Within-subject Preference Reversals in Description- and Experience-based Choice

Adrian R. Camilleri (acamilleri@psy.unsw.edu.au)
School of Psychology, University of New South Wales, Sydney, Australia

Ben R. Newell (ben.newell@unsw.edu.au)
School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Numerous studies using between-subject designs have found that different decisions can be made about identical binary choice problems depending on whether the options are described or experienced. Using a within-subjects design we examined this Description-Experience 'gap' at the level of the individual. We found that: (1) the gap could be observed both at the group and the individual levels, (2) the gap was eliminated, at least at the group level, when controlling for sampling variability, and (3) riskier decisions were made by those with more positive risk attitudes, regardless of format. We conclude that the gap is likely a statistical phenomenon due to biased samples.

Keywords: decisions from experience, decisions from description, description-experience, risk taking, risky choice.

Introduction

Murray and Kluckhohn's (1953) clever adage that "every man is in certain respects (a) like all other men, (b) like some other men, (c) like no other man" highlights three levels of investigation. The first level refers to universal cognitive or biological mechanisms, the second level to social grouping factors, and the third level to individual differences. Most studies restrict their investigation to a single level and this can become problematic if the research in an area becomes concentrated on just this one level (Lopes, 1987). A current example of where this phenomenon may be occurring is in the context of the "Description-Experience (D-E) gap" debate. The controversy lies with the observation that different decisions are made about structurally identical lotteries as a function of how information about the options is acquired. To date, all of the published studies have used between-subjects designs. This makes sense at the first level of investigation where the intent is to abstract and model universal choice mechanisms. However, several of the most interesting conclusions implied by the "gap" are at the third, and as yet largely unaddressed, level of investigation. In the present study we re-examine some of the conclusions that have been made, and add additional insights, by examining the D-E gap within-subjects while assessing individual differences in risk attitude.

Universal Choice Mechanisms

Over the last few decades the prevailing methodology used to investigate universal choice mechanisms is the decision from description paradigm (Weber, Shafir, & Blais, 2004, but see Myers & Suydam, 1964). In a decision from description (DfD) paradigm decision makers are presented with convenient descriptions of all outcomes and their respective likelihoods, and are asked to select their preferred option. For example, the decision maker might be asked to choose between: (A) a 100% chance of 3, or (B) an 80% chance of 4, else 0 (henceforth, Problem 1). Contrary to early belief (e.g., Expected Utility Theory, Savage, 1954) people often make decisions that depart from the prescriptions of rational choice axioms. For example, Kahneman and Tversky (1979) noted that in Problem 1, 80% of decision makers tended to prefer the certain option (choice A) despite it being associated with a lower expected value. In addition, people appear to make decisions as if options with very low probabilities are *overweighted* but options with moderate and high probabilities are *underweighted*. Based on the large body of data gathered from the DfD paradigm, choice behaviour appears to adhere to the 'fourfold-pattern': risk averse for gains and risk seeking for losses of high probability, but risk seeking for gains and risk averse for losses of low probability. The most successful model accounting for these patterns of choice is Prospect Theory (PT; Tversky & Kahneman, 1992).

In spite of PT's huge success and beliefs about its wide scope of generalization, recent evidence has questioned the applicability of PT, and the very occurrence of the four-fold pattern, to more ecological decisions from experience (Barron & Erev, 2003). In a decision from experience (DfE) paradigm decision makers are initially unaware of their options and must learn about potential outcomes and make estimates of their respective likelihoods through exploration and feedback. In the sampling version of the paradigm, decision makers might be presented with two options that they are asked to sample from. Each sample returns a value randomly selected from a static payoff distribution corresponding to an objective probability that is unknown to the decision-maker. For example, one option might be associated with the distribution "100% chance of 3" and the other option "80% chance of 4, else 0" (i.e., Problem 1). In this exploration stage decision makers are free to sample from each option in any order and as often as they like without consequence. Once the decision-maker has gathered enough information about their options and have formed a preference they move on to the exploitation stage where they select one option to play from for real.

Using the DfE paradigm, Hertwig et al. (2004) observed choices that were actually opposite to the predictions of PT. Indeed, strikingly different patterns of choice were observed when compared to decisions made by participants in another group presented with the same problems via the DfD paradigm. Averaged across all problems, the absolute difference in percentage points was 36. This D-E ‘gap’ has now been replicated on many occasions with a range of problem sets (Camilleri & Newell, submitted; Hau et al., in press; Hau et al., 2008; Hertwig et al., 2004; Rakow et al., 2008; Ungemach et al., 2009; Weber et al., 2004).

Within-subject Designs

Notably, all of these studies have used a between-subjects design and, appropriate to this level of investigation, a number of models have been devised to explain the universal choice mechanisms involved (for an overview, see Hau et al., 2008). However, some of the conclusions implied by the “D-E gap” may be on less solid ground. The issue boils down to what exactly we mean by, and want to infer from, the word “gap” in the context of risky choice. One inference is that, “*given the same structural decision problem, the average group of people will show a different choice preference compared to another average group of people as a function of how the two groups learn about their options*”. A second inference is that, “*given the same structural decision problem, the average person will reverse their choice preference as a function of how that person learns about their options*”. We feel that the second conclusion is at least as, if not more, interesting than the first conclusion; however, based on the current literature the second conclusion cannot be made. Thus, our first aim in the current study was to observe a D-E gap within-subjects.

The Importance of Sampling Variability

With a number of studies ruling out factors such as recency (Hau et al., in press; Ungemach et al., 2009) and judgement error (Fox & Hadar, 2006; Hau et al., 2008), the debate as to the cause of the gap is now largely focused on whether it can be explained as a statistical occurrence due to biased sampling. The freedom inherent in the sampling DfE paradigm means that decision makers often make their choices based on relatively small, and therefore biased, samples that do not accurately reflect the objective probabilities (Hertwig et al., 2004). As a consequence, there are fewer encounters with the rare event than expected. Fox and Hadar calculated that in the Hertwig et al. data, 69% of choices in the Experience group (and 63% of choices in the Description group) were predicted by cumulative PT when based on participant’s actual (biased) samples, as opposed to 27% when based on objective probabilities.

At present, the relative importance of sampling bias as a cause of the gap remains unsettled. On the one hand, the gap has been observed in studies that have increased the number of samples by manipulating incentives (Hau et al., 2008), or when forcing decision maker to sample extensively (Hau et al., 2008) or in exact accordance with the objective probabilities (Ungemach et al., 2009). On the other hand, the gap has been eliminated in studies that have yoked description-based problems to the subjective distributions observed in experience-based choices (Rakow et al., 2008), and with a binning procedure that compared description-based decisions only with experience-based decisions where the experienced distribution was roughly equal to the objective distribution (Camilleri & Newell, submitted). The methodology preferred in the current study is the binning procedure that allows participants to freely sample. Thus, our second aim was to test the statistical account by observing whether the D-E gap is eliminated when controlling for sampling bias by using the binning procedure, in the context of a within-subjects study.

Individual Differences in Risk Attitude

In the context of decisions between safe and risky options, perhaps the most fascinating individual difference is that of risk attitude. Risk attitude is broadly understood as the degree to which an individual engages in risky behaviors (Weber, Blais, & Betz, 2002). In the context of PT, risk attitude refers to the degree of concavity (or convexity) of a decision maker’s utility function. Two problems with this definition are: (1) different measures of risk attitude can classify an individual disparately, and (2) even the same measure can classify an individual disparately in different domains. Indeed, a growing body of evidence suggests that “risk attitude” is domain specific (Weber et al., 2002).

In light of these issues, some researchers have suggested that risk attitude could be better understood within a risk-return framework (Weber et al., 2002). Under this conceptualization, observed behavior is a function of two factors: (a) an evaluation of the benefits and risks, and (b) an attitude towards perceived risk (i.e., the trade-off between perceived benefits and perceived risks). Thus, an individual may be classified disparately in different domains due to inequality in either factor, but not necessarily both. A useful measure for examining each of these factors is the Domain-Specific Risk-Taking (DOSPERT) scale (Blais & Weber, 2006). The DOSPERT is a self-report questionnaire that assesses the tendency to engage in risky activities across five domains, as well as the perceived risks and benefits involved in those activities. Behavioral scores on the DOSPERT have been shown to have a significant positive relationship with risky decisions made in a DfD paradigm (Weber et al., 2002). Additional support for a link between a stable dispositional risk trait and behavior comes from the observation that choices made within a decision-under-risk paradigm (where the outcome distribution is known) predict choices made within a decision-under-ambiguity paradigm (where the outcome distribution is unknown), even after a 2 month time gap (Lauriola, Levin & Hart, 2007). Together, these findings suggest that risk attitude may be useful in predicting experience-based choice, or could interact in some way with an individual’s tendency to make choices in line with a D-E gap. Thus, our third aim was to replicate the correlation between description-based choices and individual differences in risk attitude, to determine whether this association holds with experience-based choices, and investigate the possibility that the size of the D-E gap is modulated by risk attitude.

Method

Participants

The participants were 40 undergraduate first year University of New South Wales psychology students (23 females), with an average age of 19.3 years. Participation was in exchange for course credit, plus payment contingent upon choices (range = AUD$0.00 to AUD$3.10).

Materials

Choice Problems: The ten choice problems used are shown in Table 1. Each choice problem consisted of two options with similar expected values, with at most two outcomes per option. The option predicted by PT to be preferred was labelled the "favoured" choice and the alternative option was labelled the "non-favoured" choice. Specifically, the favoured choice was the option containing the rare event when the rare event was desirable (e.g., 0 is a desirable rare event in the problem -4[.8] 0[.2]), or the alternative option when the rare event was undesirable (e.g., 0 is an undesirable rare event in the problem 4[.8] 0[.2]).

Risk Attitude Measure: The Domain Specific Risk-Taking Scale (DOSPERT; Weber et al., 2002) was used to measure individual attitudes towards risk. DOSPERT assesses an individual's risk taking in a number of scenarios within the domains of financial, health/safety, recreational, ethical, and social decisions. For each scenario respondents rate along a 7-point scale: (1) the likelihood that they would engage in the activity, (2) the perceived magnitude of the risks associated with engaging in the activity, and (3) the expected benefits from engaging in the activity. DOSPERT has been found to have adequate internal-consistency and test-retest reliability estimates, as well as good convergent/ discriminant and construct validity (Blais & Weber, 2006).

In order to minimize participant fatigue across the experiment, only the 10 scenarios from the domains of finance and recreation were used. Weber et al. (2002, p.282) state that "if risk attitudes are measured merely for predictive purposes [then] one or more of the subscales ... will suffice". An example of a scenario from the domain of finance is "*Betting a day's income at the horse races*". An example of a scenario from the domain of recreation is "*Bungee jumping off a tall bridge*".

Filler Task: A 2-minute, computerized filler task asked participants to list as many countries as they could from three different geographical regions.

Procedure and Design

The within-subjects experiment comprised four tasks that were described to the participants as independent: (A) a description-based choice task, (B) a filler task, (C) an experience-based choice task, and (D) the DOSPERT. Half of the participants completed the tasks in the order A,B,C,D and the other half in the order C,B,A,D.

At the beginning of each of the two choice tasks (i.e., tasks A and C) the instructions indicated that a number of different money machines would have to be selected between, each of which could add or subtract points from their running tally. Participants' overall task was to maximize the amount of points won. At the end of the experiment points were converted into real money according to the conversion rate of 1 point = AUD$0.10.

In the Description condition participants were instructed to compare two *labeled* money machines and to choose one to play from. In the Experience condition participants were instructed to sample from two *unlabeled* money machines in order to find out what the machine's payoff was like. Samples from each machine reflected random draws from a distribution of possible outcomes in accordance with the objective probabilities. Participants were allowed to sample each of the machines as often and in any order that they liked until they decided to choose one machine to play from. Participants were not given feedback during the experiment in order to reduce any wealth effects. In all cases allocation of safe and risky options to the left and right machines was counterbalanced and the order of the problems was random.

After both choice conditions and the filler task had been completed, the DOSPERT measure of risk attitudes was administered. Each of the three questions assessed by the DOSPERT was presented on a separate screen and in random order. The order of the scenarios on each question screen was also randomized for each participant.

At the completion of the experiment a final screen appeared informing the participant that the experiment was finished, and revealed their total points earned, as well as their corresponding real money conversion. Participants that ended up with negative point scores were treated as though they had scored zero points. Finally, participants were thanked, debriefed, and then paid.

Results

Sampling Behaviour

We computed each participant's average number of observations per problem (total sample size), average number of periods of uninterrupted observation from a single machine per problem (number of sub-samples), and the average number of observations in each of these subsamples (sub-sample size). The mean (median) values were 12.1 (9.0) for total sample size, 4.9 (3.0) for the number of sub-samples, and 3.3 (1.6) for sub-sample size.

Task order was a factor in sampling strategy adopted. Sampling was more extensive when the experience-based choice task was played *first*; the mean (median) values were 13.6 (10.0) for total sample size, 5.1 (3.0) for the number of sub-samples, and 3.5 (1.6) for sub-sample size. When the experience-based choice task was played *second* the values were 10.6 (8.0) for total sample size, 4.7 (2.0) for the number of sub-samples, and 3.0 (1.5) for sub-sample size.

Patterns of Choice

Table 1 displays the percentage of participants selecting the favoured choice in each condition. It was expected that more participants would select the favoured choice in the Description condition than in the Experience condition. When averaging across task order, the difference between Description and Experience conditions falls in the expected direction for all ten problems. Seven of these differences were significant by individual chi-square tests (p's < .05). Indeed, averaging across problems, the favoured choice was selected on 53.3% of trials in the Description condition and on 31.3% of trials in the Experience condition: a difference of 22 percentage points. The odds of selecting the favoured option in the Description condition were therefore more than 2.5 times the odds of selecting the favoured option in the Experience condition. Task order again played a role: the mean difference in the predicted direction was 16 percentage points when tasks were played *first* and 28 percentage points when tasks were played *second*.

Table 1: Percent choosing the favoured option by condition

Problem	Option		% choosing favoured option	
	Favoured	Non-favoured	Descrip.	Exper.
1	3 (1)	4 (.8)	70	30*
2	-4 (.8)	-3 (1)	53	40
3	32 (.1)	3 (1)	53	30*
4	-3 (1)	-32 (.1)	48	23*
5	9 (1)	10 (.9)	53	30*
6	-10 (.9)	-9 (1)	65	35*
7	16 (.2)	3 (1)	40	33
8	11 (.1)	1 (1)	63	38*
9	14 (.15)	2 (1)	60	28*
10	28 (.15)	4 (1)	30	28

* Significantly different from Description condition.

Choice Preference Reversals

The average percentage of problems in which participants switched their choice between Description and Experience conditions was 48.2%. Where a change in preference did occur, 72.5% of these switches were in the predicted direction, that is, from the favourable choice in the Description condition to the non-favourable choice in the Experience condition.

Taking advantage of our within-subjects design we looked at, for each individual and problem, the degree of correspondence between description- and experience-based choices to determine if there had been: (1) a preference reversal in the predicted direction, (2) a preference reversal in the non-prediction direction, or (3) no preference change. As can be seen in Figure 1, the vast majority of reversals, if they occurred, were in the predicted direction. Indeed, when we calculated a Description-Experience gap score (proportion of choice preference reversals in the *predicted* direction minus the proportion of choice preference reversals in the *non-predicted* direction) we found that thirty-two participants showed a D-E gap in the predicted direction, five participants showed no gap, and just three participants showed a gap in the non-predicted direction.

Subjective Experiences of the Rare Event

To investigate the important role of sampling variation, we divided participants' experience-based choice problems into seven 'bins' based on their subjective experience of the rare event. In order to maintain standardization between problems with rare events of differing rarity, data was binned as a function of the objective probability. Specifically, each bin constituted a certain percentage of objective probability away from objective probability. For example, the data collected in the central bin, Bin 4, represent those from participants whose subjective experience was ±10% of the objective probability away from the objective probability. Thus, when the objective probability was 10%, subjective experiences of the rare event between 9 and 11% were placed into Bin 4 (i.e., 10% of 10% = 1%; 10%±1% = 9 to 11%); when the objective probability was 15%, subjective experiences of the rare event between 13.5 and 16.5% were placed into Bin 4; and when the objective probability was 20%, subjective experiences of the rare event between 18 and 22% were placed into Bin 4. This binning procedure placed just over 16% of all trials into the central three bins.

The proportion of trials where participants selected the favoured option in each of the seven bins is displayed in

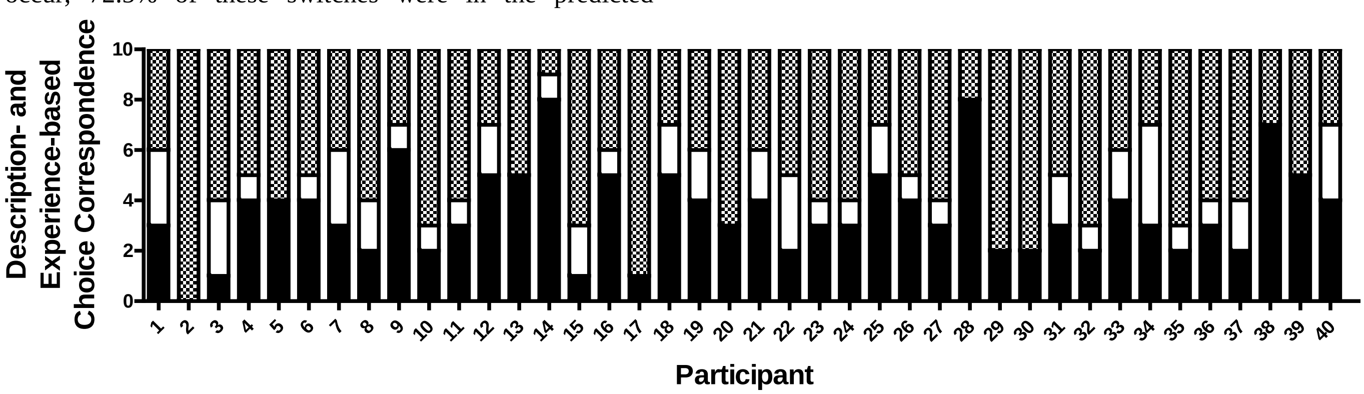

Figure 1. Degree of correspondence between description- and experience-based choice for each participant. Black bars indicate preference reversal in the predicted direction, white bars indicate preference reversal in the non-predicted direction, and dotted bars indicate that no preference reversal occurred.

Figure 2. The curve is positive and linear, indicating that participants were more likely to select the favoured option the more often they experienced the rare event. The most critical trials to consider are those located in Bin 4, since it is only on these trials that the experienced distribution was approximately (i.e., ±10%) in accordance with the objective probability. The proportion of trials in which participants selected the favoured option is remarkably similar across the Description and Experience Bin 4 data: .53 and .52 respectively ($p > 1$, one-tailed). Our power to detect a difference of the size generally reported in the literature (i.e., odds-ratio of greater than 2.5) was at least 62%.

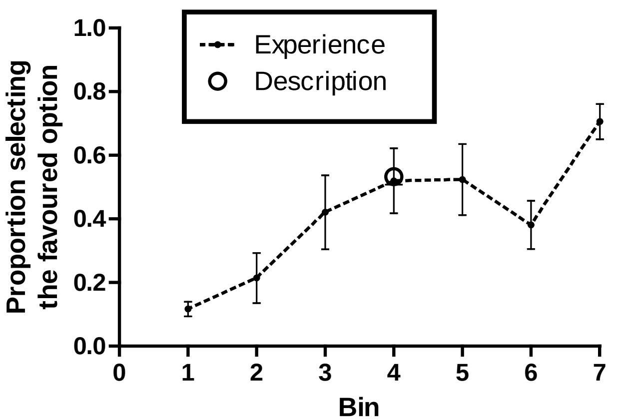

Figure 2: Percentage of participants selecting the favoured option as a function of bin.

Individual Differences in Risk Attitude

In order to examine the relationship between risk attitude and risky choices we designated the option with the greater variance the "risky" option and the alternative the "safe" option. Each participant's average score from the three DOSPERT questions were correlated with the participant's own average number of risky choices made in the Description and Experience conditions. As expected, participants' average number of risky choices made in the Description condition was significantly positively correlated with their average behavioural score ($r = .40$, $p < .05$) and negatively correlated with their averaged perceived risk score ($r = -.37$, $p < .05$). There were no significant correlations between participants' average number of risky choices made in the Experience condition and scores on the DOSPERT. Additionally, we could find no relationship between DOSPERT scores and choice preference reversals, sampling strategy adopted, or propensity to make decisions in accordance with EV.

Discussion

When making description-based choices, the majority of our participants made decisions in line with the predictions of PT and selected the favoured option (Tversky & Kahneman, 1992). In contrast, when making experience-based choices, the majority of our participants made decisions contrary to the predictions of PT and selected the non-favoured option. This apparent 'gap' between choice preferences as a function of mode of information acquisition replicates findings from numerous previous studies[1] (Camilleri & Newell, submitted; Hau et al., in press; Hau et al., 2008; Hertwig et al., 2004; Rakow et al., 2008; Ungemach et al., 2009; Weber et al., 2004). What makes our results particularly fascinating is that we observed these preference reversals at the level of the participant, that is, individual D-E gaps. Thus, we can make the conclusion that given the same structural decision problem, the average person will reverse their choice preference as a function of how that person learns about their options (i.e., described or experienced). More broadly, these results support the general hypothesis that individuals appear to make choices as if outcomes with very low probabilities have more of an impact on decisions when they are described than when they are experienced, a finding that has a wide range of real-world implications (e.g., Barron, Leider, & Stack, 2008).

The results also highlight the importance of sampling variability. When conditionalising only on those trials where participants' experienced distribution was approximately equal to the objective probability the gap disappeared: the proportion selecting the favoured choice was equivalent regardless of whether the choice was description- or experience-based. This results replicates, and extends to a within-subjects design, previous studies that have attempted to account for sampling variability using the free sampling DFE paradigm (Camilleri & Newell, submitted; Rakow et al., 2008). Equivalent choices given equivalent information supports the statistical account of the D-E gap, which suggests that the phenomenon is almost entirely due to sampling bias that occurs at the point of information acquisition. Such a proposition leads to the strong prediction that there will be no individual D-E gaps – in fact, no preference reversals at all – when comparing DfD only with central binned DfE. Unfortunately, the paucity of trials per individual that actually fell into Bin 4 severely limited our ability to conduct any meaningful inferential test of this prediction. Therefore, based on the present dataset, we cannot extend the "equivalent information equals equivalent choice" qualification to the level of the individual.

Using the DOSPERT as a measure of risk attitude, we replicated the correlation between risk attitude and description-based choices (Weber et al., 2004). Unsurprisingly, we found no evidence that scores on the DOSPERT could be used to predict experience-based choices when averaging across problem (and therefore, a myriad of experienced distributions). Such a null-finding was unsurprising given that, for the majority of problems, there was in fact no risky choice to be made: biased samples had reduced the lotteries to trivial decisions such as "100% chance of 3" versus "100% chance of 4". Only sampled

[1] The larger gap observed when the experienced-base choice task was played second is most likely due to fatigue and participants correspondingly smaller (and therefore more biased) samples.

distributions close to the objective probabilities permitted a risky choice to be made. Again, due to the small number of trials that actually fell into the central bin, we were unable to rigorously test the prediction that DOSPERT scores would predict choices made when information sampled was nearly equal to the objective probability. However, when we used a more liberal criterion and looked at the 27 participants that had at least one experienced distribution trial that fell into the central *three* bins, we found that the participants' average number of risky choices was non-significantly positively correlated with their averaged behavioural score (r = .30, p = .12) and negatively correlated with their averaged perceived risk score (r = -.36, p = .06). From these correlations we can tentatively conclude that a greater proportion of risky choices were made by those who self-reported to be more likely to perform risky behaviours, and perceived such behaviours as less risky, regardless of the mode of information acquisition.

The major limitation of the current study was the small number of trials in the experience-based condition that actually fell into the central bin. As a result, we were restricted in our ability to examine the relationship between risk attitude and choice preference reversals when information acquired was approximately equal. A methodology for overcoming this problem is to manipulate the sequence of samples that participants are exposed to in order to drive the experienced distribution towards the objective probabilities while maintaining a pseudo-random sample selection (for one such attempt see Camilleri and Newell, submitted).

In summary, we found that the Description-Experience gap phenomenon can be observed both at the individual and group levels. The gap can be eliminated, at least in the latter case, when only considering choices in which sampled observations result in experienced distributions close to the objective probabilities. Such a pattern of results strongly conforms to the predictions of a statistical account due to biased sampling. We also found that one measure of risk attitude, the DOSPERT, may be useful in predicting choices, but only when the options are presented in a description-type format or in an experienced-based format where sampling variability does not radically skew the perceived outcome distributions.

Acknowledgments

This research was supported by an Australian Postgraduate Scholarship and a UNSW Research Excellence Award to the first author, and an Australian Research Council Discovery Project Grant (DP 0770292) to the second author.

References

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16,* 215-233.

Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll. *Organizational Behavior and Human Decision Processes, 106,* 125-142.

Blais, A. R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1,* 33-47.

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.

Hau, R., Pleskac, T. J., & Hertwig, R. (in press). Decisions from experience and statistical probabilities: Do they trigger different choices than a priori probabilities? *Journal of Behavioral Decision Making*.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*, 1-26.

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,* 534-539.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-292.

Lauriola, M., Levin, I. P., & Hart, S. S. (2007). Common and distinct factors in decision making under ambiguity and risk: A psychometric study of individual differences. *Organizational Behavior and Human Decision Processes, 104*, 130-149.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 20, pp. 255-295). San Diego, CA, US: Academic Press.

Myers, J. L., & Suydam, M. M. (1964). Gain, cost, and event probability as determiners of choice behavior. *Psychonomic Science, 1,* 39-40.

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 106*, 168-179.

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.

Savage, L. J. (1954). *Foundations of Statistics*: Wiley: Oxford.

Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15,* 263-290.

Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111*, 430-445.