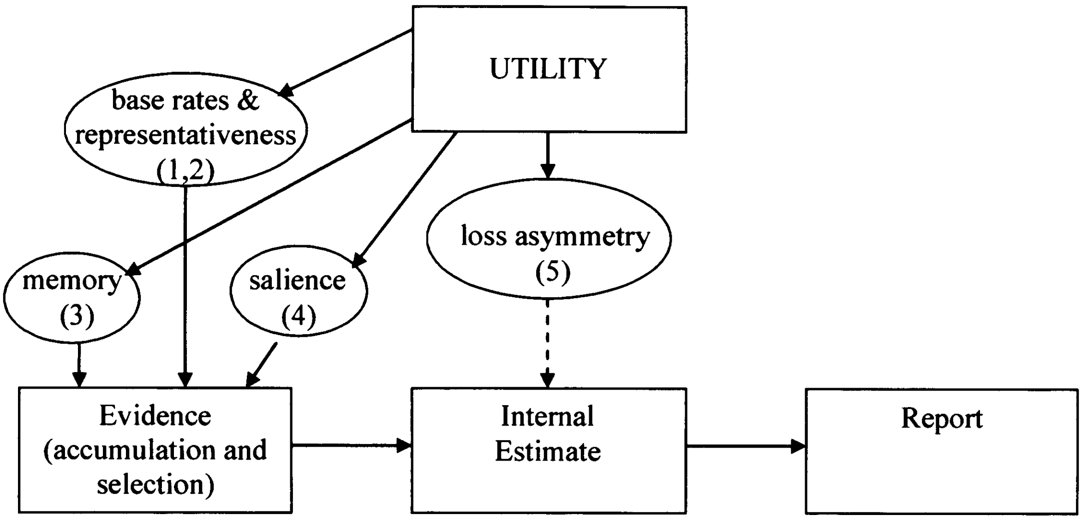
Chapter 3 extended the empirical investigation to potential biasing effects of *positive* utility, and hence provided a controlled, laboratory based, experimental test of the 'wishful thinking' effect. Having developed an asymmetric loss function account of the results in Chapter 2, any effect of positive utility on estimates of probability would have to be explained through an alternative mechanism. Across four empirical studies, however, no effect of utility was observed on participants' probability estimates.

The null results reported in Chapter 3 appeared to contradict a robust result in social psychology, that people are unrealistically optimistic, in that, as a group they tend to judge themselves as less likely than the average person to experience negative life events. In Chapter 4, we argued that the results from the paradigms employed in unrealistic optimism research are confounded by statistical effects resulting from the methodologies employed. We demonstrated that responses from a variety of hypothetical, rational participants would likely be interpreted as unrealistically optimistic according to the rationale of the unrealistic optimism methodology. Whilst some of these hypothetical participants combined imperfect information rationally in order to make their responses, others had perfect knowledge as to whether or not they would experience certain events. In both cases, seemingly rational (or perfectly accurate) individual responses were shown to lead to a bias at the group level. Two original studies (Studies 12 and 13) failed to find evidence for an unrealistic optimism bias over and above effects predicted by the statistical artifacts identified.

In addition to questioning the status of a generally accepted conclusion in the literature, the experimental work in Chapter 4 provided a natural extension of that undertaken in Chapter 3, by further enhancing the personal relevance of the utility manipulation in the experimental materials. Chapter 3 first tested for a wishful thinking effect using purely third person stimuli (Study 8). Studies 9 and 10 also used third person stimuli, but in this instance those stimuli were intended to also carry direct relevance for the global population. Study 11 then failed to find an effect of a direct, first person, affective, manipulation of positive utility (the participant could win a mars bar) on probability estimates. In Chapter 4, Study 12 investigated the potential for a direct biasing effect of utility on people's comparative probability judgments for real events in their own future. The failure to find evidence for a biasing effect of utility under these conditions adds considerable weight to the null results reported in Chapter 3. By questioning participants about their own future, specifically with respect to events rated as either very undesirable or very desirable, the materials are necessarily realistic and personally relevant.

Evidence therefore seems to be mounting that there are *no* direct effects of outcome utility on probability, whether in the positive or the negative domain (a conclusion echoed in Krizan & Windschitl, 2007). The effects of loss function asymmetry in Chapter 2 add to the experimental evidence suggesting the presence of mechanisms that can lead to the *indirect* biasing effect of utility on estimates of probability in practice. Figure 5.1 has been drawn to demonstrate where, in the overall process of producing a probability estimate, the different indirect mechanisms exert their influence. It can be seen that the majority of these mechanisms concern the evidence accumulation stage. Gordon et al. (2005) found that participants misremembering the source of predictions had a tendency to attribute more desirable predictions to the more reliable source. Bar-Hillel et al. (2008) found evidence that wishful thinking influenced information selection via *salience:* "I wish for, therefore I focus on, therefore I believe in" (Bar-Hillel et al., 2008, p. 283). Dai et al. (2008) and Mandel (2008) found evidence for a 'value heuristic', that is, base rate knowledge that the more positive an outcome, the more infrequent it is, which people use as additional evidence where information retrieval is difficult. The vulnerability of this stage of the probability estimation process to biasing factors is not surprising given recent research highlighting the effects of context and task demands on the construction of subjective probabilities (e.g., Lichtenstein & Slovic, 2006).

The failure to observe a *direct* effect of utility on probability estimates suggests that people's judgments of probability are more rational than has previously been assumed. In real life judgment tasks, asymmetric loss functions might inflate estimates of those probabilities associated with negative events, but such a mechanism seems to fulfil a protective function. The identification of the phenomenon (and the



Example Studies:

1: Dai, Wertenbroch, & Brendl (2008)
2: Mandel (2008)
3: Gordon, Franklin, & Beck (2005)
4: Bar-Hillel, Budescu, & Amar (2008)
5: Harris (Chapter 2)

Figure 5.1. Locating indirect effects of utility in the probability estimation process.

others included in Figure 5.1) enables the recognition of those situations in which such inflations might be harmful and steps can subsequently be taken to reduce their negative effects. For judgment and decision making researchers, the seeming lack of a poorly understood, likely unexplainable, general interdependence between utility and probability should inspire optimism. Researchers typically seek rational explanations, which can subsequently guide rational interventions to improve the quality of human judgment and decision making.. Consequently, research such as that presented in this thesis has implications both for theoretical development, in terms of increasing our understanding about human judgment processes, and in applied settings. As mentioned in the introduction, probability judgments about utility laden outcomes are those that are most important in human life. For example, a juror's judgment as to the likelihood of a suspect's guilt, or a clinician's judgment as to the probability of a patient having a potentially life-threatening disease are both situations in which an accurate probability estimate is desirable in order to guide rational decision making. By increasing understanding of the factors that do and do not bias such probability estimates, it might be possible for future research to improve the optimality of people's judgment and decision making. The research presented here provides further support for the rationality of human judgment by demonstrating a bias only under conditions of loss asymmetry, conditions under which such a bias may be considered rational (e.g., Batchelor & Peel, 1998; see also Whiteley & Sahani, 2008, and references therein).

Rational or not rational, the effects of asymmetric loss functions, in conjunction with the other effects cited in Figure 5.1, mean that probability estimates will often be biased in practice. Our loss asymmetry-based influence of severity occurs only in circumstances where a decision might be made. However, it is *only* in situations in which probabilities inform decisions that we really care about the accuracy of estimates in the first place. Moreover, the practical implications seem potentially even greater when the nature of the materials with which we observed this effect is considered. In these studies, participants have no personal stake in the probabilities they are providing, given that the story involves entirely fictitious third parties. Furthermore, there is a clear objective probability that is made available to participants. If a reliable and replicable effect of outcome utility on estimates of probability can be observed within such a minimal paradigm, it is likely that influences of outcome severity on estimates of probability are pervasive and it is likely they will be larger under conditions of emotional involvement (as we experience within our own lives). Finally, the observed bias could operate in conjunction with previously identified biasing influences. This suggests that further investigations under more real-world circumstances are desirable.

These implications highlight the importance of a thorough understanding of the status of the biasing effect of asymmetric loss functions. There might be situations in which such a bias cannot be considered rational. These situations should be explored more fully in future work in order to best understand human judgment, and, where appropriate, to provoke research to investigate interventions designed to enhance its rationality. The next section addresses this question in more detail.

Future Work

As stated above, where biases are observed the natural follow-up is to determine whether they can be, or indeed *should* be attenuated. It seems reasonable to suggest that an optimal judgment strategy would be one in which probabilities are never biased by utilities, for the 'badness' of an event does not affect the likelihood of an event occurring, ceteris paribus. We have, however, argued that the biasing effect of asymmetric loss functions demonstrated in Chapter 2 is a rational response to uncertainty. How can this position be reconciled with the statement that the optimal strategy would be one of no bias? Central to this issue is the recognition that the biasing effect of asymmetric loss functions is a rational response to the uncertainty of the uncertainty. Thus, in situations in which people are more confident about their probability estimates, the bias should be attenuated.

The asymmetric loss function account also predicts that there will be situations in which the effects observed in Chapter 2 can be reversed. In Studies 1, 2, 3, 6 and 7, it is assumed that there are greater costs associated with an underestimate of the probability of the chance outcome occurring (i.e., plane debris landing in a town; the farmer's daughter eating a fatally poisonous apple) than with an overestimate of that probability. In both these situations, such an assumption seems reasonable as there are no costs associated with an underestimate (aside from in Study 6, where these costs are still noticeably less negative than those associated with an overestimate). By reducing the costs associated with an overestimate, and increasing the costs associated with an underestimate, the asymmetry in the loss function might switch such that an underestimate is more costly. For example, in the orchards paradigm, the 'bad' apples could cause severe stomach cramps that last for a 24 hour period, but the only way of stopping the daughter from entering the orchard could be for the farmer to move to a recently vacated farm 50 miles away, at cost to him, and also necessitating him to move his daughter's school. In this instance, the costs associated with an underestimate may be less than those associated with an overestimate of the chance of the daughter eating a bad apple. Were probability estimates subsequently reduced in these situations, this would provide further evidence in favour of the role of asymmetric loss functions on biasing probability estimates.

The asymmetric-loss function explanation for the biasing effect of utility on probability estimates is also a decision-theoretic one, based on the costs associated with an underestimate. Such an effect can only be rational if it is further moderated in the process of actually making a decision, as opposed to simply making a probability estimate. This is because SEU (e.g., Savage, 1954) posits that both probabilities and utilities should be combined in the decision making process. Thus, if utilities influence estimates of probability and are subsequently included in the decision making process, they are effectively being 'double-counted.' Such behaviour could no longer be considered rational according to the normative framework of SEU. Were such 'double-counting' to occur, then this would be an example of a bias for which preventative strategies might usefully be devised. Consequently, it seems an important avenue for future research to address the question of whether utilities may be 'double-counted' in this way. Such a possibility could be investigated by manipulating the actor in the orchards paradigm, and the presence of a decision. In the high controllability condition of Study 6, for example, participants were told that the farmer was considering whether or not to erect an electric fence to keep his daughter from playing in the orchard. In a 'decision-present' condition, the participant could be told that they had an opportunity to erect a fence, they could be asked to estimate the probability and also to make a decision whether or not to erect the fence ('yes' or 'no'). The decision question should be visible on the same page as the one where participants make their probability estimates. If 'double-counting' does occur then participants' responses should be the same in a 'decision-present' condition as in a condition that exactly matches the high controllability condition of Study 6 ('decision-absent'). If 'double-counting' does not occur, suggesting enhanced rationality of the protective bias, then probability estimates should be lower in the 'decision-present' condition (indeed, they should be no different from those in a neutral outcome condition in Chapter 2) than in a 'decision-absent' condition.

Were the latter result to occur, it would not, however, necessarily imply enhanced rationality of the protective bias. DeKay, Patiño-Echeverri and Fischbeck (in press) found that participants who preferred a precautionary action over a non-precautionary action tended to indicate that they would prefer this action regardless of the probabilities involved. According to our decision-theoretic explanation for biased probability estimates, these people would therefore see no reason to alter their probability estimates prior to making their decision, as the probabilities are not relevant to their decision. Clearly, however, the complete ignoring of probabilistic information would be a demonstration of irrationality.

Another area of future research that might be generated by this thesis is further investigation of the true status of unrealistic optimism. We acknowledged in Chapter 4 that, although the present methodology cannot demonstrate the presence of this bias, and we were unable to find an effect within the matrix paradigm (Study 13), we have not provided conclusive evidence against its existence. Future research must develop a suitable paradigm within which this research question can be further addressed – specifically, a paradigm investigating real events in an individual's own future, but one that is not susceptible to the statistical artifacts identified here. In the absence of such a paradigm there are additional questions of applied interest that arise from our work questioning the status of unrealistic optimism. One that we shall focus on here is that of so-called 'depressive realism' (e.g., Alloy & Ahrens, 1987; Pietromonaco & Markus, 1985; Pyszczynski et al., 1987). This phenomenon is partially based on the finding that dysphoric individuals do not display unrealistic optimism to the same degree as non-dysphoric individuals. Given our critique of the status of unrealistic optimism in Chapter 4, it is no longer possible to make this conclusion pertaining to people's expectation of their future. All that can be said is that dysphoric individuals are more negative in their expectations for their future than non-dysphoric individuals. However, the degree to which depressives may be more negative can be investigated with a full study that includes not only rare negative events, but also common negative events, rare positive events and common positive events. The results of Study 12 suggest that non-dysphoric individuals rate their chances of experiencing rare events as less than the average, and their chances of experiencing common events as greater than the average. 'Depressive realism' in relation to rare negative events might suggest that dysphoric individuals are more negative, but it might equally suggest that their responses are simply more regressive towards the mean. An investigation including common negative events and rare positive events comparing dysphoric individuals and non-dysphoric individuals can disassociate the predictions of these two potential explanations for previous results. Such research might have significant implications for our understanding of the cognitive processes underlying depression.

Finally, in the introduction to this thesis, we highlighted the importance of utility considerations for consequential judgments in the real-world. Thus, a pressing extension of the current research is into more real-world settings, as hinted at above. This extension must, however, be undertaken with caution and in conjunction with further laboratory tests. Where inconsistencies are found between results obtained in different settings, reasons for these inconsistencies must be sought. Many extant real-world demonstrations of biased probability estimates resulting from utility considerations may be considered to result from biased information accumulation (see e.g., Bar-Hillel & Budescu, 1995; Gordon et al., 2005). Whilst these findings are important and interesting in their own right, it must be recognised that they are distinct from the question of whether probability estimates are genuinely biased by considerations of outcome utility.

Conclusion

The work in this thesis presented a detailed investigation of the effects of utility on probability estimates. As such this work follows other recent research in reasoning psychology investigating potential 'leakage' from decision-theoretic considerations into reasoning and judgment processes (e.g., Bonnefon, in press). The conclusion from the present research has been something of a 'thumbs-up' for the competence of human probability judgment. Consequently, this work adds to an increasing body of work suggesting, once again (c.f. Peterson & Beach, 1967), that people might meaningfully aspire to the status of intuitive statisticians after all (e.g., Dawes & Mulford, 1996; Erev et al., 1994; Gigerenzer, Hell, & Blank, 1988; Gigerenzer & Hoffrage, 1995; Griffiths & Tenenbaum, 2006; Juslin et al., 2000; Kynn, 2008; Moore & Healy, 2008).

References

Abele, A., & Hermer, P. (1993). Mood influences on health-related judgments: Appraisal of own health versus appraisal of unhealthy behaviors. *European Journal of Social Psychology, 23,* 613-625.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology, 21,* 515-550.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*, 804-825.

Alloy, L. B., & Ahrens, A. H. (1987). Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus other. *Journal of Personality and Social Psychology, 52,* 366-378.

Amgoud, L., Bonnefon, J-F., & Prade, H. (2005). An argumentation-based approach to multiple criteria decision. In L. Godo (Ed.), *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (pp. 269-280), Berlin: Springer Verlag.

Armstrong, J. S. (2001). *Principles of forecasting.* Boston: Kluwer Academic Publishing.

Arnold, J. III (1986). Assessing capital risk: You can't be too conservative. *Harvard Business Review, 64,* 113-121.

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology, 46*, 81-86.

Ayton, P., & Wright, G. (1994). Subjective probability: What should we believe. In G. Wright, & P. Ayton (Eds.), *Subjective Probability* (pp. 163-183). Chichester, UK: Wiley.

Babad, E. (1987). Wishful thinking and objectivity among sports fans. *Social Behaviour, 2,* 231-240.

Babad, E. (1995). Can accurate knowledge reduce wishful thinking in voters' predictions of election outcomes? *The Journal of Psychology, 129,* 285-300.

Babad, E., & Katz, Y. (1991). Wishful thinking–against all odds. *Journal of Applied Social Psychology, 21,* 1921-1938.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From statistical formats to cognitive structures. *Behavioral and Brain Sciences, 30,* 287-292.

Bar-Hillel, M. (1991). Commentary on Wolford, Taylor, and Beck: The conjunction fallacy? *Memory and Cognition, 19,* 412-414.

Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking and Reasoning, 1, 71*-103.

Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008). Predicting World Cup results: Do goals seem more likely when they pay off? *Psychonomic Bulletin & Review, 15,* 278-283.

Batchelor, R., & Peel, D. (1998). Rationality testing under asymmetric loss. *Economics Letters, 61,* 49-54.

Bauman, L. J., & Siegel, K. (1987). Misperception among gay men of the risk for AIDS associated with their sexual behavior. *Journal of Applied Social Psychology, 17,* 329-350.

Bernoulli, J. (1713). *Ars Conjectandi [Probability Calculation].* Basel, Switzerland: Thurnisorum.

Berscheid, E. (1983). Emotion. In H. H. Kelley, E. Berscheid, A. Christensen, J. H. Harvey, T. L. Huston, G. Levinger, E. McClintock, L. A. Peplau, & D. R. Peterson (Eds.), *Close Relationships* (pp. 110-168). San Francisco: Freeman.

Beyth-Marom, R. (1982). How probable is probable? a numerical translation of verbal probability expressions. *Journal of Forecasting, 1*, 257-269.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85-94.

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48,* 87-106.

Birnbaum, M. H. (2004b). Human research and data collection via the internet. *Annual Review of Psychology, 55,* 803-832.

Birnbaum, M. H., Coffey, G., Mellers, B. A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 331-346.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37*, 48-74.

Blake, R. R., & Mouton, J. S. (1961). Comprehension of own and of outgroup positions under intergroup competition. *Journal of Conflict Resolution, 5,* 304-310.

Bonnefon, J-F. (in press). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review.*

Bonnefon, J-F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science, 17,*747-751.

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifiers: Usage and interpretation. *Journal of Experimental Psychology, 102,* 335-338.

Bottom, W. P., & Paese, P. W. (1999). Judgment accuracy and the asymmetric cost of errors in distributive bargaining. *Group Decision and Negotiation, 8,* 349-364.

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology.* Oxford, UK: Oxford University Press.

Bradburn, N. M. (1969). *The Structure of Psychological Well-Being.* Chicago, IL: Aldine.

Brand, A. (2005). iPsychExpts: Web experiments for psychologists. *http://www.ipsychexpts.com.*

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113,* 409-432.

Brase, G. L. (2008). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review, 15,* 284-289.

Brown, C. E. (1982). A false consensus bias in 1980 presidential preferences. *Journal of Social Psychology, 118,* 137-138.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* Cambridge, UK: Cambridge University Press.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: general principles and empirical evidence. In J. Busemeyer, & R. Hastie (Eds.), *The Psychology of learning and motivaton: Decision making from a cognitive perspective* (pp. 275-318). San Diego, CA: Academic Press.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 281-294.

Buehler, R., & Griffin, D. (2003). Planning, personality, and prediction: The role of future focus in optimistic time predictions. *Organizational Behavior and Human Decision Processes, 92,* 80-90.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychological Bulletin, 23,* 238-247.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67,* 366-381.

Burger, J. M., & Burns, L. (1988). The illusion of unique invulnerability and the use of effective contraception. *Personality and Social Psychology Bulletin, 14,* 264-270.

Campbell, J., Greenauer, N., Macaluso, K., & End, C. (2007). Unrealistic optimism in internet events. *Computers in Human Behavior, 23,* 1273-1284.

Cantin, I., & Dubé, L. (1999). Attitudinal moderation of correlation between food liking and consumption. *Appetite, 32,* 367-381.

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299,* 999-1001.

Causse, P., Delhomme, P., & Kouabenan, D. R. (2005). Absolute and comparative judgements of two situated road risk and reasons evocated for these judgements. *Psychologie Du Travail et Des Organisations, 11,* 191-208.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*, 813-838.

Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, event frequency, and comparative optimism: When what happens frequently is "more likely to happen to me". *Personality and Social Psychology Bulletin, 29*, 1343-1356.

Chang, E. C., Asakawa, K., & Sanna, L. J. (2001). Cultural variations in optimistic and pessimistic bias: Do Easterners really expect the worst and Westerners really expect the best when predicting future life events. *Journal of Personality and Social Psychology, 81*, 476-491.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences, 2*, 206-214.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*, 335-344.

Chater, N., & Oaksford, M. (Eds.) (2008). *The Probabilistic Mind: Prospects for Bayesian Cognitive Science.* Oxford, UK: Oxford University Press.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: onceptual foundations. *Trends in Cognitive Sciences, 10*, 287-291.

Christensen, I. P. (1979). Distributional and non-distributional uncertainty. In C. R. Bell (Ed.), *Uncertain Outcomes* (pp. 49-59). Lancaster, UK: MTP Press Ltd.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist, 39*, 75-78.

Clutterbuck, C. S. (2008). Depression as a moderator of the optimism bias: "Realism" or pessimism-in-diguise?. Final year project, BSc Psychology. Cardiff University.

Cohen, B. L., & Wallsten, T. S. (1991). The effect of outcome desirability on comparisons of numerical and linguistic probabilities. *Journal of Behavioral Decision Making, 5,* 53-72.

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd Edition).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., Dearnley, E. J., & Hansel, C. E. M. (1958). A quantitative study of meaning. *British Journal of Educational Psychology, 28,* 141-148.

Cohn, L. D., Macfarlane, S., Yanez, C., & Imai, W. K. (1995). Risk-perceptions: Differences between adolescents and adults. *Health Psychology, 14*, 217-222.

Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepeneurs' perceived chances for success. *Journal of Business Venturing, 3,* 97-108.

Cornell, C. E., Rodin, J., & Weingarten, H. (1989). Stimulus-induced eating when satiated. *Physiology and Behavior, 45,* 695-704.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1-73.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38,* 668-678.

Coventry, K. R., & Norman, A. C. (1998). Arousal, erroneous verbalizations and the illusion of control during a computer-generated gambling task. *British Journal of Psychology, 89,* 629-645.

Covey, J. A., & Davies, A. D. M. (2004). Are people unrealistically optimistic? It depends how you ask them. *British Journal of Health Psychology, 9,* 39-49.

Cox, R. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics, 14,* 1-13.

Crandall, V. J., Solomon, D., & Kellaway, R. (1955). Expectancy statements and decision themes as functions of objective probabilities and reinforcement values. *Journal of Personality, 24,* 192-203.

Dai, X., Wertenbroch, K., & Brendl, C. M. (2008). The value heuristic in judgments of relative frequency. *Psychological Science, 19,* 18-20.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65,* 201-211.

De Finetti, B. (1974). *Theory of Probability.* New York: John Wiley & Sons.

DeKay, M. L., Patiño-Echeverri, D., & Fischbeck, P. S. (in press). Distortion of probability and outcome information in risky decisions. *Organizational Behavior and Human Decision Processes.*

Delfabbro, P. (2004). The stubborn logic of regular gamblers: Obstacles and dilemmas in cognitive gambling research. *Journal of Gambling Studies, 20,* 1-21.

Delfabbro, P. H., & Winefield, A. H. (2000). Predictors of irrational thinking in slot-machine gambling. *Journal of Psychology, 134,* 17-28.

Dewbery, C., Ing, M., James, S., Nixon, M., & Richardson, S. (1990). Anxiety and unrealistic optimism. *Journal of Social Psychology, 130,* 151-156.

Dewbery, C., & Richardson, S. (1990). Effect of anxiety on optimism. *Journal of Social Psychology, 130,* 731-738.

Drake, R. A. (1984). Lateral asymmetry of personal optimism. *Journal of Research in Personality, 18,* 497-507.

Drake, R. A. (1987). Conception of own versus others' outcomes: Manipulation of monaural attentional orientation. *European Journal of Social Psychology, 17,* 373-375.

Drake, R. A., & Ulrich, G. (1992). Line bisecting as a predictor of personal optimism and desirability of risky behavior. *Acta Psychcologica, 79,* 219-226.

Dubois, D., & Prade, H. (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* New York: Plenum Press.

Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition, 9,* 85-110.

Edwards, W. (1953). Probability preferences in gambling. *American Journal of Psychology, 66,* 349-364.

Edwards, W. (1962). Utility, subjective probability, their interaction, and variance preferences. *Journal of Conflict Resolution, 6,* 42-51.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal Representations of Human Judgment* (pp. 17-52). New York: Wiley.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92,* 433-461.

Eiser, J. R., Eiser, C., & Pauwels, P. (1993). Skin cancer: Assessing perceived risk and behavioural attitudes. *Psychology and Health*, 8, 393-404.

Erev, I., & Cohen, B. L. (1990). Verbal verus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes, 45,* 1-18.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519-527.

Evans, J. St. B. T., Handley, S. H., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition, 77,* 197-213.

Eysenck, M. W., & Derakshan, N. (1997). Cognitive biases for future negative events as a function of trait anxiety and social desirability. *Personality and Individual Differences, 22,* 597-605.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science, 299*, 1898-1902.

Fischer, K., & Jungermann, H. (1996). Rarely occurring headaches and rarely occurring blindness: Is rarely = rarely? Meaning of verbal frequentistic labels in specific medical contexts. *Journal of Behavioral Decision Making, 9,* 153-172.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*, 239-260.

Fisk, J. E., & Pidgeon, N. (1996). Component probabilities and the conjunction fallacy: Resolving signed summation and the low component model in a contingent approach. *Acta Psychologica, 94,* 1-20.

Fisk, J. E., & Pidgeon, N. (1997). The conjunction fallacy: The case for the existence of competing heuristic strategies. *British Journal of Psychology, 88,* 1-27.

Fisk, J. E., & Pidgeon, N. (1998). Conditional probabilities, potential surprise, and the conjunction fallacy. *The Quarterly Journal of Experimental Psychology, 51A,* 655-681.

Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making, 1*, 159-161.

Fox, J., Krause, P., & Ambler, S. (1992). Arguments, contradictions and practical reasoning. In B. Neumann (Ed.), *Proceedings of the 10th European Conference on Artificial Intelligence* (pp. 623-627). Chichester, UK: John Wiley and Sons.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116,* 439-453.

Gendolla, G. H. E. (1997). Surprise in the context of achievement: The role of outcome valence and importance. *Motivation and Emotion, 21,* 165-193.

Gerrard, M., Gibbons, F. X., Benthin, A. C., & Hessling, R. M. (1996). A longitudinal study of the reciprocal nature of risk behaviors and cognitions in adolescents: What you do shapes what you think, and vice versa. *Health Psychology, 15*, 344-354.

Gerrard, M., Gibbons, F. X., & Bushman, B. J. (1996). Relation between perceived vulnerability to HIV and precautionary sexual behavior. *Psychological Bulletin, 119*, 390-409.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright, & P. Ayton (Eds.), *Subjective Probability* (pp. 129-162). Chichester, UK: Wiley.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8,* 53-96.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513-525.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684-704.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart.* Oxford: Oxford University Press.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition, 78,* 247-276.

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106*, 325-344.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior.* Garden City, NY: Anchor Books.

Goodwin, P. (1996). Subjective correction of judgemental point forecasts and decisions. *Omega, 24,* 551-559.

Gordon, R., Franklin, N., & Beck, J. (2005) Wishful thinking and source monitoring. *Memory and Cognition*, 33, 418-429.

Granberg, D., & Brent, E. (1983). When prophecy bends: the preference-expectation link in U.S. presidential elections, 1952-1980. *Journal of Personality and Social Psychology, 45*, 477-491.

Granger, C. (1969). Prediction with a generalised cost of error function. *Operational Research Quarterly, 20,* 199-207.

Grice, H. P. (1975/2001). Logic and conversation. In A. P. Martinich (Ed.), *The Philosophy of Language* (4th ed., pp. 165-175). Oxford: Oxford University Press.

Grice, H. P. (2001). *Aspects of Reason.* Oxford: Clarendon Press.

Griffiths, M. D. (1994). The role of cognitive bias and skill in fruit machine gambling. *British Journal of Psychology, 85,* 351-369.

Griffiths, M. D. (1995). *Adolescent Gambling.* London: Routledge.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 334-384.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767-773.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116,* 454-461.

Hampson, S. E., Andrews, J. A., Barckley, M., Lichtenstein, E., & Lee, M. E. (2000). Conscientiousness, perceived risk, and risk-reduction behaviors: A preliminary study. *Health Psychology, 19,* 496-500.

Hansen, W. B., & Malotte, C. K. (1986). Perceived personal immunity: The development of beliefs about susceptibility to the consequences of smoking. *Preventive Medicine, 15,* 363-372.

Harris, A. J. L., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1366-1373.

Harris, D. M., & Guten, S. (1979). Health-Protective Behavior: An exploratory study *Journal of Health and Social Behavior, 20*, 17-29.

Harris, P., & Middleton, W. (1994). The illusion of control and optimism about health: On being less at risk but no more in control than others. *British Journal of Social Psychology, 33*, 369-386.

Harris, P. R., Griffin, D. W., & Murray, S. (2008). Testing the limits of optimistic bias: Event and person moderators in a multilevel framework. *Journal of Personality and Social Psychology, 95,* 1225-1237.

Hattori, M., & Nishida, Y. (in press). Why does the base rate appear to be ignored? The equiprobability hypothesis. *Psychonomic Bulletin and Review.*

Hawe, E., Talmud, P. J., Miller, G. J., & Humphries, S. E. (2003). Family history is a coronary heart disease risk factor in the second Northwick Park heart study. *Annals of Human Genetics, 67,* 97-106.

Heine, S. J., & Lehman, D. R. (1995). Cultural variation in unrealistic optimism: Does the West feel more invulnerable than the East? *Journal of Personality and Social Psychology*, 68, 595-607.

Helweg-Larsen, M. (1999). (The lack of) optimistic biases in response to the Northridge earthquake: The role of personal experience. *Basic and Applied Social Psychology, 21,* 119-129.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 5,* 74-95.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534-537.

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of *and. Cognition, 108,* 740-753.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors *Journal of Behavioral Decision Making, 12*, 275-305.

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 621-642.

Hertwig, R., Zangerl, M. A., Biedert, E., & Margraf, J. (2008). The public's probabilistic numeracy: How tasks, education and exposure to games of chance shape it. *Journal of Behavioral Decision Making, 21,* 457-470.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*, 248-271.

Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science, 36,* 780-803.

Hoorens, V., & Buunk, B. P. (1993). Social comparison of health risks: Locus of control, the person-positivity bias, and unrealistic optimism. *Journal of Applied Social Psychology, 23,* 291-302.

Hoorens, V., Smits, T., & Shepperd, J. A. (2008). Comparative optimism in the spontaneous generation of future life-events. *British Journal of Social Psychology, 47,* 441-451.

Horvitz, E. J., Heckerman, D., & Langlotz, C. P. (1986). A framework for comparing alternative formalisms for plausible reasoning. In *Proceedings AAAI-86 Fifth National Conference on Artificial Intelligence* (pp. 210-214). San Mateo, CA: Morgan Kaufmann.

Howell, D. C. (1997). *Statistical Methods for Psychology (4th Edition).* Belmont, CA: Duxbury Press.

Howson, C., & Urbach, P. (1996). *Scientific Reasoning: The Bayesian Approach (2nd Edition).* Chicago, IL: Open Court.

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality, 21,* 329-335.

Irwin, F. W., & Graae, C. N. (1968). Tests of the discontinuity hypothesis of the effects of independent outcome values upon bets. *Journal of Experimental Psychology, 76,* 444-449.

Irwin, F. W., & Snodgrass, J. G. (1966). Effects of independent and dependent outcome values upon bets. *Journal of Experimental Psychology, 71,* 22-285.

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. W. jr. & T. K. Skrull (Eds.), *Handbook of Social Cognition* (Vol. 3, pp. 179-236). Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62-88.

Jourden, F. J., & Heath, C. (1996). The evaluation gap in performance perceptions: Illusory perceptions of groups and individuals. *Journal of Applied Psychology, 81,* 369-379.

Juslin, P., Winman, A., & Olsson, H. (2000). Naïve empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107,* 384-396.

Kadane, J. B., & Winkler, R. L. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association, 83,* 357-363.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science, 39,* 17-31.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

Kahneman, D., & Tversky, A. (1979a). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science, 12,* 313-327.

Kahneman, D., & Tversky, A. (1979b). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11,* 123-141.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773-795.

Keppel, G., Saufley, W. H. Jr., & Tokunaga, H. (1992). *Introduction to Design and Analysis: A Student's Handbook (2nd Edition).* New York: W. H. Freeman and Company.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77,* 217-273.

Kirscht, J. F., Haefner, D. P., Kegeles, S. S., & Rosenstock, I. M. (1966). A national study of health beliefs. *Journal of Health & Human Behaviour, 7,* 248-254.

Klar, Y., & Ayal, S. (2004). Event frequency and comparative optimism: Another look at the indirect elicitation method of self-others risks. *Journal of Experimental Social Psychology, 40,* 805-814.

Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25,* 585-594.

Klar, Y., Medding, A., & Sarel, D. (1996). Nonunique invulnerability: Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes, 67*, 229-245.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *The journal of Neuroscience, 25*, 4806-4812.

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10,* 319-326.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin, 133,* 95-121.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77,* 221-232.

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40*, 332-340.

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology, 44,* 220-232.

Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society A, 171*, 239-264.

Lachman, M. E., Röcke, C., Rosnick, C., & Ryff, C. D. (2008). Realism and illusion in Americans' temporal views of their life satisfaction: Age differences in reconstructing the past and anticipating the future. *Psychological Science, 19,* 889-897.

Ladouceur, R., Gaboury, A., Dumont, D., & Rochette, P. (1988). Gambling: Relationship between the frequency of wins and irrational thinking. *Journal of Psychology, 122,* 409-414.

Landy, M. S., Goutcher, R., Trommershäuser, J., & Mamassian, P. (2007). Visual estimation under risk. *Journal of Vision, 7*(6):4, 1-15.

Langer, E. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*, 311-328.

Langer, E.J. & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task *Journal of Personality and Social Psychology, 32*, 951-955.

Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting, 21,* 3-14.

Lawrence, M., O'Connor, M., & Edmundson, R. (2000). A field study of forecasting accuracy. *European Journal of Operational Research, 122,* 151-160.

Lee, C. (1989). Perceptions of immunity to disease in adult smokers. *Journal of Behavioral Medicine, 12,* 267-277.

Lek, Y., & Bishop, G. D. (1995). Perceived vulnerability to illness threats: The role of disease type, risk factor perception and attributions. *Psychology and Health, 10*, 205-217.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81,* 146-159.

Letarte, A., Dubé, L., & Troche, V. (1997). Similarities and differences in affective and cognitive origins of food likings and dislikes. *Appetite, 28*, 115-129.

Levy, A. G., & Hershey, J. C. (2008). Value-induced bias in medical decision making. *Medical Decision Making, 28*, 269-276.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and biases* (pp. 306-334). Cambridge: Cambridge University Press.

Lichtenstein, S., & Slovic, P. (2006). *The Construction of Preference.* Cambridge, UK: Cambridge University Press.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 551-578.

Lindley, D. V. (1982). Scoring rules and the inevitability of probability. *International Statistical Review, 50,* 1-26.

Lindley, D. (1994). Foundations. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 3-15). Chichester, UK: John Wiley & Sons

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92,* 805-824.

Lorch, R. F., & Myers, J. L. (1990). Regression analysis of repeated measures data in cognitive psychology. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 149-157.

Lund, F. H. (1925). The psychology of belief. *Journal of Abnormal and Social Psychology, 20,* 63-81; 174-196.

Macdonald, R. R. (1986). Credible conceptions and implausible probabilities. *British Journal of Mathematical and Statistical Psychology, 39,* 15-27.

Maguire, P., & Maguire, R. (2009). Investigating the difference between surprise and probability judgments. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings* *of the 31st Annual Conference of the Cognitive Science Society* (pp. 2359-2364). Austin, TX: Cognitive Science Society.

Mandel, D. R. (2008). Violations of coherence in subjective probability: A representational and assessment processes account. *Cognition, 106,* 130-156.

Marks, R. W. (1951). The effect of probability, desirability and 'privilege' on the stated expectations of children. *Journal of Personality, 19,* 332-351.

Markus, H., & Zajonc, R. B. (1985). The cognitive perspective in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (3rd edition, Vol. 1, pp. 137-230). New York: Lawrence Erlbaum Associates.

Marteau, T. M. (1989). Framing of information: Its influence upon decisions of doctors and patients. *British Journal of Social Psychology, 28,* 89-94.

Martha, C., Sanchez, X., & Gomà-i-Freixanet, M. (2009). Risk perception as a function of risk exposure amongst rock climbers. *Psychology of Sport and Exercise, 10,* 193-200.

Matute, H., Vadillo, M. A., Blanco, F. Musca, S. C. (2007). Either greedy or well-informed: The reward maximization – unbiased evaluation trade-off. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of The European Cognitive Science Conference* (pp. 341-346). Hove, UK: Lawrence Erlbaum Associates.

McGuire, W. J. (1960a). Cognitive consistency and attitude change. *Journal of Abnormal and Social Psychology, 60,* 345-353.

McGuire, W. J. (1960b). Direct and indirect persuasive effects of dissonance-producing messages. *Journal of Abnormal and Social Psychology, 60,* 354-358.

McKenna, F. P. (1993). It won't happen to me: Unrealistic optimism or illusion of control? *British Journal of Psychology, 84*, 39-50.

McKenna, F. P., Warburton, D. M., & Winwood, M. (1993). Exploring the limits of optimism: The case of smokers' decision making. *British Journal of Psychology, 84,* 389-394.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review, 10,* 596-602.

McNeil, B. J., Pauker, S. G., Sox, H. C., Jr., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine, 306,* 1259-1262.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? *Psychological Science, 12,* 269-275.

Mellers, B. A., & McGraw, P. A. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage, 1995. *Psychological Review, 106,* 417-424.

Merz, J. F., Druzdzel, M. J., & Mazur, D. J. (1991). Verbal expressions of probability in informed consent litigation. *Medical Decision Making, 11,* 273-281.

Millar, M. G., & Millar, K. U. (1990). Attitude change as a function of attitude type and argument type. *Journal of Personality and Social Psychology, 59,* 217-228.

Miller, A. G., Ashton, W. A., & McHoskey, J. W. (1990). What price attractiveness? Stereotype and risk factors in suntanning behavior. *Journal of Applied Social Psychology, 20,* 1272-1300.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502-517.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92,* 972-989.

Moore, D., & Small, D. (2008). When it is rational for the majority to believe that they are better than average. In J. I. Krueger (Ed.), *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes* (pp. 141-174). New York: Psychology Press.

Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task-specific phenomenon. *Personality and Social Psychology Bulletin, 10,* 243-252.

Morlock, H. (1967). The effect of outcome desirability on information required for decisions. *Behavioral Science, 12,* 296-300.

Morlock, H. C., & Hertz, K. J. (1964). Effect of the desirability of outcomes on decision making. *Psychological Reports, 14,* 11-17.

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron, 43*, 133-143.

Mullet, E., & Rivet, I. (1991). Comprehension of verbal probability expressions in children and adolescents. *Language and Communication, 11*, 217-225.

Murphy, A. H., & Daan, H. (1984). Impacts of feedback and experience on the quality of subjective probability forecasts: Comparison of results from the first and second years of the Zierikzee experiment. *Monthly Weather Review, 112*, 413-423.

Murphy, A. H., & Winkler, R. L. (1982). Subjective probabilistic tornado forecasts: Some experimental results. *Monthly Weather Review, 110*, 1288-1297.

Nagel, S., Lamm, D., & Neef, M. (1981). Decision theory and juror decision-making. In B. D. Sales (Ed.), *The Trial Process*. New York, NY: Plenum Press.

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences, 32*, 69-120.

Oaksford, M., Chater, N., & Stewart, N. (in press). Reasoning and decision making. In K. Frankish & W. Ramsey (Eds.), *The Cambridge handbook of cognitive science*. Cambridge: Cambridge University Press.

Office for National Statistics (2000). Registrations of cancer diagnosed in 1994-1997, England and Wales. *Health Statistics Quarterly, 7,* 71-82.

Otten, W., & van der Pligt, J. (1996). Context effects in the measurement of comparative optimism in probability judgments. *Journal of Social and Clinical Psychology, 15,* 80-101.

Parsons, S. (2001). *Qualitative Methods for Reasoning Under Uncertainty.* Cambridge, MA: The MIT Press.

Pawlak, Z. (1982). Rough sets. *International Journal of Information and Computer Sciences, 11,* 341-356.

Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the Juror: The psychology of juror decision making* (pp. 192-221). Cambridge: Cambridge University Press.

Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology, 50,* 502-510.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29-46.

Peterson, C. R., & Miller, A. J. (1965). Sensitivity of subjective probability revision. *Journal of Experimental Psychology, 70*, 117-121.

Peterson, C. R., Schneider, R. J., & Miller, A. J. (1965). Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology, 69*, 522-527.

Pfeifer, P. E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes, 58,* 203-213.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72*, 346-354.

Phillips, L. D., Hays, W. L., & Edwards, W. (1966). Conservatism in complex probabilistic inference. *IEEE Transactions on Human Factors in Electronics, HFE-7,* 7-18.

Pietromonaco, P. R., & Markus, H. (1985). The nature of negative thoughts in depression. *Journal of Personality and Social Psychology, 48,* 799-807.

Politzer, G., & Bonnefon, J-F. (2009). Let us not put the probabilistic cart before the uncertainty bull. *Behavioral and Brain Sciences, 32,* 100-101.

Politzer, G., & Noveck, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research, 20,* 83-103.

Price, P. C. (2000). Wishful thinking in the prediction of competitive outcomes. *Thinking and Reasoning, 6,* 161-172.

Price, P. C., Pentecost, H. C., & Voth, R. D. (2002). Perceived event frequency and the optimistic bias: Evidence for a two-process model of personal risk judgments. *Journal of Experimental Social Psychology, 38*, 242-252.

Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality and Social Psychology, 90,* 382-398.

Pruitt, D. G., & Hoge, R. D. (1965). Strength of the relationship between the value of an event and its subjective probability as a function of the method of measurement. *Journal of Experimental Psychology, 69,* 483-489.

Pyszczynski, T., Holt, K., & Greenberg, J. (1987). Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. *Journal of Personality and Social Psychology, 52*, 994-1001.

Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank-Dependent model.* Boston: Kluwer.

Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. *Annual Review of Psychology, 23,* 131-176.

Regan, P. C., Snyder, M., & Kassin, S. M. (1995). Unrealistic optimism, self-enhancement or person positivity. *Personality and Social Psychology Bulletin, 21*, 1073-1082.

Rescher, N. (1976). *Plausible reasoning.* Amsterdam: Van Gorcum.

Rise, J., Strype, J., & Sutton, S. (2002). Comparative risk ratings and lung cancer among Norwegian smokers. *Addiction Research and Theory, 10,* 313-320.

Robins, R. W., & Craik, K. H. (1993). Is there a citation bias in the judgment and decision literature. *Organizational Behavior and Human Decision Processes, 54*, 225-244.

Rogers, P., & Webley, P. (2001). "It could be us!": Cognitive and social psychological factors in UK National Lottery play. *Applied Psychology: An International Review, 50,* 181-199.

Rose, J. P., Endo, Y., Windschitl, P. D., & Suls, J. (2008). Cultural differences in unrealistic optimism and pessimism: The role of egocentrism and direct versus indirect comparison measures. *Personality and Social Psychology Bulletin, 34,* 1236-1248.

Rosenthal, R. (1979). The "File Drawer Problem" and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* London: Sage Publications.

Rosnow, R. L., & Rosenthal, R. (1995). "Some things you learn aren't so": Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science, 6,* 3-9.

Rothman, A. J., & Kiviniemi, M. T. (1999). Treating people with information: an analysis and review of approaches to communicating health risk information. *Journal of the National Cancer Institute Monograph, 25,* 44-51.

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: on the affective psychology of risk. *Psychological Science, 12,* 185-190.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16,* 225-237.

Ryff, C. D. (1989). In the eye of the beholder: Views of psychological well-being among middle-aged and older adults. *Psychology and Aging, 4,* 195-210.

Salovey, P., & Birnbaum, D. (1989). Influence of mood on health-relevant cognitions. *Journal of Personality and Social Psychology, 57,* 539-551.

Sanfey, A. G. (2007). Decision neuroscience: New directions in studies of judgment and decision making. *Current Directions in Psychological Science, 16*(3), 151-155.

Savage, L. J. (1954). *The Foundations of Statistics.* New York: John Wiley & Sons.

Schum, D. A. (1988). Probability and the processes of discovery, proof, and choice. In P. Tillers, & E. D. Green (Eds.), *Probability and inference in the law of evidence: the uses and limits of Bayesianism* (pp. 213-270). Dordrecht, NL: Kluwer Academic Publishers.

Schum, D. A. (1994). *The Evidential Foundations of Probabilistic Reasoning.* New York: John Wiley & Sons.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press.

Shaffer, V. A., & Arkes, H. R. (in press). Preference reversals in evaluations of cash versus non-cash incentives. *Journal of Economic Psychology.*

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology, 70,* 844-855.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101,* 467-494.

Sher, S., & McKenzie, C. R. M. (2008). Framing effects and rationality. In N. Chater, & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 79-96). Oxford, UK: Oxford University Press.

Sherif, M., & Sherif, C. W. (1956). *An Outline of Social Psychology (revised edition).* New York: Harper & Brothers.

Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences, 23,* 351-379.

Sloman, S. A., & Over, D. E. (2003). Probability judgement from the inside and out. In D. E. Over (Ed.), *Evolution and the Psychology of Thinking: The Debate* (pp. 145-169). Hove, UK: Psychology Press.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91,* 296-309.

Slovic, P. (1966). Value as a determiner of subjective probability. *IEEE Transactions on Human Factors in Electronics, HFE-7,* 22-28.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases (pp. 463-489).* Cambridge, UK: Cambridge University Press.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of human information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649-744.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48,* 813-838.

Smits, T., & Hoorens, V. (2005). How probable is probably? It depends on whom you're talking about. *Journal of Behavioral Decision Making, 18*, 83-96.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65,* 117-137.

Solow, A. R., & Broadus, J. M. (1988). A simple model of over-forecasting. *Monthly Weather Review, 116,* 1371-1373.

Snow, P. (1998). On the correctness and reasonableness of Cox's Theorem for finite domains. *Computational Intelligence, 14,* 452-459.

Spohn, W. (1990). A general non-probabilistic theory of inductive reasoning. In R. D. Shachter, T. S. Levitt, L. N. Kanal, & J. F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 4* (pp. 149-158). Amsterdam: Elsevier.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association, 54,* 30-34.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1-26.

Stewart, N., & Simpson, K. (2008). A decision-by-sampling account of decision under risk. In N. Chater & M. Oaksford. (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 261-276). Oxford: Oxford University Press.

Steyvers, M., Tenenbaum, J., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27,* 453-489.

Sutton, S. R. (1999). How accurate are smokers' perceptions of risk? *Health, Risk and Society, 1,* 223-230.

Sutton, S. R. (2002). Influencing optimism in smokers by giving information about the average smoker. *Risk Decision and Policy, 7,* 165-174.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143-148.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110,* 67-85.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Taylor, K. M., & Shepperd, J. A. (1998). Bracing for the worst: Severity, testing and feedback as moderators of optimistic bias. *Personality and Social Psychology Bulletin*, 24, 915-926.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Science, 104*, 19156-19159.

Teigen, K. H., & Keren, G. (2002). When are successes more surprising than failures? *Cognition and Emotion, 16,* 245-268.

Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts? *Cognition, 87,* 55-71.

Tellegen, A. (1984). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma, & J. D. Maser (Eds.), *Anxiety and the Anxiety Disorders.* Hillsdale, NJ: Erlbaum.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10,* 309-318.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science, 28,* 467-477.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435-502.

Thagard, P. (2000). *Coherence in thought and action.* Cambridge, MA: MIT Press.

Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence, 18,* 231-249.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *Journal of Neurophysiology, 97,* 1621-1632.

Trommershäuser, J., Maloney, L. T. & Landy, M. S. (2003), Statistical decision theory and rapid, goal-directed movements. *Journal of the Optical Society A*, 1419-1433.

Trommershäuser, J., Maloney, L. T. & Landy M. S. (2008), The expected utility of movement. In P. Glimcher, C. Camerer, E. Fehr, & R. Poldrack (Eds.), *Neuroeconomics* (pp. 95-111). New York: Academic Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453-458.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty:*

Heuristics and Biases (pp. 84-98). Cambridge, UK: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90*, 293-315.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science, 20,* 473-479.

van der Velde, F. W., Hooykas, C., & van der Pligt, J. (1992). Risk perception and behavior: Pessimism, realism, and optimism about AIDS-related health behavior. *Psychology and Health, 6,* 23-38.

van der Velde, F. W., & van der Pligt, J. (1991). AIDS-related health behavior: Coping, protection motivation, and previous behavior. *Journal of Behavioral Medicine, 14,* 429-451.

van der Velde, F. W., van der Pligt, J., & Hooykas, C. (1994). Perceiving AIDS-related risk: Accuracy as a function of differences in actual risk. *Health Psychology, 13*, 25-33.

Verplanken, B. (1997). The effect of catastrophe potential on the interpretation of numerical probabilities of the occurrence of hazards. *Journal of Applied Social Psychology, 27,* 1453-1467.

Wagenaar, W. A. (1988). *Paradoxes of Gambling Behavior.* Hove, UK: Erlbaum.

Walker, M. B. (1992). Irrational thinking among slot machine players. *Journal of Gambling Studies, 8,* 245-288.

Wallsten, T. S. (1981). Physician and medical student bias in evaluating diagnostic information. *Medical Decision Making, 1*, 145-164.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science, 29,* 151-173.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348-365.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science, 39*, 176-190.

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society, 31,* 135-138.

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base-rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language, 25,* 571-587.

Walter, F. M., & Emery, J. (2006). Perceptions of family history across common diseases: A qualitative study in primary care. *Family Practice, 23,* 472-480.

Warr, P., Barter, J., & Brownbridge, G. (1983). On the independence of positive and negative affect. *Journal of Personality and Social Psychology, 44,* 644-651.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465-490.

Watson, D., Clark, L. A., & Tellegen, A. (1984). Cross-cultural convergence in the structure of mood: A Japanese replication and a comparison with U.S. findings. *Journal of Personality and Social Psychology, 47,* 127-144.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063-1070.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219-235.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 115,* 228-242.

Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 781-789.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality & Social Psychology, 39,* 806-820.

Weinstein, N. D. (1982). Unrealistic optimism about susceptibility to health problems. *Journal of Behavioural Medicine, 5,* 441-460.

Weinstein, N. D. (1984). Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology, 3,* 431-457.

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine, 10*, 481-500.

Weinstein, N. D. (1989a). Optimistic biases about personal risks. *Science, 246,* 1232-1233.

Weinstein, N. D. (1989b). Effects of personal experience on self-protective behavior. *Psychological Bulletin, 105,* 31-50.

Weinstein, N. D. (1999). What does it mean to understand a risk? Evaluating risk comprehension. *Journal of the National Cancer Institute Monographs, 25,* 15-20.

Weinstein, N. D. (2000). Perceived probability, perceived severity, and health-protective behavior. *Health Psychology, 19,* 65-74.

Weinstein, N. D., & Klein, W. M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychology, 14*, 132-140.

Weinstein, N. D., & Klein, W. M. (1996). Unrealistic optimism: Present and future. *Journal of Social and Clinical Psychology, 15*, 1-8.

Weinstein, N. D., & Lachendro, E. (1982). Egocentrism as a source of unrealistic optimism. *Personality and Social Psychology Bulletin, 8,* 195-200.

Weinstein, N. D., Marcus, S. E., & Moser, R. P. (2005). Smokers' unrealistic optimism about their risk. *Tobacco Control, 14,* 55-59.

Weinstein, N. D., Sandman, P. M., & Roberts, N. E. (1990). Determinants of self-protective behaviour: Home radon testing. *Journal of Applied Social Psychology*, *20*, 783-801.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5,* 598-604.

Welkenhuysen, M., Evers-Kiebooms, G., Decruyenaere, M., & Van Den Berghe, H. (1996). Unrealistic optimism and genetic risk. *Psychology and Health, 11,* 479-492.

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision, 8*(3):2, 1-15.

Whitley, B. E., & Hern, A. L. (1991). Perceptions of vulnerability to pregnancy and the use of effective contraception. *Personality and Social Psychology Bulletin, 17,* 104-110.

Wilson, D. K., Kaplan, R. M., & Schneidermann, L. J. (1987). Framing of decisions and selections of alternatives in health care. *Social Behaviour, 2,* 51-59.

Windschitl, P. D. (2002). Judging the accuracy of a likelihood judgment: The case of smoking risk. *Journal of Behavioral Decision Making, 15,* 19-35.

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (in press). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes.*

Windschitl, P. D., & Weber, E. U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70% - right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1514-1533.

Wisco, B. E. (2009). Depressive cognition: Self-reference and depth of processing. *Clinical Psychology Review, 29,* 382-392.

Wolford, G., Taylor, H. A., & Beck, J. R. (1990). The conjunction fallacy? *Memory and Cognition, 18,* 47-53.

Yates, J. F. (1990). *Judgment and Decision Making.* Englewood Cliffs, NJ: Prentice Hall.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.

Zakay, D. (1984). The influence of perceived event's controllability on its subjective occurrence probability. *The Psychological Record, 34,* 233-240.

Zakay, D. (1996). The relativity of unrealistic optimism. *Acta Psychologica, 93,* 121-131.

Zanna, M. P., & Rempel, J. K. (1988). Attitudes: a new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The Social Psychology of Knowledge* (pp. 315-334). Cambridge, UK: Cambridge University Press.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111-122.

Appendix

To calculate the average response of the entire population, it is necessary to sum the average responses of those people receiving a positive test result and those receiving a negative test result. These averages are obtained by multiplying the respective posterior degrees of belief (Equations 1 and 2) with the proportions of people expressing them (Equations 3 and 4). It can be seen from Equations 1 and 3 and from Equations 2 and 4 that this multiplication process will cancel out the denominators in Equations 1 and 4, leaving the average response of the population equal to Equation 5. Due to the complimentary nature of $P(\neg e \mid h)$ and $P(e \mid h)$, Equation 5 reduces to $P(h)$, which equals the base rate.

$$P(h \mid e) = \frac{P(h)P(e \mid h)}{P(h)P(e \mid h) + P(\neg h)P(e \mid \neg h)} \tag{1}$$

$$P(h \mid \neg e) = \frac{P(h)P(\neg e \mid h)}{P(\neg h)P(\neg e \mid \neg h) + P(h)P(\neg e \mid h)} \tag{2}$$

$$P(h)P(e \mid h) + P(\neg h)P(e \mid \neg h) \tag{3}$$

$$P(h)P(\neg e \mid h) + P(\neg h)P(\neg e \mid \neg h) \tag{4}$$

$$P(h)P(\neg e \mid h) + P(h)P(e \mid h) \tag{5}$$