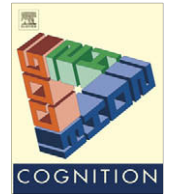




ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Estimating the probability of negative events

Adam J.L. Harris *, Adam Corner, Ulrike Hahn

Department of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom

ARTICLE INFO

Article history:

Received 17 January 2007

Revised 7 October 2008

Accepted 10 October 2008

Keywords:

Probability

Utility

Severity

Asymmetric loss functions

Control

ABSTRACT

How well we are attuned to the statistics of our environment is a fundamental question in understanding human behaviour. 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. One question that arises here is whether or not our probability estimates for negative events are systematically biased by their severity. In a minimal experimental context involving an unambiguous, objective representation of probability, we found that participants judged a controllable event as more likely to occur when its utility was extremely negative than when it was more neutral. A decision-theoretic explanation based on loss function asymmetries is advanced which supports the claim that probability estimates are not *intrinsically* biased by utilities.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Decision theory (Pratt, Raiffa, & Schlaifer, 1995) 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 decision theory 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 re-

search querying whether probabilities and utilities are in fact assessed independently.

1.1. 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. Hence, the results from studies utilising decision-making paradigms can generally be explained in terms of non-linear utility functions (e.g., Kahneman & Tversky, 1979). No interdependence between probability and utility is therefore necessary to account for these findings (see also Edwards, 1962; Slovic, 1966).

Some support for the idea that utilities 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

* Corresponding author. Tel.: +44 2920874007.

E-mail address: harrisaj@cardiff.ac.uk (A.J.L. Harris).

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., Bonnefon & Villejoubert, 2006; Grice, 2001). Context effects on the interpretation and selection of vague terms are ubiquitous. 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 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* (Fischer & Jungermann, 1996; Weber & Hilton, 1990) 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.

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 relation-

ship 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 frequently than penalised outcomes. This hypothesis would predict the 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). However, the most extensive of these, by Bar-Hillel and Budescu (1995), found no evidence for an effect of positive outcome utility on probability estimates.¹ They 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." Bar-Hillel and Budescu 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; Slovic, 1966). 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 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, other direct tests of the wishful thinking phenomenon have reported conflicting results with some finding evidence for wishful thinking (Price, 2000; Pruitt & Hoge, 1965) and others finding the opposite, a pessimism bias (Dai, Wertenbroch, & Brendl, 2008; Mandel, 2008). 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.

¹ Price (2000) reported evidence of a wishful thinking effect in the laboratory. In his study, participants were divided into two teams, and required 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 the wishful thinking hypothesis.

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.

1.2. Overview

In the following, we describe five experiments testing the proposition that the severity of negative events directly influences their perceived probability. Experiment 1 provides a demonstration that severe (extremely negative) events are assigned higher probability estimates than neutral events, a finding that is replicated twice (Experiments 1–3). Experiments 4 and 5 support an explanation for this effect in terms of loss asymmetry.

1.3. 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 realworld 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.

2. Experiment 1

The purpose of the first experiment 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 supplying probability 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; black 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.

2.1. Method

2.1.1. Participants

100 participants took part in Experiment 1. The experiment was conducted remotely using an internet host, iPsychExpts.com (Brand, 2005). 55 female and 45 male participants with a median age of 25 completed the experiment in an average time of 2.54 min. 50 participants provided probability estimates of severe outcomes, and 50 provided probability estimates of neutral outcomes.

2.1.2. Design

The hypothesis that probability estimates of severe outcomes differ from probability estimates of neutral outcomes was tested using visual response matrices containing varying proportions of black 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 estimates of three probabilities (low, medium, and high).

2.1.3. Materials

A visual display containing 2236 cells with a random distribution of black and yellow cells was constructed with a simple JAVA program designed specifically for the exper-

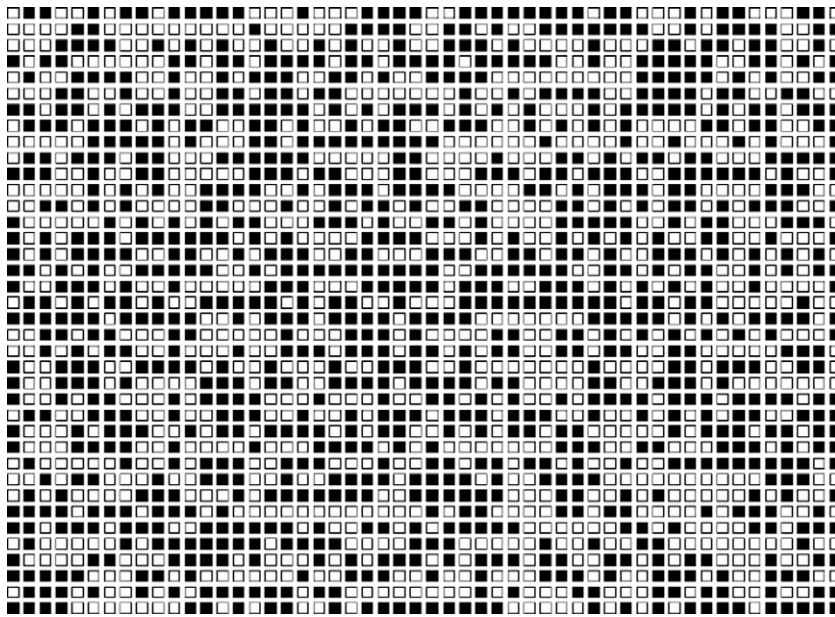


Fig. 1. A black and white example of a cell matrix (from the medium probability level).

iment (see Fig. 1). In the low probability condition of the experiment, 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.

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 BLACK 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 black 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 BLACK 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 black 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 numerical scale from 0% (Absolutely Impossible) to 100% (Absolutely Certain) with 5 point increments. Participants responded by clicking on a radio button.

2.1.4. Procedure

The experiment was run through ipsyhexpts.com. Having chosen to participate in the experiment, participants first viewed the consent screen which was followed by a screen containing the general instructions for the experiment. The next three screens contained the experimental materials. On completion participants entered their age and sex before being presented with a debriefing screen.

2.2. Results

As the experiment was conducted remotely using an internet host, we followed Birnbaum (2004) 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 elim-

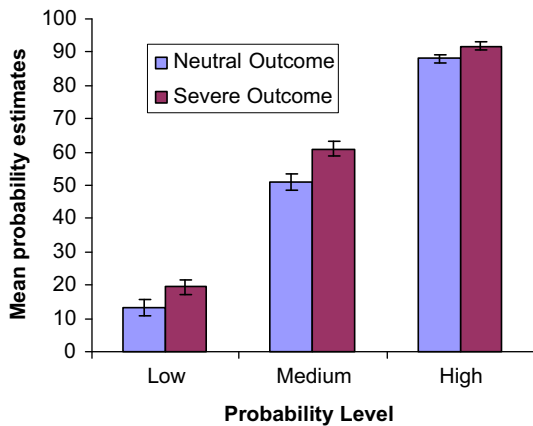


Fig. 2. The effect of outcome utility on probability judgments. Error bars are plus and minus 1 standard error.

inated 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 min were also excluded, to ensure that people were *estimating*, and not *counting* the cells. Following these exclusions, 76 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, Fig. 2 displays these probability estimates (a complete table of means for all experiments is presented in the Appendix), 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, $F(2, 142) = .75$, $p > .05$, $MSE = 87.90$.

3. Experiment 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. Experiment 2 was a direct replication of Experiment 1 with a different set of participants.

3.1. Method

3.1.1. Participants

In this study, data collection continued until there were 50 participants after exclusions (criteria as in Experiment

1) in each experimental group. 52 participants were female and 48 male, with a median age of 22.5 years. The average time to complete the experiment was 2.56 min.

3.1.2. Design, materials and procedure

Experiment 2 was an exact methodological replication of Experiment 1.

3.2. Results

The results matched those of Experiment 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 \times severity interaction was, again, not significant, $F(2, 196) = .004$, $p > .05$, $MSE = 88.75$.

4. Experiment 3

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

4.1. Method

4.1.1. Participants

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

4.1.2. Design, materials and procedure

The same mixed 3×2 design was employed as in Experiments 1 and 2.

Three blue and black matrices were constructed using the JAVA program. In this experiment, 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 Experiments 1 and 2, all participants viewed the same 3 matrices.

The same basic orchard premise was used in the cover stories. Some 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'. The procedure was identical to that in Experiments 1 and 2.

4.2. Results

Participants were excluded prior to data analysis using the same criteria as before. Following 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 Fig. 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 \times severity interaction was, once again, not significant, $F(1.9, 389.6) = 2.95$, $p > .05$, $MSE = 94.81$ (Greenhouse–Geisser corrections applied).

4.3. Discussion

The results of Experiment 3 replicate exactly the findings of Experiments 1 and 2 with changes to the colours and probabilities associated with the matrices, further suggesting that this is a robust effect, despite the minimal nature of the paradigm.

How might this effect be explained? One possibility is that it stems from sensitivity to loss function asymmetries. Loss asymmetries arise from the fact that there are two types of error 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). An asymmetric loss function account predicts that probability estimates will be biased in order to reduce the likelihood of making the more costly error. Applied to the current context this decision-theoretic account suggests that participants perceived the experimental situation as one in which the farmer could potentially prevent his daughter from entering the orchard, and as such 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

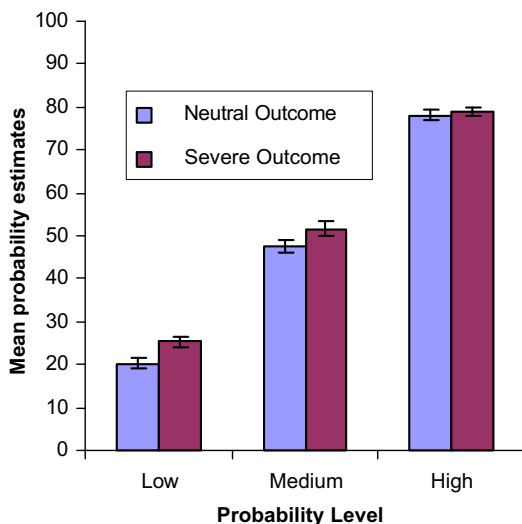


Fig. 3. Mean probability estimates made in Experiment 3. Error bars are plus and minus 1 standard error.

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.

5. Experiment 4

The purpose of this experiment was to test an asymmetric loss function based explanation for the effect observed in Experiments 1–3. Within such a decision-theoretic account, probability estimates are inflated as a preventative measure against the negative effects of a decision based on an underestimate. Hence the loss asymmetry account makes different predictions depending on whether or not there are decisions associated with the event.

One way to eliminate decisions is to consider an event where all aspects of the situation are beyond an individual’s control. Here, whether you overestimate, underestimate, or correctly estimate the probability makes no difference to the costs incurred so there can be no asymmetry in the loss function. Thus, if the asymmetric loss function explanation of our results is correct, higher probability estimates should be seen for controllable than uncontrollable events. In Experiment 4 we therefore used the severe events from Experiments 1 and 2 and introduced a manipulation of outcome control.

5.1. Method

5.1.1. Participants

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

5.1.2. Design

A 3×2 (probability \times 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.

5.1.3. Materials and procedure

This experiment used the same materials as in the severe outcome condition of Experiments 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 experiments.

5.2. 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 Fig. 4. 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 (Fig. 4).

5.3. Discussion

The results of Experiment 4 provide support for an asymmetric loss function based explanation of the biasing impact of negative utility on probability judgments. This experiment 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 ver-

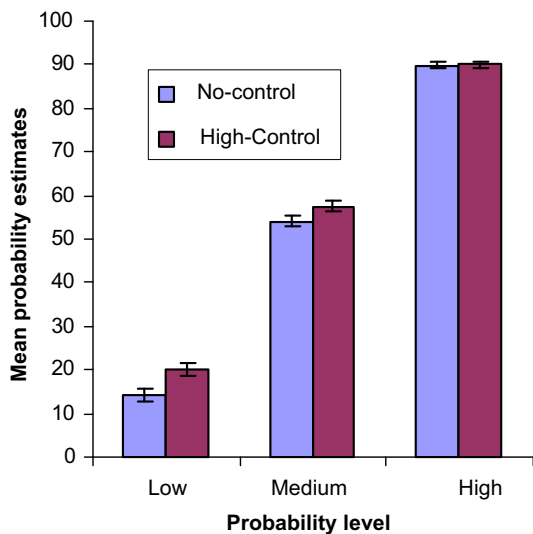


Fig. 4. Mean probability estimates in the high-control and no-control conditions of Experiment 4. Error bars are plus and minus 1 standard error.

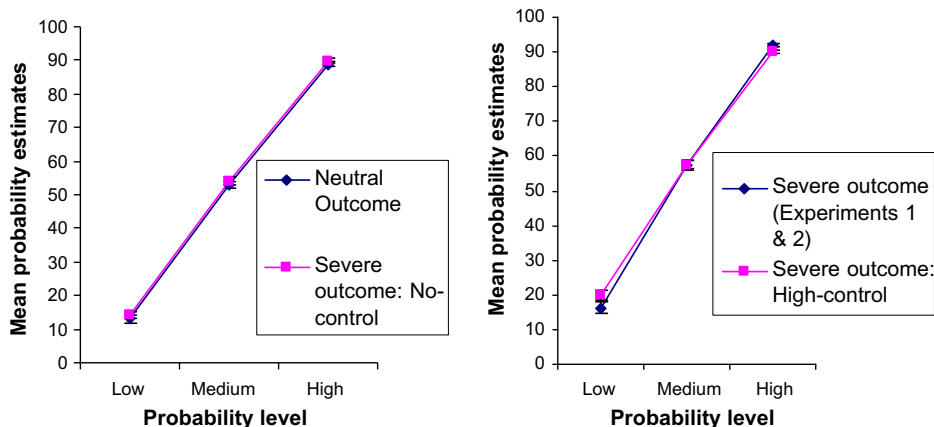


Fig. 5. Plotted are the mean probability estimates of Experiments 1 and 2 combined and the means of Experiment 4 in those conditions where no asymmetric loss function exists (left panel) and where an asymmetric loss function does exist (right panel). Error bars are plus and minus 1 standard error.

sion of the orchard cover story used in Experiments 1 to 3. Hence it should be possible to match the data of the present experiment (with its explicit controllability manipulation) to the data from Experiments 1 and 2 which used the same probability matrices (see Fig. 5).

For there to be an asymmetry in the loss function associated with the probability estimate of the event it must be both controllable and severe; thus there is no asymmetry in the no-control condition of this experiment or the neutral outcome condition of Experiments 1 and 2. A meta-analytic comparison (following Rosenthal, 1991) of the relationships depicted in Fig. 5 confirmed that the results for the respective conditions of Experiment 4 were analogous to the results of Experiments 1 and 2. As the left panel of Fig. 5 shows, the means of the *neutral* conditions of Experiments 1 and 2 and the no-control condition of Experiment 4 are virtually indistinguishable, and statistically there is no difference. Comparing the severe outcome conditions and the high-control condition of Experiment 4 (Fig. 5 right panel) there was *some* difference in that there was a significant interaction between probability and experiment, $F(1.8, 333.2) = 3.95, p < .05, MSE = 106.67$ (Greenhouse–Geisser correction applied). This is driven entirely by higher estimates at the low probability level in Experiment 4, $F(1, 428) = 4.24, p < .05, MSE = 163.53$, and would seem to reflect the greater explicitness of event-controllability in this experiment. A further meta-analytic comparison found no difference in effect sizes (r) across the experiments. The perceived controllability of the event in Experiments 1 and 2 thus seems necessary for the observed severity effect; this supports an asymmetric loss function explanation.

6. Experiment 5

The final test of the loss asymmetry explanation would be to factorially combine severity and controllability within a new scenario to demonstrate the combined importance of the two factors in biasing probability estimates. The loss asymmetry account predicts an interaction between controllability and severity. Severe events should receive higher estimates than neutral events under conditions of high control. However, under conditions of no control, no such difference should be observed. These predictions are illustrated in Fig. 6.

6.1. Method

6.1.1. Participants

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

6.1.2. Design

The design was a $3 \times 2 \times 2$ (probability \times outcome severity \times outcome controllability) mixed design in which probability was the within-subjects variable and outcome

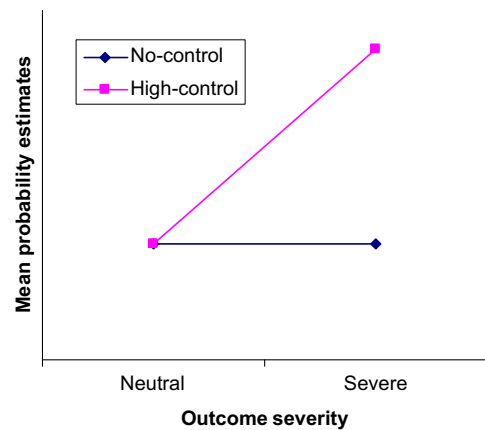


Fig. 6. The interaction predicted in Experiment 5.

severity and controllability were combined factorially between subjects. Participants were randomly assigned to one of the four conditions.

6.1.3. Materials and procedure

To test the generality of the present effect, we created three new visual displays using Microsoft's "Paint" application and Adobe Photoshop (see Fig. 7). Three probability levels were created using this display by manipulating the thickness of the blue 'river'. A new cover story was also created. The procedure was identical to that in Experiments 1–4.

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:

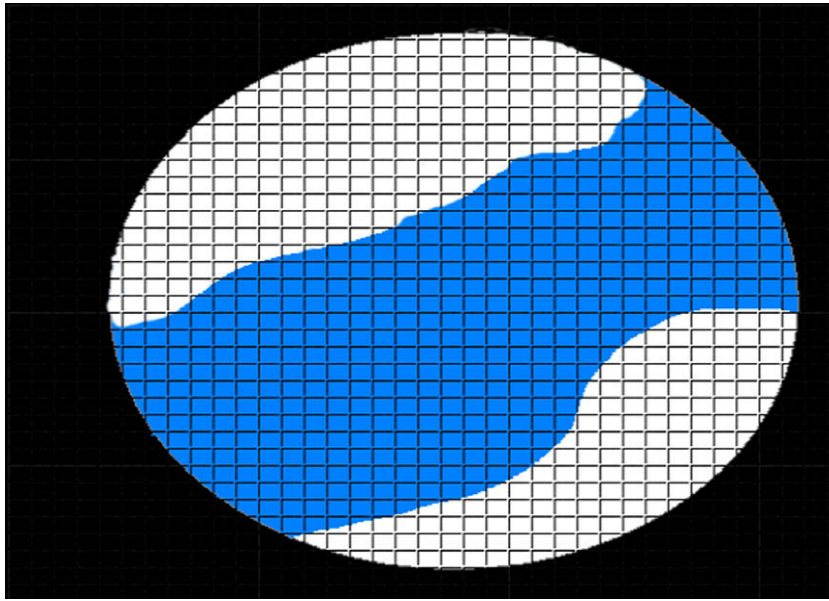


Fig. 7. An example of the visual stimulus in Experiment 5 (from the medium probability level).

'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'.

6.2. Results and discussion

Following exclusions (criteria as before), 205 participants were retained for analysis. Participants' mean probability estimates in the four experimental conditions are shown in Fig. 8. Visually, these results appear to be in line with our predictions (see Fig. 6).

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, adds

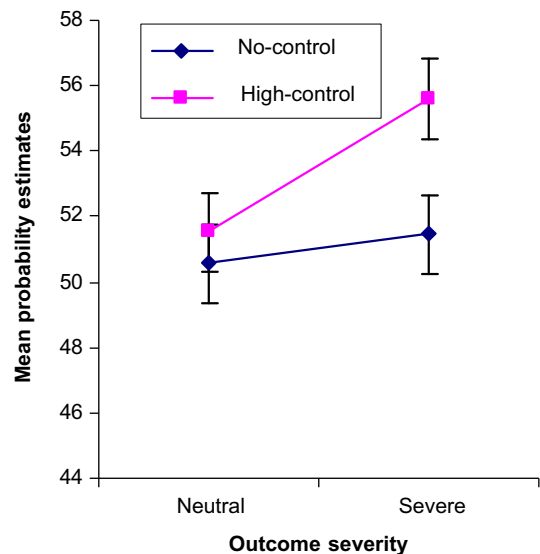


Fig. 8. Mean probability estimates in the four experimental conditions of Experiment 5. Error bars are plus and minus 1 standard error.

further support to our explanation of the utility/probability interdependence.

Experiment 5 therefore provides a clear 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 an element of controllability is crucial for the effect of severity to obtain, thus directly supporting the asymmetric loss function account.²

² Further evidence of the crucial role of controllability also came from two replications of a no-control version of this paradigm in which no effect of outcome severity was again observed.

7. General discussion

We presented five experiments investigating the effect of outcome severity on probability estimates. Our experiments used a minimal paradigm, in which an objective representation of the probability to be estimated was constantly available to participants. Experiments 1–3 showed that severe events were rated as more likely to occur than neutral events. Experiments 4 and 5 demonstrated the importance of event-controllability for the effect of outcome severity on probability estimates such that the effect was only observed for controllable events.

The results presented are explained by asymmetric loss functions. 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 preventive 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. This account is a decision-theoretic explanation and a loss asymmetry only exists if there is some element of control associated with the event. Without an element of control, a decision cannot make a difference.

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. However, 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, & 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 experiments 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, it seems, is simply the possibility 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 Fischer and Jungermann (1996) and Weber and Hilton (1990) 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.

Weber and Hilton's (1990) results were, however, 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 et al. (1986) (see Table 1 [from Wallsten et al., 1986, p. 574]), showed a reliable positive influence of severity on probability. However, these were also the only scenarios that

Table 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?...

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 1, the materials of Wallsten et al. (1986) compare 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. Realizing, 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.

7.1. Locating the effect

The next issue to address is where in the overall process of generating and reporting a probability estimate the influence of loss asymmetry occurs. Fig. 9 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 (Fig. 9) 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 all 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 Experiments 1 and 2 the objective probability is less than 5%. However, the mean estimates in the neutral outcome con-

dition lie at 13% and they become even higher, not lower, in the severe condition.

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, 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, a conclusion in line with Weber's (1994) proposal that asymmetric loss functions affect mental simulation processes that generate the estimates themselves (as in Einhorn & Hogarth, 1985; Hogarth & Einhorn, 1990). In studies where participants have been assumed to be sensitive to asymmetric loss functions, for example, participants' estimates have shown bias even once any advantage of consciously increasing a report of their internally held probabilities is removed (e.g., Bottom & Paese, 1999; Weber & Hilton, 1990).

Regardless of 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 in...' 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,

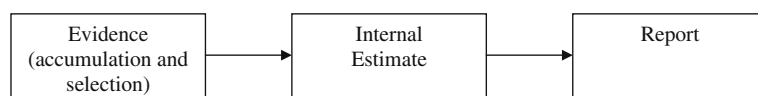
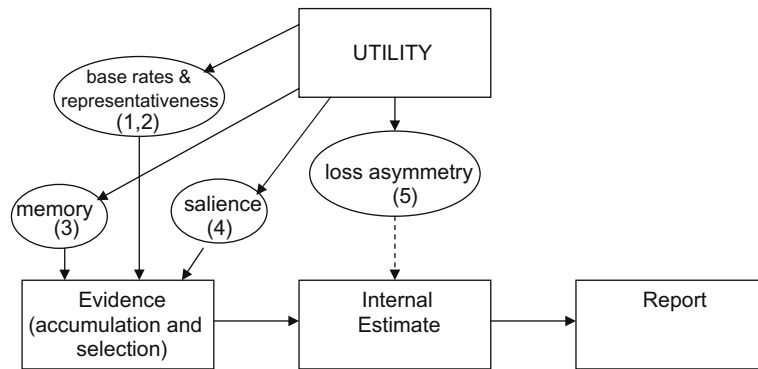


Fig. 9. The process of making and reporting a probability estimate.



Example Studies:

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

Fig. 10. Locating indirect effects of utility in the probability estimation process.

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.

Whilst the present results suggest that probability estimates are not inherently biased, they add to a considerable volume of research suggesting that they will often be biased *in practice*. At the evidence accumulation stage of the probability estimation process (see Fig. 10), 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 *saliency*: "I wish for, therefore I focus on, therefore I believe in" (Bar-Hillel et al., 2008, p. 283). Finally, Dai et al. (2008) and Mandel (2008), mentioned above, 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 current loss asymmetry-based influence of outcome severity occurs only in situations 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 even greater when the nature of our materials is considered. In our experiments, 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.

8. 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.

Acknowledgments

Adam Harris and Adam Corner were supported by the Economic and Social Research Council. A preliminary report of the results of Experiments 1 and 2 is found in the Proceedings of EuroCogSci 07. We thank Cardiff University's School of Psychology Cognitive Group for helpful discussions and the reviewers for their constructive comments.

Appendix

Mean probability estimates across all conditions of all experiments.

Probability level	Neutral outcome ^a			Severe outcome ^a		
	Low	Medium	High	Low	Medium	High
Experiment 1	13.03	51.21	88.03	17.13	58.25	91.5
Experiment 2	13	54.1	89.5	15.5	56.8	92.2
Experiment 3	20.34	47.55	78.17	25.29	51.68	78.89
Experiment 4	14.03	54.09	89.89	19.9	57.53	90
Experiment 5 (low-control)	25.19	45.67	80.87	26.79	46.23	81.32
Experiment 5 (high-control)	26.18	47.94	80.49	30	52.04	84.69

^a For Experiment 4, neutral outcome represents the no-control condition and severe outcome represents the high-control condition.

References

- Armstrong, J. S. (2001). *Principles of forecasting*. Boston: Kluwer Academic Publishing.
- Babad, E., & Katz, Y. (1991). Wishful thinking—against all odds. *Journal of Applied Social Psychology*, 21, 1921–1938.
- 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 and Review*, 15, 278–283.
- Batchelor, R., & Peel, D. (1998). Rationality testing under asymmetric loss. *Economics Letters*, 61, 49–54.
- Birnbaum, M. H. (2004). 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., & 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.
- Brand, A. (2005). iPsychExpts: Web experiments for psychologists. <<http://www.ipsyhexpts.com>>.
- Cohen, J., Dearnley, E. J., & Hansel, C. E. M. (1958). A quantitative study of meaning. *British Journal of Educational Psychology*, 28, 141–148.
- 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.
- 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., Wertebroch, K., & Brendl, C. M. (2008). The value heuristic in judgments of relative frequency. *Psychological Science*, 19, 18–20.
- 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.
- Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, 92, 433–461.
- 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.
- 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.
- Granger, C. (1969). Prediction with a generalised cost of error function. *Operational Research Quarterly*, 20, 199–207.
- Grice, H. P. (2001). *Aspects of reason*. Oxford: Clarendon Press.
- Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science*, 36, 780–803.
- Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.
- Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality*, 21, 329–335.
- 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.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, 133, 95–121.
- Landy, M. S., Goutcher, R., Trommershäuser, J., & Mamassian, P. (2007). Visual estimation under risk. *Journal of Vision*, 7(6):4, 1–15.
- 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.
- 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.
- Merz, J. F., Druzdzel, M. J., & Mazur, D. J. (1991). Verbal expressions of probability in informed consent litigation. *Medical Decision Making*, 11, 273–281.
- 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.
- Pratt, J., Raiffa, H., & Schlaifer, R. (1995). *Introduction to statistical decision theory*. Massachusetts: MIT Press.
- Price, P. C. (2000). Wishful thinking in the prediction of competitive outcomes. *Thinking and Reasoning*, 6, 161–172.
- 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.
- 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.
- Sherif, M., & Sherif, C. W. (1956). *An Outline of Social Psychology* (revised ed.). New York: Harper & Brothers.
- 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.
- Solow, A. R., & Broadus, J. M. (1988). A simple model of over-forecasting. *Monthly Weather Review*, *116*, 1371–1373.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.
- 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.
- 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., 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.
- 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.
- Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision*, *8*(3):2, 1–15.
- 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.