The Psychological Mechanisms Underpinning Experience-based Choice

A dissertation by
Adrian Ryan Camilleri

Submitted in Fulfillment of the Requirements for the Degree of
Doctor of Philosophy / Masters of Psychology (Organisational)

September 2011

School of Psychology
The University of New South Wales
Sydney, Australia

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed……

Date……...

The key to success is perseverance – and good luck.

Acknowledgements

This thesis is a culmination of a perfect working relationship with my supervisor, Ben Newell, to whom I am eternally grateful. Ben provided unreserved support during my PhD and generously paved the way for my development as a research scientist. Perhaps most importantly, I thank Ben for being my companion on our quest to discover what lies in the shadow of the statue.

I am also greatly indebted to the many people who in some way contributed to the progress and publication of the work contained herein. First and foremost, I thank my co-authors, Guy Hawkins and Scott Brown. I also appreciate the help provided by Fred Westbrook, Brett Hayes, Kevin Bird, Jonathan Baron, Ralph Hertwig, Craig Fox, Timothy Rakow, Robin Hau, Daniel Gottlieb, Timothy Schofield, Amos Schurr, Michael Hill, Chris Tynan, Frank Yates, Ryan Jessup, Claudia Gonzalez-Vallejo, Joanne Earl, Erik Altmann, Ulrike Hahn, the members of the UNSW cognition lab, and numerous other anonymous reviewers and online technical experts.

Writing this thesis was not the lonely experience it could have been because of cherished friends who provided enthusiasm and empathy in just the right doses. The wonderful companionship of Zayra Millan, Anna McCarrey, Genevra Hart, Emma Campbell-Smith, Melissa Onden Lim, Gloria Lau, Tony Wang, Nairita Bhattacharya, Xerox Tang, Danielle Mathersul, Emma Fabiansson, and many others ensures that I can only think back upon the last few years with feelings of fondness and reminisce .

The unconditional love and encouragement provided by my family served as a secure anchor during the hard and easy times; thank you.

Lastly, to my *Constant*, Joy Francisco, for always being there.

Included Papers and Contribution of the Candidate

Paper 1 Camilleri, A. R., & Newell, B. R. (2009). Within-subject preference reversals in description- and experience-based choice. In N. Taatgen, J. v. Rijn, J. Nerbonne & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 449-454). Austin, TX.

80% contribution.

The candidate led the experimental design, programming, experimental testing, data analysis, and manuscript preparation.

Paper 2 Camilleri, A. R., & Newell, B. R. (2011). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276–284.

80% contribution.

The candidate led the experimental design, programming, experimental testing, data analysis, and manuscript preparation.

Paper 3 Camilleri, A. R., & Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*, 377–384.

75% contribution.

The candidate led the experimental design, programming, experimental testing, data analysis, manuscript preparation, and conducted part of the computational modelling.

Paper 4 Camilleri, A. R., & Newell, B. R. (submitted). The long and short of it: Closing the description-experience "gap" by taking the long run view.

80% contribution.

The candidate led the experimental design, programming, experimental testing, data analysis, and manuscript preparation. The manuscript was submitted to *Psychological Science* in September 2011.

Paper 5 Camilleri, A. R., & Newell, B. R. (2009). The role of representation in experience-based choice. *Judgment and Decision Making, 4*, 518-529.

75% contribution.

The candidate led the experimental design, programming, experimental testing, computational modelling, manuscript preparation, and conducted part of the data analysis.

Paper 6 Camilleri, A. R., & Newell, B. R. (2011). The relevance of a probabilistic mindset in risky choice. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2794-2799). Austin, TX: Cognitive Science Society.

80% contribution.

The candidate led the experimental design, programming, experimental testing, data analysis, and manuscript preparation.

Paper 7 Hawkins, G., Camilleri, A. R., Newell, B. R., & Brown, S. D. (submitted). Modeling probability estimates and choice in decisions from experience.

30% contribution.

The candidate contributed to the experimental design, experimental testing, data analysis, and manuscript preparation. The manuscript was submitted to *Psychological Review* in September 2011.

Abstract

Most decisions occur in the context of uncertainty. Usually we do not possess explicit knowledge of all the outcomes and their associated probabilities; instead, we must estimate this outcome distribution information from our own personal experience with similar past situations. The primary goal motivating the work contained within was to reveal the psychological mechanisms underlying such experience-based choices. The phenomenon inspiring this goal was the observation that preferences tend to reverse depending on whether information about alternative outcome distributions is learnt from a summary description or from the experience of sequentially sampling outcomes. In the first experimental chapter it is argued that much of this description-experience "gap" can be attributed to non-representative samples serving as the basis of experience-based choice. Such non-representative samples can occur externally – because of frugal sampling efforts – and internally – because of limited cognitive resources. Both of these sources of bias have the effect of under-representing rare events. However, as discussed in the second experimental chapter, these explanations are sufficient only when costless sampling is followed by a single choice. In contrast, the gap remains in situations where each of many samples is a repeated, consequential choice. It is argued that the sequential nature of these repeated choices induces a short horizon and heavy reliance on recent outcomes. The final experimental chapter demonstrates that decision-makers appear to integrate their experience in such a way as to overestimate rare events and under estimate common events. It is argued that this judgment error reflects the processes of a noisy, instance-based memory system. The system is mechanised in a new and

successful model of experience-based choice: the exemplar-confusion model. It is concluded that description- and experience-based choice formats lie along a continuum of uncertainty and share important core features, including the explicit representation of probability, the combining of this probability information with outcome information, and utility maximization. The implication of this conclusion is that the differences between description- and experience-based choices emerge from how uncertainty information is acquired, rather than how it is represented or used.

Table of Contents

Chapter 1: Introduction

Nearly every decision we make occurs in the context of uncertainty. Where to live? Whom to marry? Arsenal or Manchester United? One question that arises when thinking about this subject is whether it matters how uncertainty information is acquired.

Consider the decision of how close to live to a nuclear reactor. During the first decade of this century numerous Gallup polls showed that support for nuclear energy by Americans adults was nearly evenly split (Jones, 2010). Although there are several obvious variables mediating this preference, including gender, income, and political party affiliation, one less obvious variable that may be just as important is the information format that people tend to rely on: experiences or descriptions.

Recent estimates indicate that nearly 1 in 3 Americans live within 50 miles of the 104 nuclear reactors powering 20% of the United States (Dedman, 2011; US Energy Information Administration, 2009). These individuals have the opportunity to assess the likelihood of a nuclear incident based on their own personal experience of living close to a nuclear reactor. In contrast, individuals without experience living close to a nuclear reactor must rely on summary descriptions presented by others to assess of the likelihood of a nuclear incident. For example, an interdisciplinary team from MIT have estimated that given the expected growth scenario for nuclear power from 2005 – 2055, at least four serious nuclear accidents will occur in that period (Beckjord et al., 2003). Interestingly, of those Americans living in close proximity to an operating nuclear power plant, 82% are in favour of nuclear energy (Nuclear Energy Institute, 2007). This rate of nuclear energy endorsement is much greater

than the national average of 57% and suggests that the availability and reliance on personal experience to assess risk may produce preferences that diverge from those that rely on summary descriptions provided by others. The possibility that the experience-based choices and their underlying processes are unique is of particular consequence given that in the last three decades the focus of most risky choice research and theoretical development has occurred with description-based choices (Weber, Shafir, & Blais, 2004).

Thus, the aim in the present research was to reveal the processes underlying experience-based choices. The investigation was limited to the study of binary choice monetary gambles. This approach was adopted to permit comparison with the extensive description-based choice literature, which has demonstrated the diagnostic viability of binary choice monetary gambles.

The thesis is presented in the form of seven papers bounded between a literature review chapter and conclusion chapter. Note that because the thesis is a collection of published papers, a reference section concludes each chapter and paper. Chapter 2 provides a review of the current literature contrasting description- and experience-based risky choices and reveals a systematic "gap" between the two formats of choice. The next three chapters outline different factors that were investigated in this thesis as potential causes of this description-experience choice gap. Chapter 3 consists of two papers which demonstrated that reliance on unrepresentative samples is a primary reason for the choice gap. Chapter 4 consists of two papers that implicate repeated, consequential choice as another important driver of the choice gap. Chapter 5 consists of three papers that examined the importance of probabilistic representation in experience-based choice. The last of these three papers outlines a new model – the exemplar confusion (Ex-CON) model – to account for experience-

based choices and probability estimates. Chapter 6 consists of a brief conclusion that draws together the insights gained from the work.

References

Beckjord, E., Ansolabehere, S., Deutch, J., Driscoll, M., Gray, P., Holdren, J., et al. (2003). *The Future of Nuclear Power: An Interdisciplinary MIT Study*. Cambridge: MIT Press.

Dedman, B. (2011). *Nuclear neighbors: Population rises near US reactors*. Retrieved August 12, 2011, from http://www.msnbc.msn.com/id/42555888/ns/us_news-life/t/nuclear-neighbors -population-rises-near-us-reactors/

Jones, J. M. (2010). *U.S. Support for Nuclear Power Climbs to New High of 62%*. Retrieved August 12, 2010, from http://www.gallup.com/poll/126827/support-nuclear-power-climbs-new-high.aspx

Nuclear Energy Institute. (2007). *Nuclear Power Plant Neighbors Accept Potential for New Reactor Nearby by Margin of Nearly 3 to 1*. Retrieved August 12, 2011, from http://www.nei.org/newsandevents/newsreleases/nuclearpowerplantneighborsacceptpotentialfornewreactornearby/

US Energy Information Administration. (2009). *Nuclear Energy Review.* Retrieved August 12, 2011, from http://www.eia.gov/emeu/aer/pdf/pages/sec9.pdf

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.

Chapter 2: Literature Review and Overview of Included Papers

A Literature Review of Decisions under Uncertainty

Life continually confronts us with choice. Often these choices are trivial, like what to eat for lunch, but sometimes these choices can be life changing, like whom to marry. A critical feature of all these choices is the degree of uncertainty associated with the possible outcomes and their likelihoods of occurring. Knight (1921), in his pioneering book, *Risk, Uncertainty and Profit*, introduced a continuum of uncertainty that has provided much of the scaffolding for subsequent developments in economic and psychological theories of choice (Rakow, 2010). One end of the continuum is anchored by a "decision under risk", where outcome probabilities are known or measurable. The other end of the continuum is anchored by a "decision under uncertainty", where outcome probabilities are unmeasurable and can only be estimated. In recent years this continuum has returned to the fore in light of the distinction between description-based and experience-based choice (Hau, Pleskac, & Hertwig, 2010; Rakow & Newell, 2010).

Consider the simple choice of whether to carry an umbrella to work in the morning. A quick scan of the online weather forecast will reveal the chance of rain. In this situation the possible outcomes – rain or no rain – and their associated probabilities are explicitly presented in a summary format. This is an example of a description-based choice and lies close to the "decision under risk" anchor in Knight's (1921) continuum. Alternatively, one could examine the appearance of the

clouds above and estimate the probability of rain based on previous encounters with similar clouds. In this situation the outcomes and their associated probabilities are gauged from a sample of outcomes sequentially observed over time. This is an example of an experience-based choice and lies more toward the "decision under uncertainty" anchor in Knight's continuum.

The distinction between description- and experience-based choices has become of particular interest in the past few years because of substantial evidence demonstrating that preferences systematically vary depending on whether information regarding choice options is learnt from a description or garnered from experience (Hertwig, in press; Hertwig & Erev, 2009).

Decisions from Description

Expected Value Theory and the St. Petersburg Paradox

In the laboratory, decisions from description have been operationalized as choices between simple monetary gambles. For example, a decision-maker may be presented with a choice between a safe option offering $3 for sure and a risky option offering a 10% chance of $32 (else nothing). Discussions between the 17th century mathematicians Blaise Pascal and Pierre de Fermat served as the basis for the earliest normative theory of description-based choice: expected value (EV) theory (see Machina, 1987, for an historical review). According to EV theory, a rational decision-maker should prefer the option that provides the highest expected value, which can be thought of as the expected long-run average and is calculated as the sum of the each outcome multiplied by its probability. For example, in the choice outlined above, the safe option has an EV of 3.0 (i.e., $3 x 1.0) whereas the risky

option has an EV of 3.2 (i.e., [$32 x 0.1] + [$0 x 0.9]). According to EV theory, the decision maker should prefer the risky option.

The inadequacy of EV theory was made apparent by the St. Petersburg Paradox, which asks how much a decision-maker is willing to pay to participate in the following gamble: a fair coin is tossed repeatedly until tails appears, at which point the decision-maker is paid a sum equal to $2*n*, where *n* is the toss number upon which tails appears (Bernoulli, 1738 / 1967). Note that in this particular problem the EV of the gamble is actually infinite and so a decision-maker, according to EV theory, would be expected to pay quite a large sum of money in order to play the gamble. In practice, however, few are willing to pay more than a few dollars to play, which highlights a critical problem with EV theory (Lopes, 1981).

Expected Utility Theory and the Allais and Ellsberg Paradoxes

In order to maintain the rationality of combining outcomes and probabilities, a slight modification to EV theory was proposed that realigned theory with behaviour: the concept of objective value was replaced with the concept of subjective utility, specifically, a diminishing marginal utility of money was assumed (Bernoulli, 1738 / 1967). Under this assumption, utility increases with value but at an ever decreasing rate. This modification was formalised into Expected Utility Theory (EUT), which outlined the axioms required to preserve rationality of choice, including rules of completeness, transitivity, continuity, and independence (Von Neumann & Morgenstern, 1947). While EUT treated probability information as an objective measure of an outcome's relative frequency, Subjective EUT (SEUT) treated probability information as the decision maker's degree of belief in the likely realization of outcomes (Savage, 1954).

Several observations have subsequently demonstrated the inadequacy of S/EUT as a model of human choice. The Allais Paradox, for example, shows clear violation of the independence axiom, which asserts that two identical outcomes within a gamble should be treated as irrelevant to the analysis of the gamble as a whole. The Ellsberg Paradox, moreover, shows that decision-makers do not make assumptions of probability when information is lacking but instead display an irrational ambiguity aversion in that they strictly prefer to bet on a gamble with definite rather than indefinite probability information (Ellsberg, 1961). Kahneman and Tversky (1979) demonstrated a number of additional S/EUT violations including the certainty effect (preference for certain outcomes), the reflection effect (divergent preferences between gain and loss domains), and loss aversion (preference to avoid losses over acquiring equivalent gains).

Prospect Theory

In the recent decades the investigation of description-based choice has revealed a fourfold pattern of risk preferences: risk aversion for gains and risk seeking for losses of high probability but risk seeking for gains and risk aversion for losses of low probability (Kahneman & Tversky, 1979, 2000). These patterns of preference led Kahneman and Tversky to advance a new, descriptive theory of choice called Prospect Theory (PT).

Although PT maintains EUT's idea of utility maximization, it introduces a number of key modifications that serve to account for the fourfold pattern of risk preferences. First, according to PT, decision-makers assess gambles in terms of deviations from a reference point as opposed to absolute states of wealth. Moreover, PT asserts that decision-makers value relative gains and losses according to an 'S' shape curve that is concave in the domain of gains but convex in the domain of

losses. That is, the marginal impact of a change in value diminishes with the distance from the reference point. Second, the shape of the S-curve is much steeper for losses than for gains, implying that losses loom larger than gains. Third, a prospect's value is distorted by a nonlinear decision weight function that modifies the impact of probabilities. The decision-weight function follows an inverse S-curve which implies that very low probabilities are overweighted and that moderate and high probabilities are underweighted. Under cumulative prospect theory (CPT), a more recent edition of the theory that avoids violation of first order stochastic dominance, weighting is applied to the cumulative probability distribution function rather than to the probabilities of individual outcomes (Fox & Tversky, 1998; Tversky & Fox, 1995; Tversky & Kahneman, 1992).

PT has proved to be highly successful accounting for a wide range of risky choice phenomena (Kahneman & Tversky, 2000). However, the majority of these phenomena have been in observed description-based tasks. Intriguingly, a growing literature has found that many of these phenomena are absent or reversed in experience-based choice.

Decisions from Experience

A decision from experience is one in which the decision-makers' incomplete knowledge of possible outcomes and/or their likelihoods is derived, at least in part, from a sampling process (Hadar & Fox, 2009). Experience-based choices have primarily been studied using the feedback paradigm and the sampling paradigm. Choices in both of these paradigms have been shown to produce preferences that systematically diverge from the preferences observed in description-based tasks, a

phenomenon that has since been termed the "description-experience gap" (Hertwig, Barron, Weber, & Erev, 2004; Hertwig & Erev, 2009; Rakow & Newell, 2010).

The Description-Experience Gap: The Feedback Paradigm

In the typical feedback paradigm task, the decision-maker is presented with the alternative options and encouraged to sample outcomes from each option in any order (e.g., Barron & Erev, 2003). Each sample briefly reveals a randomly selected outcome, with replacement, from a hidden distribution associated with the option. The distribution is often simple, containing just one or two outcomes. The decision-maker is encouraged to sample from all options in order to learn the outcomes available and also their probability of occurring. Crucially, each sampled outcome adds to a running total that is constantly displayed to the decision-maker. The decision-maker is not informed how many samples will be granted but is encouraged to earn the highest score. Thus, the decision-maker is faced with a tension between the objectives of learning more about the options ("explore") while also trying to maximise earnings across an unknown number of repeated, consequential choices ("exploit"; Cohen, McClure, & Yu, 2007). Preference is usually inferred as the more frequently selected option in the final block of samples. The exploration-exploitation tension can be eliminated by also providing feedback for the unselected, or foregone, alternative (e.g., Yechiam & Busemeyer, 2006).

Barron and Erev (2003) first systematically contrasted decisions from experience and decisions from description. In their study, half of the participants were randomly allocated to the Description group and were presented with a number of typical description-based problems (e.g., $3 for sure vs. 10% chance of $32, else nothing). The other half were allocated to the Experience group and were presented with a computerized "money machine" consisting of two unlabelled buttons. Each button

was associated with a different static payoff distribution that corresponded to the outcome distribution of the problems faced by those in the Description group. For example, a participant faced with a money machine reflecting the problem outlined above would find that one button always returned $3 and that the other button returned, on average, $32 on 10% of trials and $0 on 90% of trials. There were a total of 400 trials for each problem and feedback was limited to the button selected on each trial. Preference was inferred from the modal choice in the last 100 trials.

There were systematic differences in preferences between the two groups. For example, in the problem outlined above, 47% of participants in the Description group preferred the sure $3 compared to just 32% of participants in the Experience group. Indeed, participants in the Experience group showed opposite patterns of choice to those normally observed in decisions from description: certain outcomes were less attractive rather than more attractive, risk aversion was displayed in the loss domain rather than in the gain domain, and decisions were made as if rare events were underweighted rather than overweighted. Incredibly, the reverse predictions of CPT were more useful in trying to forecast experience-based choice preferences.

Many studies have since confirmed the robustness of preferences in the repeated-choice feedback paradigm, their propensity to seemingly underweight the impact of rare events, and also their independence from description-based preferences (Barron, Leider, & Stack, 2008; Barron & Yechiam, 2009; Biele, Erev, & Ert, 2009; Erev & Barron, 2005; Erev, Ert, & Yechiam, 2008; Erev, Glozman, & Hertwig, 2008; Ert & Erev, 2007; Ert & Yechiam, 2010; Fujikawa, 2009; Munichor, Erev, & Lotem, 2006; Newell & Rakow, 2007; Rakow & Miler, 2009; Shafir, Reich, Tsur, Erev, & Lotem, 2008; Yechiam, Barron, & Erev, 2005; Yechiam & Busemeyer, 2006; Yechiam, Erev, & Barron, 2006).

The Description-Experience Gap: The Sampling Paradigm

In the typical sampling paradigm task the goals of exploring and exploiting the options are separated into distinct phases (e.g., Hertwig, et al., 2004; Weber, Shafir, & Blais, 2004). During the sampling phase, the decision-maker is encouraged to sample outcomes from each option in any order. Importantly, each sampled outcome during this phase is without financial consequence and is purely for the purpose of learning the outcome distribution associated with the option. At any point during the sampling phase the decision-maker can elect to stop sampling and move on to the choice phase. During the choice phase, the decision-maker selects the option that they prefer with the goal of earning the highest score. The outcome of this single choice is added to a running tally that is hidden from the decision-maker until the end of the task.

In the seminal study of this approach, Hertwig et al. (2004) presented participants with six different problems sets containing a range of expected values and encompassing both positive and negative prospects. Half of the participants saw the problems numerically described, and the other half experienced the problems via the sampling paradigm. A strikingly different pattern of results were observed between the two groups. For example, in the problem described above ($3, 1.0; $32, 0.1), 48% preferred the sure $3 compared to just 20% of participants in the Experience group. Indeed, large and significant differences were observed for five of the six problems, all as a function of whether the rare event was positive or negative. Averaged across all six problems, the absolute difference in percentage points was 36%. Moreover, the pattern of preferences between the sampling and feedback paradigms of the experience-based choice was very similar (r = .93 between the Hertwig et al., 2004 and Barron & Erev, 2003 datasets).

Many studies have now found evidence consistent with the idea that rare events are given more weight when described than when experienced via the sampling paradigm, thus leading to a description-experience choice gap (Abdellaoui, L'Haridon, & Paraschiv, in press; Broomell, submitted; Dutt & Gonzalez, 2011; Fiedler & Unkelbach, 2011; Fox & Hadar, 2006; Gottlieb, Weiss, & Chapman, 2007; Hadar & Fox, 2009; Hau, et al., 2010; Hau, Pleskac, Kiefer, & Hertwig, 2008; Lejarraga, 2010; Ludvig & Spetch, 2011; Rakow, Demes, & Newell, 2008; Rakow & Rahim, 2010; Ungemach, Chater, & Stewart, 2009; Weber, et al., 2004).

Factors Contributing to the Description-Experience Gap

Misleading Samples

Sequentially sampling outcomes from a static distribution does not ensure that the observed sample will be representative of the underlying distribution (Hertwig, et al., 2004). This issue of misleading, or biased, samples is particularly relevant when small samples are taken from a skewed binomial distribution, which can be shown to result in fewer encounters with the rare event than expected from the objective probability (Hertwig & Pleskac, 2010). For example, if 1000 people each draw 20 samples from an option containing a rare outcome with an objective probability of 0.1, just 28.5% will encounter the rare event as expected. In contrast, 32.3% people will see the rare outcome more than expected and the majority of people – 39.2%– will experience the rare event less than expected, if at all. This threat of misleading samples is particularly relevant in the sampling paradigm because participants often display very frugal sampling behaviors and usually take a median of just 5 to 10 samples per option (Hau, et al., 2010). Such frugal sampling is thought to make choices easier by amplifying the differences between options (Hertwig & Pleskac,

2008). Consistent with this hypothesis, Hertwig et al. (2004) found that 78% of participants had sampled the rare event less than expected, and this experience had a distinct impact on choices. For example, when the rare event was a positive outcome – like the 10% chance of 32 in the problem described above – under-sampling the rare event led 23% to prefer this risky option compared to 57% when the rare event had been sampled equal to or greater than expected.

Subsequent research has debated whether the description-experience gap can be entirely explained as a statistical phenomenon caused by misleading samples. Fox and Hadar (2006) conducted a reanalysis of the Hertwig et al. (2004) data and found that PT could satisfactorily account for both description and experience-based choices when based on the outcome probabilities actually experienced by the participants (as opposed to the objective, underlying outcome probabilities). Also in support of the statistical account, Rakow et al. (2008) yoked the description-based problems faced by one group of participants to the actual outcome distributions observed by another group of participants facing experience-based problems. They found that elimination of misleading samples also eliminated the choice gap. However, Hau et al. (2010) subsequently showed that this null effect was carried predominately by cases in which samples had been particularly frugal and had rendered the choice trivial (e.g., 100% chance of $3 vs. 100% chance of $0).

Other studies have observed the choice gap even in the absence of misleading samples. Ungemach, Chater, and Stewart (2009) removed the impact of sampling bias by obliging participants to sample 40 times from each option while ensuring that all samples were representative of the underlying outcome distribution. For example, a participant faced with problem described above would eventually select the risky options 40 times and observe $32 exactly 4 times and $0 exactly 36 times.

Participants were free to sample the options in any order, and the order of the outcomes was random. They found that although the size of the gap was reduced when compared to those in a free sampling condition, it was not eliminated. This finding was replicated in a study by Hau et al. (2008) who in one experiment incentivized participants to sample extensively with the prospect of large potential rewards and in another experiment obliged participants to sample 100 times. Although in both cases the choice gap closed in size, it nevertheless remained apparent when averaging across problems. Together, these results suggest that misleading samples is a primary, but not sole, cause of the choice gap.

Memory Order Effects

Sequentially observing a sample that is representative of the underlying distribution does not ensure that all outcomes will be weighted equally, or even considered, when making a choice. Such potential memory order effects are particularly relevant given that research on memory (Atkinson & Shiffrin, 1968) and belief updating (Hogarth & Einhorn, 1992) have demonstrated that the order in which outcomes are experienced can influence the weight accorded to those outcomes. Moreover, according to Kareev's narrow window hypothesis (1995, 2000), people tend to make inferences based on a limited number of items in working memory, and hence, decisions are often based on a subset of experiences. Memory order effects could contribute to the choice gap if later sampled outcomes are weighted more heavily than earlier sampled outcomes because rare events are less likely than common events to have occurred recently and thus less likely to affect choice.

In support of the importance of memory order effects, Hertwig et al. (2004) found that the second half of sampled outcomes did indeed predict choices better

than the first half of sampled outcomes (75% vs. 59%, respectively). Thus, participants demonstrated a recency effect whereby outcomes observed later in the sequence were given relatively more weight when making the choice. However, subsequent experiments have produced mixed support for recency as a contributor to the choice gap. Rakow et al. (2008) found a recency effect for participants in an active sampling condition but not for those in a passive sampling condition. Rakow and Rahim (2010) found a recency effect for children but the opposite effect for adults. In addition, the description-experience gap has been observed in absence of order effects (Hau, et al., 2008; Ungemach, et al., 2009) and in cases without memory burden at all (Hau, et al., 2010). Together, these results suggest that memory order effects, especially in the form of recency, can contribute to the choice gap.

Judgment Errors

Sequentially observing a sample that is representative of the true distribution does not ensure that that the frequencies of events will be accurately represented in the mind of the decision-maker. Although frequency information appears to be automatically stored (Hasher & Zacks, 1984), estimates of probability can often be inaccurate (Erev, Wallsten, & Budescu, 1994; Lichtenstein, Slovic, Fischhoff, & Combs, 1978; Zacks & Hasher, 2002) and even the same information presented in a different format can be represented and subsequently used quite differently (Gigerenzer & Hoffrage, 1995). The gap could therefore be explained if decision-makers systematically misrepresent distributions by underestimating the probability of rare events (Fox & Hadar, 2006).

Several experiments have in fact asked participants to provide subjective estimates of probability. The methods used to gauge these estimates have included both verbal and non-verbal probes. Verbal probes have asked the decision-maker to

explicitly state the probability of each outcome occurring. Non-verbal probes have asked the decision-maker to adjust the density of a grid to correspond to the probability of each outcome occurring. In general, decision-makers have produced estimates that are well calibrated (Fox & Hadar, 2006) or that overestimate rare events (Barron & Yechiam, 2009; Broomell, submitted; Gottlieb, et al., 2007; Hau, et al., 2008; Hertwig, Pachur, & Kurzenhauser, 2005; Ungemach, et al., 2009). Interestingly, there is also evidence demonstrating that deferred frequency judgments are more accurate when probabilities are garnered from experience than learned from description (Lejarraga, 2010). Together, these results suggest that judgments of outcome distributions tend to overestimate the probability of rare events and are therefore unlikely to be a cause of choice gap given that rare events tend to be underweighted.

Format-Dependent Cognitive Strategies

The most radical explanation for the description-experience gap is that the description and experience formats of choice recruit fundamentally different cognitive strategies. Whereas description-based choice tasks seem to naturally induce strategies that combine outcomes and their probabilities, experience-based choice tasks do not demand explicit consideration of probabilities and may therefore rely on quite difference cognitive architecture. The range of these potential alternatives has been recently showcased in two model competitions (Erev et al., 2010; Hau, et al., 2008).

Models of Experience-based Choice

A number of models have been proposed to account for experience-based choices ranging from Bernoullian-inspired models that combine outcome and

probability information, to heuristics that rely on simple rules and limited information, to associative-learning models that rely on feedback and reinforcement, to instance-based learning models that rely on the recollection of stored instances. The development of these models recently culminated with two choice prediction competitions (Erev, et al., 2010; Hau, et al., 2008). The most comprehensive competition was the one organised by Erev and his colleagues, who collected two extensive datasets encompassing a broad range of 60 problems in the description, sampling, and feedback paradigms. The problems were all binary and contrasted a safe option offering a medium value with certainty with a risky option offering a high and a low value, each with some probability. Problems were evenly distributed in terms of gain, loss, or mixed frame. The estimation dataset was made public and researchers were invited to submit a model that was later tested against the competition dataset. The submitted models varied considerably in terms of underlying processes and whether probability information was explicitly represented. Note that because one of the aims of the current thesis is to put forward a new model of experience-based choice, the sections below contrast the existing models in some detail.

Bernoullian-inspired Models

Bernoullian-inspired models are those that retain the core feature of Daniel Bernoulli's EUT: the combining of information about outcomes and probabilities. The most common developments from EUT have been ''weighted utility'' models, which assume that the psychological impact or weight of probabilities is a nonlinear function (e.g., N. H. Anderson & Shanteau, 1970; Edwards, 1962; Kahneman & Tversky, 1979; Karmarkar, 1978; Luce & Narens, 1985).

The leading weighted utility model of decision making under risk is Prospect Theory (PT), which assumes that the value of a prospect is equal to the subjective value of the outcome multiplied by the impact of its probability (Kahneman & Tversky, 1979). Crucially, the value and decision weight functions are nonlinear: the value function implies diminishing sensitivity to increases in the absolute payoffs from the reference point, and the decision weight function implies that decision-makers overweight low probabilities and underweight moderate and high probabilities. Two key improvements are made with the two-stage model of cumulative PT (CPT; Fox & Tversky, 1998; Tversky & Fox, 1995; Tversky & Kahneman, 1992). First, the decision weight adjustment is applied to the decision maker's judged outcome probability as opposed to the objective or experienced probability. Second, the weight accorded to the probability of a potential outcome is a function of both the magnitude of probability and the position of the outcome in the set of possible outcomes (referred to as rank dependent weighting).

Prospect Theory has enjoyed great success accounting for description-based choices (Kahneman & Tversky, 2000). As a result, it came as little surprise when the best performing model in Erev et al.'s (2010) description-based choice competition was a stochastic variant of CPT (SCPT). SCPT assumes that the probability of selecting one particular option over another increases with the relative advantage of that option. Indeed, nearly all the top performing models in the competition were variants of PT.

Variants of PT have also been applied to experience-based choices with mixed success. Fox and Hadar (2006) reanalysed the data from Hertwig et al. (2004) and found that CPT, using the same set of parameters across choice formats, could predict choices in the description and experience formats equally well when applied

to the outcome distributions actually observed by the participants. Hau et al. (2008) conducted their own model competition with a more limited set of problems and contrasted 12 heuristics, two associative-learning models, and the two-stage model of CPT. They found that CPT was one of the two best performing models. Note, however, that the set of parameters used were fit to decisions from experience and implied nearly linear weighting of probabilities. Abdellaoui et al. (in press) used a certainty equivalents method to elicit the utility and probability weighting parameters of CPT for both description and experience (sampling) formats of choice. They found less pronounced overweighting of small probabilities and more pronounced underweighting of moderate and high probabilities for experience-based decisions. These findings suggest that PT may be able to account for experience-based choice, but only if its parameters are recalibrated with experience-based choice data. However, as evidenced by the results of the Erev et al. (2010) model completion, even when fit to experience-based choice data, the performance of PT variants is inferior to other models of choice.

Cognitive Heuristics

Cognitive heuristics are choice strategies that either entirely ignore or use only rudimentary probability information (Hau, et al., 2008). In the model competition conducted by Hau et al. (2008), the equal best performing model was a minimax strategy that simply selected the option with largest experienced minimum outcome. Another heuristic that has been shown to be very successful is the natural mean heuristic (Hertwig & Pleskac, 2008). This strategy simply selects the option that produces the largest average outcome during sampling. A variant of the natural mean model is the *k*-sampler model, which takes *k* samples from each option and then selects the option with the highest summed outcome (Erev, Glozman, et al., 2008).

This model did very well in the competition with *k* fitted to 5. Note that both of these models are successful without any conceptual parameters taking into account probability information.

An example of another successful heuristic that makes use of some probability information is the priority heuristic (Brandstatter, Gigerenzer, & Hertwig, 2006). Although the priority heuristic was initially developed for description-based choices, it did not perform statistically worse than the best strategy in the Hau et al. (2008) model competition. According to the priority heuristic, decision-makers look for an adequate reason to prefer one option over another by way of the following three-step sequence: compare minimum outcomes, then compare the minimum outcomes' probabilities, and then compare maximum outcomes. If the result of a comparison exceeds a pre-specified threshold, then a choice is made and the other steps are skipped.

The winner of Erev et al.'s (2010) sampling competition was the ensemble model, which assumes that each choice is made based on the average prediction of four equally likely rules: two versions of the *k*-sampler model, SCPT, and a stochastic version of the priority heuristic. The two *k*-sampler models took *k* as either a variable number between 1 and 9 for each participant, or drew it from the distribution of observed sample sizes in the estimation dataset. SCPT was fit with parameters that implied underweighting of rare events and a reversed S-shape value function. The modified version of the priority heuristic also included a second search order and fitted new parameters for threshold values. The victory of the ensemble model is somewhat dimmed by its complexity – 4 sub-models and 40 free parameters – although such complexity may reflect the fact that decision-makers employ a number of different strategies. This speculation should not overshadow the finding

that experience-based choices are consistent with a number of heuristic strategies that do not explicitly represent probability information or else do so at a very basic level.

Associative Learning Models

Associative learning models conceptualize choice as a learned response to experienced outcomes (Bush & Mosteller, 1955; Sutton & Barto, 1998). Under these models good experiences increase the propensity to select the same option again whereas bad experiences decrease the propensity to select the same option again (Denrell, 2007; March, 1996). Importantly, the concepts of value and probability are not distinct entities and are therefore not reinforced; rather it is the "attractiveness" of each option that is reinforced.

The value-updating model calculates the value of an option as the weighted average of the previously estimated value and the value of the most recently experienced outcome (Hertwig, Barron, Weber, & Erev, 2006). The estimated value for each option starts at zero and is then updated after each sample. The option with the highest value is preferred. The model had been shown to successfully predict the choices made by participants reported in Hertwig et al. (2004). When fit with a parameter indicating a substantial recency effect, the model did not perform statistically worse than the best strategy in the Hau et al. (2008) model competition.

The fractional-adjustment model modifies the propensity to select an option by the magnitude of the observed outcome and a learning rate parameter (March, 1996). The initial propensity for each option is equalised across options and then updated after each sample. The option with the highest propensity is preferred. The model was shown to successfully predict the choices made by participants reported in Weber et al. (2004). When fit with a parameter indicating a slow change of choice

propensity, the model did not perform statistically worse than the best strategy in the Hau et al. (2008) model competition.

Although a normalised reinforcement learning model did well in Erev et al.'s (2010) feedback model competition, it was not the best performing model. Nevertheless, associative learning models provide another class of choice strategies that are unique to predicting experience-based choices and perform well without explicit reference to probability information.

Instance-Based Learning Models

Case- or exemplar- or instance-based learning (IBL) models suggest that decision-makers' choices occur through a learning process comprising of the accumulation, recognition, and refinement of instances (Gilboa & Schmeidler, 1995; Gonzalez, Lerch, & Lebiere, 2003). Instances contain information on the decision-making situation, the action, and the result of a decision.

The winner of the Erev et al. (2010) feedback competition was a variant of IBL theory utilizing the ACT-R architecture with sequential dependencies and blended memory (J. R. Anderson & Lebiere, 1998). Since the conclusion of the competition a simpler IBL model has been put forward that produces superior performance than any of the competition models (Lejarraga, Dutt, & Gonzalez, 2011). This simpler IBL model compares and then selects the alternative with the highest "blended value". Similar to aspects of Bernoullian-inspired accounts, the blended value is the summation of all observed outcomes weighted by their probability of retrieval. The probability of retrieval is a function of the activation of each specific outcome relative to the total amount of activation for all outcomes. Similar to associative learning accounts, activation is a function of an outcome's frequency of occurrence and also the time since the outcome was last observed. However, unlike associative

learning models, the IBL model maintains the separation between the concept of values and probabilities.

The IBL model is distinctive in that it is the only model that successfully generalises across a range of experience-based choices tasks: The model has been shown to accurately predict behaviour in the feedback paradigm, a probability learning task, a repeated binary choice task within a changing environment (Gonzalez & Dutt, 2011; Lejarraga, et al., 2011) and, with the addition of an inertia parameter, also to decisions from experience in market entry games (Gonzalez, Dutt, & Lejarraga, 2011). With an additional modification to introduce a stopping rule, the IBL model has also been shown to predict the Erev et al. (2010) competition dataset better than any of the models submitted to the sampling competition (Gonzalez & Dutt, 2011). In this way, the IBL model begins to answer the challenge to create models that accurately predict behaviour across a range of tasks with similar underlying mechanisms (Cassimatis, Bello, & Langley, 2010).

The choice competition revealed that fundamentally different models are successful in accounting for description and experience choice (Erev, et al., 2010). For example, SCPT – the winner of the description competition, which combined outcomes and their probabilities – did not perform well in either of the experience competitions. Indeed, several of the more successful models used to account for experience-based choices did not explicitly represent probability information. If these types of models do in fact reflect the process of experience-based choice, then format-dependent cognitive strategies may indeed be a cause of the description-experience gap.

Aims of the Thesis and Overview of Included Papers

The broad aim of the current thesis was to reveal the psychological mechanisms underlying experience-based choice with specific subgoals to (1) explain why preferences made under the experience format often diverge from the preferences made under the description format, (2) illuminate the similarities and differences between the different types of experience-based choice, (3) reveal whether probabilistic representation is relevant in the context of experience-based choice, and (4) design a new model that can successfully account for experience-based choices. These subgoals were achieved through a series of parallel investigations, culminating in the seven papers outlined in Table 1.

Table 1 (continued on next page)

Publications included in the current thesis

Chapter	Paper	Reference
3: The Effect of Sampling Biases	1	Camilleri & Newell (2009). Within-subject preference reversals in description- and experience-based choice. In N. Taatgen, J. van Rijn, J. Nerbonne & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 449-454). Austin, TX: Cognitive Science Society.
	2	Camilleri, A. R., & Newell, B. R (2011). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276-284.
4: The Role of Repeated Choices	3	Camilleri, A. R., & Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*, 377-384.
	4	Camilleri, A. R. & Newell, B. R. (submitted). The long and short of it: Closing the description-experience "gap" by taking the long run view.

Table 1 (continued from last page)

Publications included in the current thesis

Chapter	Paper	Reference
5: The Importance of Probabilistic Representation	5	Camilleri & Newell (2009). The role of representation in experience-based choice. *Judgment and Decision Making, 4*, 518-529.
	6	Camilleri, A. R., & Newell, B. R. (2011). The relevance of a probabilistic mindset in risky choice. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2794-2799). Austin, TX: Cognitive Science Society.
	7	Hawkins, G., Camilleri, A. R., Newell, B. R., & Brown, S. (submitted). Modeling probability estimates and choice in decisions from experience.

The Effect of Sampling Biases

The first series of papers examined the issue of whether the description-experience gap can be explained as statistical phenomenon – a sample-population gap – due to misleading samples that are not representative of the underlying outcome distribution.

The first paper – Camilleri and Newell (2009) "Within-subject preference reversals in description- and experience-based choice" – used a novel within-subjects design. We found that the individuals tended to reverse their preference to the same problem presented in description and experience formats. We also found that a pen-and-pencil measure of risk attitude – the DOSPERT (Blais & Weber, 2006; Weber, Blais, & Betz, 2002) – predicted the number of description-based risky choices but not experience-based risky choices. However, because participants were free to terminate sampling at any time, many did not observe a representative sample. To account for such sampling biases, we used a novel binning procedure that sorted choices depending on how near or far their observed distribution was from the true

distribution. We found no evidence of a description-experience gap when focusing only on the choices where a representative sample was taken. We concluded that when samples are representative then the gap all but disappears; thus, the gap appears to primarily be a statistical phenomenon, at least when tested using the sampling paradigm.

The second paper – Camilleri & Newell (2011) "Description- and experience-based choice: Does equivalent information equal equivalent choice?" – contained two experiments. The first experiment was primarily a replication of the experiment reported in the first paper using a between-subjects design. To improve the number of representative samples and reduce memory order effects whilst preserving the freedom to terminate sampling, we used a novel procedure that manipulated the outcomes presented on each sample to drive the sample to be representative. Although the results again pointed to the gap being a statistical problem, the retained subset of data was not representative across participants or problems. To rectify this problem, we replicated the experiment but this time entirely eliminated misleading samples by allowing participants the freedom to select the number of perfectly representative sample sets to observe. Under these conditions the choice gap was again all but eliminated and there was no evidence of memory order effects. We concluded that decision-makers' choices are often the same regardless of whether examined in the description or sampling paradigm when equivalent information is relied upon. However, we also noted that established description-based models will never be able to provide a complete understanding of the processes underlying experience-based choice because they lack "modules" for cognitions unique to experience-based choice, including search and stopping strategies in sampling, and the roles of memory and learning.

The Role of Repeated Choices

The second series of papers examined the persistence of the description-experience gap in the feedback paradigm, where the impact of misleading samples is small. In particular, we examined the importance of the distinction between single-shot choices, which is common to both the description and sampling paradigms, and repeated choices, which is the case in the feedback paradigm. This line of research was motivated in part by the observation that description-based choices can be shifted when presented as a choice that is repeated (Keren & Wagenaar, 1987; Lopes, 1981; Wedell & Böckenholt, 1990).

The third paper – Camilleri and Newell (2011) "When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms" – focused primarily on the difference between three versions of the experience-based choice task: the sampling paradigm, the feedback paradigm, and the "full" feedback paradigm. To facilitate comparisons, the experience-based paradigms were equated in terms of the number of trials, problems, and instructions. The full feedback paradigm was a crucial comparison group because it eliminated the participants' need to explore options yet preserved the element of making repeated consequential choices. We found that preferences were very similar between the two feedback conditions, but both of these groups produced preferences quite distinct from those produced by the sampling task. When we fit the data to CPT, we found clear evidence of underweighting of rare events only in the two feedback paradigms. We concluded that the persistence of a gap between the description- and experience-based choice formats when observed in the absence of sampling biases may be related to the act of making repeated, consequential choices.

The fourth paper – Camilleri and Newell (submitted) "The long and short of it: Closing the description-experience "gap" by taking the long run view" – was an attempt to confirm the important role of repeated choice as a key difference between description and feedback choice tasks. We hypothesised that the one-shot play in the description task induced overweighting of rare events through the presentation of the small probability outcome, and that the repeated-play in the feedback task induced underweighting of rare events through a heavy reliance on recent outcomes. We hypothesised that these two effects might be reduced or eliminated if decision-makers were asked to consider their choice with a long-run horizon. To test this idea, we again contrasted description and full feedback groups but this time implemented a novel dependent measure that gave all participants the opportunity to allocate 100 plays between the two alternatives after learning about the alternatives. Using this long horizon measure, the choice differences between the description and feedback groups was all but eliminated. We concluded that the persistence of preferences consistent with underweighting in the feedback paradigm may depend critically on the sequential nature of the repeated choices.

The Importance of Probabilistic Representation

The third series of papers examined the issue of whether the description-experience gap can be explained in terms of judgment errors and also how probability information is represented when acquired from sequential samples of experience.

The fifth paper – Camilleri and Newell (2009) "The role of representation in experience-based choice" – examined whether the choice gap could be at least partially explained as a representational phenomenon, that is, the distortion of the

observed outcome distribution at the time of encoding. To assess this possibility, we probed participants to provide estimates of the outcome probability for each outcome. To allow for the possibility that decision-makers do not explicitly represent probabilities when options are learned from a sampling task, judgment probes were either verbal or non-verbal. Consistent with the idea of distinct representational formats, the verbal percentage probe produced better calibrated judgments for those in the description format whereas the non-verbal grid probe produced better calibrated judgments for those in the experience format. Although we found that participants in the description group showed a greater tendency to overestimate rare events, there was no evidence that the effect of presentation format on choice was mediated by its effect on judgment. We concluded that description and experience formats may indeed induce different probability representations, but these are not necessarily a direct cause of the choice gap. We speculated that gap may derive from a probabilistic focus in the description format and a non-probabilistic focus in the experience format.

The sixth paper – Camilleri and Newell (2011) "The relevance of a probabilistic mindset in risky choice" – explored whether the description-experience gap could indeed be explained as a difference in probabilistic mindset, that is, the explicit consideration of probability information in the former but not the latter. We induced a non-probabilistic mindset in the description condition by presenting information in frequency format and a probabilistic mindset in the sampling condition by requiring participants to periodically report the probability of different outcomes. We again found a description-experience gap and a tendency to overestimate rare events in both tasks, but particularly in the description format. However, there was no reliable effect for mindset and estimated probabilities were again not a good predictor of

choice. We concluded that explicit consideration of outcome probabilities in the description but not the experience format is unlikely to be a direct cause of the choice gap. The gap in our data appeared to be primarily a result of a memory order effect.

The seventh paper – Hawkins, Camilleri, Newell, and Brown (submitted) "Modeling probability estimates and choice in decisions from experience" – advanced a new model of experience-based choice: the exemplar confusion (ExCON) model. The goal of the ExCON model was to provide an account of how sequentially sampled outcomes are used to form a representation of the outcome distribution, and how that representation is used to form a preference. The model was broadly aimed and designed to account for both probability estimates and choices in the feedback and sampling paradigms. The ExCON is an instance-based model that stores instances – what we call "exemplars" – on each trial. Crucially, the storage of each exemplar is associated with a small probability of memory interference such that currently stored exemplars become "confused". The memory store is envisioned to be limitless and all stored exemplars – veridical or otherwise – are equally considered at the point of choice. Building from the implication of the previous experiment that probability information may indeed be explicitly represented in experience-based choice, the ExCON combines each outcome with its estimated probability of occurring and then selects the option that maximises utility.

In order to rigorously test the ExCON model, we conducted a new experiment that presented participants with binary choices between 5-outcome options in the sampling and feedback paradigms. We also asked each participant to estimate the probability associated with each outcome. The ExCON model was able to account for the tendency to overestimate rare outcomes and also did well at predicting choice preferences. When the ExCON model was entered into the Technion Prediction

Tournament (Erev et al., 2010), it won the sampling competition and came close to winning the feedback competition. We concluded that a general class of models comprising of noisy, instanced-based memory, such as those incorporated within the ExCON and the IBL models, are key when modelling experience-based choice.

References

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (in press). Experienced vs. described uncertainty: Do we need two Prospect Theory specifications? *Management Science*.

Anderson, J. R., & Lebiere, C. L. (1998). *The Atomic Components of Thought*. Hillsdale, NJ Erlbaum (Lawrence).

Anderson, N. H., & Shanteau, J. C. (1970). Information integration in risky decision making. *Journal of Experimental Psychology, 84*, 441-451.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 89-195). New York: Academic Press.

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.

Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making, 4*, 447-460.

Bernoulli, D. (1738 / 1967). *Exposition of a New Theory on the Measurement of Risk* (L. Sommer, Trans.). Farnborough Hants, England: Gregg Press.

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology, 53*, 155-167.

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

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

Broomell, S. B. (submitted). The effects of information and incentives on sampling in experience based decision making.

Bush, R. R., & Mosteller, F. (1955). *Stochastic Models for Learning*. New York: Wiley.

Camilleri, A. R., & Newell, B. R. (2009a). The role of representation in experience-based choice. *Judgment and Decision Making, 4*, 518-529.

Camilleri, A. R., & Newell, B. R. (2009b). Within-subject preference reversals in description- and experience-based choice. In N. Taatgen, J. v. Rijn, J. Nerbonne & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 449-454). Austin, TX.

Camilleri, A. R., & Newell, B. R. (2011a). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276–284.

Camilleri, A. R., & Newell, B. R. (2011b). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*, 377–384.

Camilleri, A. R., & Newell, B. R. (submitted). The long and short of it: Decisions made from description and from experience.

Camilleri, A. R., & Newell, B. R. (2011). The relevance of a probabilistic mindset in risky choice. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of*

the 33rd Annual Conference of the Cognitive Science Society (pp. 2794-2799). Austin, TX: Cognitive Science Society.

Cassimatis, N. L., Bello, P., & Langley, P. (2010). Ability, breadth, and parsimony in computational models of higher-order cognition. *Cognitive Science, 32*, 1304–1322.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society, 362*, 933–942.

Denrell, J. (2007). Adaptive learning and risk taking. *Psychological Review, 114*, 177-187.

Dutt, V., & Gonzalez, C. (2011). Why do we want to delay actions on climate change? Effects of probability and timing of climate consequences. *Journal of Behavioral Decision Making, 24*, n/a.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69*, 109-135.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics, 75*, 643–669.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112*, 912-931.

Erev, I., Ert, E., Roth, A. E., Haruvy, E. E., Herzog, S., Hau, R., et al. (2010). A choice prediction competition, for choices from experience and from description. *Journal of Behavioral Decision Making, 23*, 15-47.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making, 21*, 575–597.

Erev, I., Glozman, I., & Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk and Uncertainty, 36*, 153-177.

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

Ert, E., & Erev, I. (2007). Replicated alternatives and the role of confusion, chasing, and regret in decisions from experience. *Journal of Behavioral Decision Making, 20,* 305-322.

Ert, E., & Yechiam, E. (2010). Consistent constructs in individuals' risk taking in decisions from experience. *Acta Psychologica, 134*, 225–232.

Fiedler, K., & Unkelbach, C. (2011). Lottery attractiveness and presentation mode of probability and value information. *Journal of Behavioral Decision Making, 24,* 99-115.

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

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science, 44*, 879-895.

Fujikawa, T. (2009). On the relative importance of the hot stove effect and the tendency to rely on small samples. *Judgment and Decision Making, 4*, 429-435.

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

Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics, 110,* 605-639.

Gonzalez, C., & Dutt, V. (2011). Instance-Based learning: Integrating sampling and repeated decisions from experience. *Psychological Review.*

Gonzalez, C., Dutt, V., & Lejarraga, T. (2011). A loser can be a winner: Comparison of two instance-based learning models in a market entry competition. *Games,* 2, 136-162.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591–635.

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18*, 240-246.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making, 4*, 317-325.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372-1388.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23,* 48 - 68.

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.

Hawkins, G., Camilleri, A. R., Newell, B. R., & Brown, S. D. (submitted). The substituted exemplar model: A model of estimation and choice in a sequential sampling paradigm.

Hertwig, R. (in press). The psychology and rationality of decisions from experience. *Synthese*.

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.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2006). The role of information sampling in risky choice. In K. Fiedler & P. Juslin (Eds.), *Information Sampling and Adaptive Cognition* (pp. 75-91). New York: Cambridge University Press.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13,* 517-523.

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

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Charter & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 209-236). Oxford, England: Oxford University Press.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*, 225-237.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1-55.

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

Kahneman, D., & Tversky, A. (2000). *Choices, Values, and Frames*. Cambridge University Press: New York.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56,* 263-269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107,* 397-402.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance, 21*, 61-72.

Keren, G., & Wagenaar, W. A. (1987). Violation of utility theory in unique and repeated gambles. J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 387-391.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit.* New York: Sentry Press.

Lejarraga, T. (2010). When experience is better than description: Time delays and complexity *Journal of Behavioral Decision Making, 23*, 100-116.

Lejarraga, T., Dutt, V., & Gonzalez, C. (2011). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*, *24.*

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

Lopes, L. L. (1981). Notes, comments, and new findings. Journal of Experimental Psychology: *Human Learning and Memory, 7*, 377-385.

Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according scale type. *Journal of Mathematical Psychology, 29*, 1-72.

Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events? *PLoS ONE, 6*, e20262.

Machina, M. (1987). Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives, 1*, 121-154.

March, J. G. (1996). Learning to be risk averse. *Psychological Review, 103*, 309–319.

Munichor, N., Erev, I., & Lotem, A. (2006). Risk attitude in small timesaving decisions. *Journal of Experimental Psychology: Applied, 12*, 129-141.

Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review, 14*, 1133-1139.

Rakow, T. (2010). Risk, uncertainty and prophet: The psychological insights of Frank H. Knight. *Judgment and Decision Making, 5*, 458–466.

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.

Rakow, T., & Miler, K. (2009). Doomed to repeat the successes of the past: History is best forgotten for repeated choices with nonstationary payoffs. *Memory & Cognition, 37*, 985-1000.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*, 1-14.

Rakow, T., & Rahim, S. B. (2010). Developmental insights into experience-based decision making. *Journal of Behavioral Decision Making, 23*, 69-82.

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

Shafir, S., Reich, T., Tsur, E., Erev, I., & Lotem, A. (2008). Perceptual accuracy and conflicting effects of certainty on risk-taking behaviour. *Nature, 453*, 917-920

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge; MA: The MIT Press.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269-283.

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.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

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.

Wedell, D. H., & Böckenholt, U. (1990). Moderation of preference reversals in the long run. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 429-438.

Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks. *Journal of Conflict Resolution, 49*, 430-439.

Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making, 19*, 1-16.

Yechiam, E., Erev, I., & Barron, G. (2006). The effect of experience on using a safety device. *Safety Science, 44*, 515-522.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & B. Tilmann (Eds.), *Frequency Processing and Cognition* (pp. 21–36). New York, NY, US: Oxford University Press.

Chapter 3: The Effect of Sampling Biases

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

Adrian R. Camilleri[1]
Ben R. Newell[1]

[1] 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.

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

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, 2011; Hau et al., 2010; 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., 2010; 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, 2011). 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

Percentage choosing the favoured option by condition

Problem	Option		% choosing favoured option	
	Favoured	Non-favoured	Description	Experience
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.

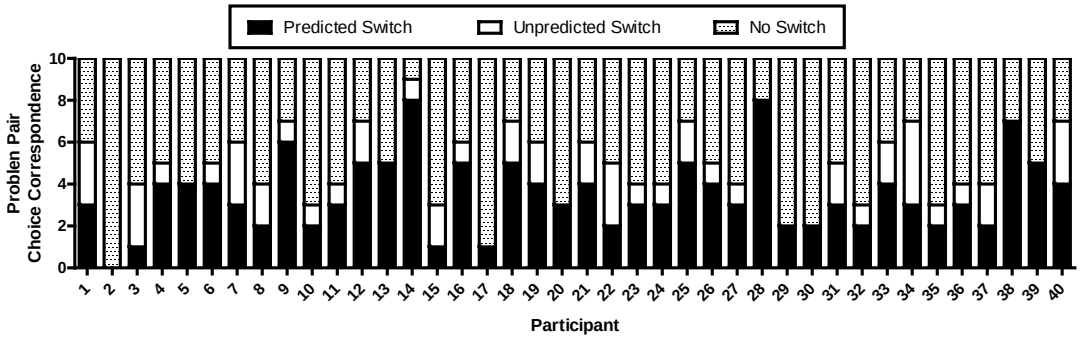


Figure 1. Degree of correspondence between description- and experience-based choices for each participant. Black bars indicate preference reversal in the predicted direction, white bars indicate preference reversal in the non-predicted direction, and dotted bars indicate no preference reversal occurred.

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 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%.

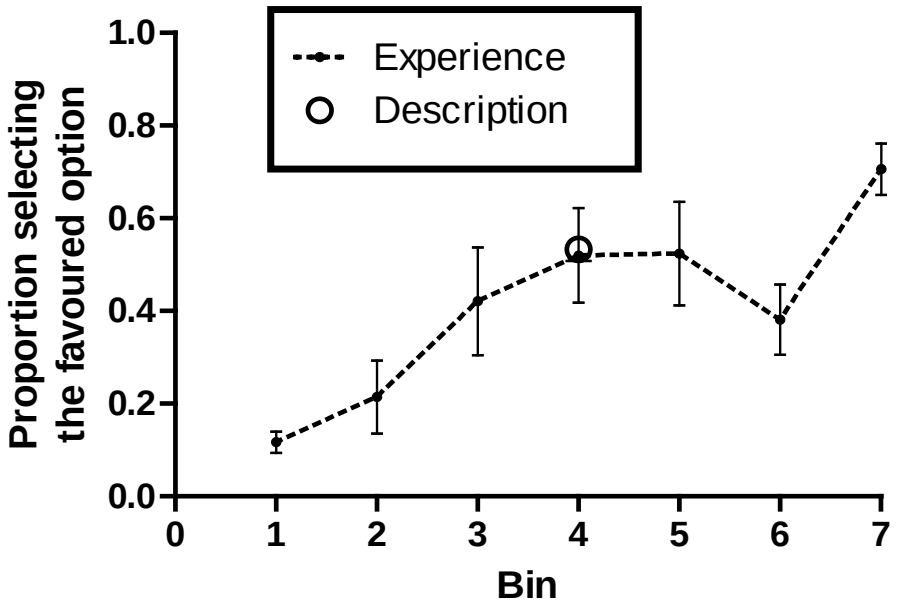


Figure 2: Percentage of participants selecting the favoured option as a function of bin.

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 (Camilleri & Newell, 2011; Hau et al., 2010; 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, 2011; 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 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, 2011).

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.

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.

Camilleri, A. R., & Newell, B. R. (2011a). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276–284.

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. (2010). Decisions from experience and statistical probabilities: Do they trigger different choices than a priori probabilities? *Journal of Behavioral Decision Making, 23*, 48 - 68.

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.

Description- and Experience-based Choice: Does Equivalent Information Equal Equivalent Choice?

Adrian R. Camilleri[1]
Ben R. Newell[1]

[1] School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Does the manner in which people acquire information affect their choices? Recent research has contrasted choices based on summary descriptions (e.g. a 100% chance of $3 vs. an 80% chance of $4) with those based on the 'experience' of drawing samples from environments that do (or should) match those provided by descriptions. Intriguingly, decision-makers' preferences differ markedly across the two formats: the so-called description-experience "gap" – but debate over the cause of this gap continues. We employed novel techniques to ensure strict control over both external and internal biases in the samples of information that people used to make decisions from experience. In line with some other recent research, we found a much diminished gap in both experiments suggesting that the divergence in choices based on description and sequentially acquired (non-consequential) samples is largely the result of non-equivalent information at the point of choice. The implications for models of risky choice are discussed.

Description- and Experience-based Choice: Does Equivalent Information Equal Equivalent Choice?

Driving is an activity that many of us undertake. Speeding is common: according to some estimates, one in every six drivers will receive a speeding ticket each year (Dallah, 2008). The decision to speed usually results in positive outcomes (e.g., destination is reached sooner) and only rarely results in negative outcomes (e.g., becoming involved in a car accident). It is therefore quite likely that you could find people, most likely males under the age of 25, making the argument that speeding is basically a good decision, especially if they have never received a speeding ticket or been involved in an accident. Such a choice could be called a decision made from *experience*. Road accident statisticians, in contrast, are probably less likely to speed. They are familiar with the statistic that speed is a related factor in more than 32% of fatal road accidents (RTA, 2007). Their choice could be called a decision made from *description*.

Since the 1970s and 1980s, spurred by the work of Daniel Kahneman and Amos Tversky (e.g., Kahneman & Tversky, 1979), decision scientists have been particularly interested in studying decisions from description. Although a genuinely productive workbench from which to examine how people choose between different monetary gambles, this paradigm ignores a range of other cognitive factors central to everyday decision-making, including the roles of experience, sampling, memory and learning. In more recent years there has been resurgence in the study of these cognitive factors and how they relate to decision-making under uncertainty. Examination of such decisions from experience has prompted decision scientists to consider more general psychological processes, including the acquisition,

representation, weighting and the integration of information prior to choice (e.g., see Rakow & Newell, 2010).

Perhaps the most interesting phenomenon to emerge from this literature so far is that description- and experience-based choices typically lead to different decisions – this has been called the description-experience "gap". Should we be surprised that young males and road accident statisticians make different decisions? At first blush, maybe not. After all, it seems obvious that there is a difference between choice based on a description specifying objective outcome probabilities and choice based on learnt contingencies between events from one's personal experience. The interesting question is *how* the mode by which information is garnered influences choice.

The Description-Experience "Gap"

Hertwig, Barron, Weber and Erev (2004) contrasted these two choice formats by presenting decision-makers with the same structural problem in either the description or the experience format. The task was to select between computerised money machines that were each associated with different static payoff distributions. In the description format of the task, each machine was clearly labeled with a specification of the outcomes and their probabilities and the participant was required to choose the alternative they preferred to play from. For example, the machine on the left may have provided a "100% chance of 3" whereas the machine on the right may have provided an "80% chance of 4, else 0" (henceforth, Problem 1). In the experience format, each machine was unlabeled and the participant was required to sample from the alternative machines by clicking on them. Each sample revealed a randomly selected outcome from the unknown payoff distribution (e.g., Figure 1). The participant was given the opportunity to freely sample from the machines in any order and as often as they liked until they were ready to choose the alternative they

preferred to play from. Importantly, the payoff distribution corresponded to the objective descriptions provided to those playing the task in the description format.

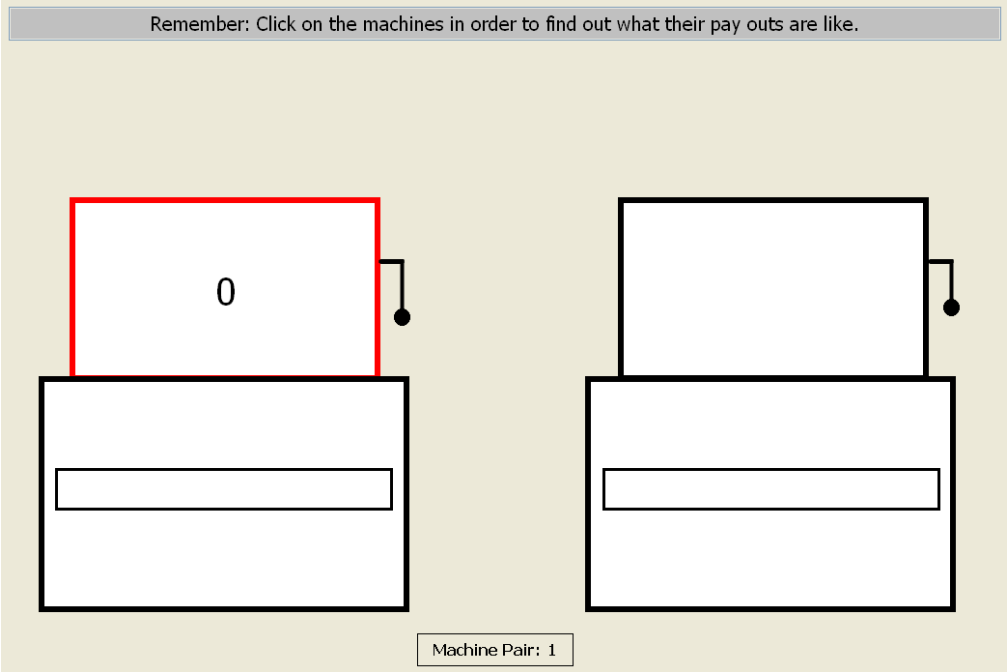


Figure 1. Screenshot of the experience-based version of the task just as the left option had been selected (revealing a 0).

A strikingly different pattern of choices was observed depending on the way the choice was presented. In Problem 1, for example, 36% of the participants selected the risky option when the decision was made from description but 88% preferred this option when the decision was made from experience. Indeed, when averaged across the six problems, the so-called description-experience "gap" was 36 percentage points in magnitude and consistent with the idea that rare events have more impact on decisions when described than when experienced (Hertwig, et al., 2004; Weber, Shafir, & Blais, 2004). This finding, combined with analogous results when using an experience-based choice paradigm where samples are also financially consequential (e.g., Barron & Erev, 2003), has led some to call for the development of separate and distinct theories of risky choice for description and experience formats.

Biased Samples

Hertwig et al. (2004) identified two sources of bias that clearly contributed to the differences observed between description- and experience-based choices.

External Biases. External sampling biases occur when an observed sample of outcomes does not accurately reflect the true outcome distribution. Just as many members of the general public who speed have never been involved in an accident, many playing the experience version of the task never encountered the rare event. Indeed, Hertwig et al. (2004) noted that experience-based choices often relied on small samples: the median number of samples taken by their participants was just 15. It can be shown that small samples, due to the skewed binomial distribution inherent in risky choice problems, results in fewer encounters with the rare event than is expected from the payoff distribution (i.e., the number of samples [N] x the probability of the rare event [p]; Hertwig & Pleskac, 2010). Hertwig et al. found that 78% of the participants had observed the rare event less than expected, and this had a distinct impact on choices. For example, when the rare event was undesirable (e.g., 0, 0.2 in Problem 1) under-sampling led 92% to prefer this risky option, compared to just 50% when sampling of the rare event was equal to or greater than expected. In light of these results, it has been argued by some that the gap is due entirely to external sampling bias and has little to do with the mode of presentation (Fox & Hadar, 2006; Hadar & Fox, 2009; Rakow, Demes, & Newell, 2008).

Internal Biases. Internal sampling biases occur when a mental sub-sample of outcomes does not accurately reflect the outcome distribution. Even members of the general public who have been involved in a speed-related accident may fail to take this experience into account when making a choice. They may simply forget about the event (Atkinson & Shiffrin, 1968) or may classify the event as irrelevant to the

current decision (Gilboa & Schmeidler, 1995). Indeed, according to Kareev's narrow window hypothesis (1995, 2000), people make inferences based on a limited number of items in working memory, and hence, decisions may often be based on a subset of experiences. Evidence for mental sample subsets was also found by Hertwig et al. (2004), who observed that participants showed a "recency" effect: outcomes observed more recently were better predictors of choice than outcomes observed earlier (see also Stewart, Chater, & Brown, 2006)[1].

The Current Study

A number of approaches have been employed in attempt to empirically eliminate sampling biases, each associated with its own set of advantages and disadvantages. For example, one popular method is to fix the sample size – typically to something large – thereby reducing external sampling bias by ensuring that a highly representative sequence is presented (e.g., Hau, et al., 2008). A consequence of this manipulation, however, is to increase internal sampling bias. This is because people prefer to rely on small samples (Hertwig & Pleskac, 2008), which they believe accurately represent the objective probability (Tversky & Kahneman, 1971) and make choices easier (Hertwig & Pleskac, 2010). Since the manipulation obliges participants to take an artificially large number of samples, it is feasible that they pay attention to, or make their choice based on, merely a subsample of the presented outcomes[2]. Another method has been to yoke described problems to the outcome

[1] Not all studies have found a recency effect (e.g., Hau, Pleskac, Kiefer, & Hertwig, 2008; Ungemach, Chater, & Stewart, 2009).

[2] Ungemach et al. (2009) found that people made accurate frequency judgements, suggesting that information from across all experienced outcomes is available at the point of choice. This evidence does not rule out internal sampling bias for two reasons. First, judgements made by those in the Ungemach et al. study comprised only in participants stating how frequently the rare outcome had been observed. This is quite distinct from participants appreciating the probability of the rare event being observed on the next sample, which additionally involves knowing of the number of samples

distribution actually experienced by participants in a free sampling experience paradigm (e.g., Rakow, et al., 2008). A problem with this approach, however, is that participants often draw very small samples that trivialise many choices (e.g., the equivalent of 100% chance of $3 vs. 100% chance of $4), which can mask any true differences (Hau, Pleskac, & Hertwig, 2010). An additional issue associated with previous experience-based choice tasks is that the outcome presented on each sample is randomly generated. As a result, any mental subset of outcomes that accords more weight to recent observations will be biased and tend to underweight rarer outcomes due to the statistical characteristics of the binomial distribution (discussed above). A more complete summary of previous attempts to account for external and internal sampling biases is presented in Table 1.

Inspection of the fifth column of Table 1 shows that these studies have produced mixed, inconclusive results (see also Hertwig & Erev, 2009; Rakow & Newell, 2010, for reviews). Preferences may differ between the two choice formats because information acquisition results in different information or because equivalent information is treated differently to arrive at a decision (or both). Our aim was to eliminate the first possibility (differences in acquired information) in order to test the second (differences in the use of information at choice).We achieved this aim by setting up two highly controlled experimental situations that employed three novel methods that largely eliminated external and internal sampling biases, thereby equating information.

taken. Second, recent evidence suggests that peoples' choice behaviour can be unrelated to their probability judgments (Barron & Yechiam, 2009; Camilleri & Newell, 2009).

Table 1 (continued on next page)

Summary of the Methods Previously used to Account for Sampling Biases in Experience-Based Choice (see text for additional details)

Experiment	Manipulation	External Sampling Bias†	Internal Sampling Bias‡	Results*	Notes
Hadar & Fox (2009). Exp. 1.	Obliged small samples and then revealed all potential outcomes.	Moderate – Sample sequences, on average, unrepresentative of the objective outcome distribution (but all outcomes known).	High – Sample sequence randomly generated. Fixed small sample length.	No gap.	-
Hau et al. (2008): Exp. 1.	Large incentives.	Moderate – Sample sequences, on average, moderately representative of the objective outcome distribution.	Moderate – Sample sequence randomly generated.	Small gap.	-
Hau et al. (2008): Exp. 2. Hau, Pleskac, and Hertwig (2010): Exp. 1. Camilleri & Newell (2011)	Obliged large samples.	Low – Sample sequences, on average, highly representative of the objective outcome distribution.	High - Sample sequence randomly generated. Fixed large sample length.	Small gap.	-

† External sampling bias refers to observation of a sample sequence that does not reflect the objective outcome distribution.

‡ Internal sampling bias refers to use of a mental subset of outcomes that does not reflect the observed outcome distribution.

* The "gap" refers to different patterns of choice observed as a function of whether options are presented as descriptions or learnt from experience.

Table 1 (continued from last page)

Summary of the Methods Previously used to Account for Sampling Biases in Experience-Based Choice (see text for additional details)

Experiment	Manipulation	External Sampling Bias†	Internal Sampling Bias‡	Results*	Notes
Ungemach, Chater and Stewart (2009).	Obliged large samples and fixed the outcome pool.	None.	High - Sample sequence randomly generated. Fixed large sample length.	Small gap.	-
Rakow, Demes, and Newell (2008).	Yoked described problems to samples that were freely taken.	None.	Moderate – Sample sequence randomly generated.	No gap.	Due to small samples, many of the gambles become trivial.
Hau, Pleskac, and Hertwig (2010): Exp. 2.	Yoked described problems to obliged large samples (with repeated choice) and access to all previous outcomes.	None.	Moderate - Sample sequence randomly generated. Fixed large sample length (but with access to previous outcomes).	Moderate gap.	Potential concern over the impact of choice inertia.

† External sampling bias refers to observation of a sample sequence that does not reflect the objective outcome distribution.

‡ Internal sampling bias refers to use of a mental subset of outcomes that does not reflect the observed outcome distribution.

* The “gap” refers to different patterns of choice observed as a function of whether options are presented as descriptions or learnt from experience.

Different choices across the two formats would support the idea that equivalent information is used non-equivalently at the point of choice. In contrast, similar choices regardless of format would suggest that the choice gap is primarily an external and internal sampling biases phenomenon and that equivalent information produces equivalent choice. Our study is therefore a response to the recommendation for the development of acquisition-specific theories of risky choice. Although existing models of description-based choice may be insufficient to explain the *process* underlying experience-based choices, we investigate whether such models are nevertheless adequate to explain the *outcome* of experience-based choices.

Experiment 1

Any new method for dealing with the problems of sampling bias must account for (1) external sampling bias, (2) internal sampling bias, and (3) trivial choices. The method we used in Experiment 1 to achieve these goals permitted participants to freely select the size of their sample and then conditionalised on the subset of occasions where participants observed an outcome distribution approximately equal to the objective distribution. In order to increase the likelihood of this match, and to directly target the threat of internal sampling bias, we also included a group in which the sequence of outcomes was manipulated. Rather than allowing each sample to reflect a random draw from a pool of numbers based on the objective probabilities, sampled outcomes were selected to improve the match between the participants' experienced outcome distribution and the objective outcome distribution. This manipulation ensured that rare outcomes were semi-evenly distributed across the entire sample.

There are five important benefits that follow from our novel method. First, the number of trials in which the experienced distribution is near or equal to the objective probability is greater than in a standard free sampling condition, thus improving statistical power. Second, the method reduces, and subsequently attempts to account for, external sampling bias while simultaneously allowing the participant to freely sample. Third, because participants are allowed to freely sample and terminate their search, factors associated with participants becoming bored and inattentive are reduced, thereby limiting the amount of internal sampling bias. Fourth, the proportion of trials upon which the choice comparisons are rendered trivial is minimised. Fifth, the impact of internal sampling bias, primarily in the form of a recency effect, is minimised because the outcome sequences observed earlier are congruent with those taken later.

Method

Participants

The participants were 102 undergraduate first year University of New South Wales psychology students (69 females), with an average age of 20.3 years and a range of 17 to 61 years. Participation was in exchange for course credit, plus payment contingent upon choices.

Materials

Decision Task. The decision task was a virtual money machine game. In the description-based version of the task, two alternative money machines were presented labelled with an explicit specification of the outcome payouts and their probabilities (e.g., 80% chance of 4, else 0). In the experience version of the task, the two alternative money machines were unlabelled. Each of the machines was

associated with a distribution of possible outcomes in accordance with the objective probabilities. Samples from each machine were draws from the respective outcome distributions (see Figure 1). Allocation of safe and risky options to the left and right machines was counterbalanced and the order of the problems was randomized.

Choice Problems. The ten choice problems are shown in first three columns of Table 2. Each problem consisted of a risky option that probabilistically paid out one of two values, and a safe option that always paid out the same value. There were seven problems in the gain domain and three problems in the loss domain to ensure that the majority of participants won money.

Design

The independent variable was the decision task (description or experience). The dependent variable was the choice made (risky or safe option). The sequence of sampled outcomes for 31 participants playing the experience version of the task was randomly generated (the Random Experience group). In contrast, the sequence of sample outcomes for another 35 participants playing the experience version of the task was pseudo-randomly generated (the Pseudo-random Experience group). For this latter experience group, outcomes presented on each individual sample were selected in order to improve the match between the objective probabilities and the participant's actual experience. Specifically, an algorithm was constructed that calculated the sequence of outcomes that would minimise the difference between the objective distribution and the participant's experienced distribution at each sample. The resulting sequence produced a repeating pattern of outcomes: for problems in which the rare event occurred 20% of the time the pattern repeated itself in blocks of five. For problems in which the rare event occurred 15% of the time, the pattern

repeated itself in blocks of twenty. For problems in which the rare event occurred 10% of the time, the pattern repeated itself in blocks of ten.

To eliminate this regularity, the samples taken by participants in the Pseudo-random Experience group were actually random draws from within each repeating block of outcomes. For example, the repeating block for the risky option in Problem 3 was 0, 0, 0, 0, 0, 32, 0, 0, 0, 0. The order of outcomes within this block was randomised for each repetition of the block and for each participant. This "jitter" prevented any systematic pattern in the sample to form, whilst nevertheless maintaining a close match between objective and actually experienced outcome probabilities. Importantly, participants did not know that there were repeating blocks nor the length of each one.

Procedure

The participant's job was to maximise the amount of points won. The instructions indicated that at the end of the experiment earned points would be converted into real money (1 point = AUD$0.10). Instructions for participants in the Description group were to compare the two labelled money machines and to choose one to play from. Instructions for participants in the Experience groups made explicit that the unlabelled machines should be clicked on in order to find out what their payoffs were like. 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. The outcome of this true play was hidden until the end of the experiment in order to reduce any wealth effects.

Results

Patterns of Choice

Table 2 displays the percentage of participants choosing the risky option in each of the three groups[3]. The difference between the Description and Experience groups falls in the expected direction, assuming rare events have more impact when described than experienced, for 19 out of the 20 comparisons. Ten of these differences were significant by individual chi-square tests (all p's < .05). Overall, the mean difference between description- and experience-based choices in the expected direction was 23.7 percentage points for the Random Experience group and 27.7 percentage points for the Pseudo-random Experience group. The only significant difference in the choices between the two experience groups was Problem 8 (χ^2 = 4.02, p = .045).

We mapped patterns of choice onto a single directional scale by re-categorizing choices in terms of whether the predicted option was preferred. The "predicted" option was the alternative that would be preferred assuming that rare events are overweighted, as is typical for description-based choices. As shown in the leftmost of Figure 2, when averaging across problems, the predicted choice was selected on 57.2% of trials in the Description group, which was significantly larger than the 33.7% of trials in the Random Experience group (χ^2 = 37.1, p < .001) and 29.9% of trials in the Pseudo-random Experience group (χ^2 = 53.7, p < .001). The odds of selecting the predicted option in the Description group were more than 2.6 times the odds of selecting the predicted option in either of the Experience groups. Thus, taken

[3] Technical error resulted in 1 trial from the Random and 2 trials from the Pseudo-random experience groups to be removed. Thus, the analyses are based on 360 trials from the Description group, 309 trials from the Random Experience group, and 348 trials from the Pseudo-random Experience group.

as a whole, our data replicate previous studies and demonstrate a description-experience "gap".

Table 2

Percentage of Participants Choosing the Risky Option in Experiments 1 and 2

Problem	Option		Percentage Choosing the Risky Option			
	Risky	Safe	Exp. 1			Exp. 2
			Description	Experience		
			(*n*=36)	Random (*n*=31)	Pseudo-random (*n*=35)	(*n*=36)
1	4 (.8)	3 (1.0)[P]	33	68[*]	77[*]	53
2	-4 (.8)[P]	-3 (1.0)	61	37[*]	46	33[*]
3	32 (.1)[P]	3 (1.0)	42	35	34	61
4	-32 (.1)	-3 (1.0)[P]	28	81[*]	83[*]	36
5	10 (.9)	9 (1.0)[P]	36	68[*]	74[*]	44
6	-10 (.9)[P]	-9 (1.0)	69	35[*]	26[*]	56
7	16 (.2)[P]	3 (1.0)	39	32	49	61
8	11 (.1)[P]	1 (1.0)	58	35	14[*]	61
9	14 (.15)[P]	2 (1.0)	50	39	37	-
10	28 (.15)[P]	4 (1.0)	50	39	29	-
Mean difference in predicted direction[#]:				23.7[*]	27.7[*]	4.1

[P] Indicates the predicted option, that is, the more favourable option if rare events are overweighted.

[*] Denotes significantly different from Description group by χ^2 ($p < .05$).

- Due to a programming error, the data for Problems 9 and 10 in Experiment 2 were lost.

[#] Predicted direction is that rare events have more of an impact on decisions when they are described than when they are experienced.

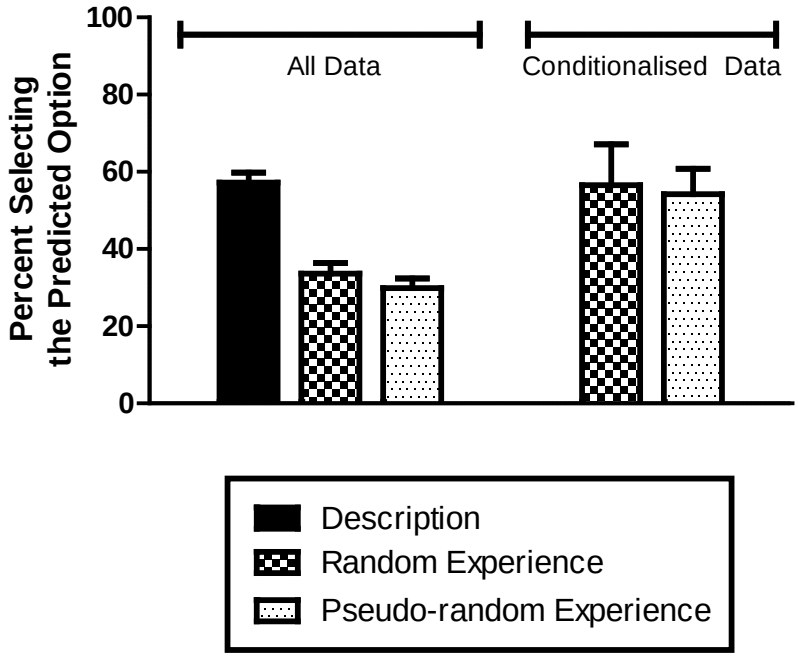


Figure 2. The percentage of participants selecting the predicted option, assuming rare events are overweighted, in the Description and two Experience groups of Experiment 1. The conditionalised data were those trials where the participants' experienced distribution was within 10% of the rare event objective probability. Error bars indicate the standard error of the mean.

Matching Experienced to Objective Outcome Distribution

To account for the impact of external sampling bias, we focused on those trials where participants' experienced distribution was ±10% of the objective distribution. In order to maintain standardization across problems with rare events of differing rarity, data were categorised as a function of the rare event objective probability: when the objective probability was 10%, subjective experiences of the rare event between 9 and 11% were conditionalised upon (i.e., 1/10 of 10% = 10±1%), when the objective probability was 15%, subjective experiences of the rare event between 13.5 and 16.5% were conditionalised upon (i.e., 1/10 of 15% = 15±1.5%), and when the objective probability was 20%, subjective experiences of the rare event between 18 and 22% were conditionalised upon (i.e., 1/10 of 20% = 20±2%). As shown in

Table 3, the amount of trials that satisfied this criterion in the Random Experience group was very low (7%). In contrast, the amount of trials that satisfied this criterion in the Pseudo-random Experience group was larger (17%). Thus, our sample manipulation successfully decreased external sampling bias and more participants freely observed a representative sample.

Table 3

Extent of Sampling and Contribution to Conditionalised Data for each Problem in Experiment 1

Problem	Option		Median number of samples taken across both options		Number of participants contributing to conditionalised data*	
	Risky	Safe	Random Experience	Pseudo-random Experience	Random Experience	Pseudo-random Experience
1	4 (.8)	3 (1.0)	9	10	3/31	12/35
2	-4 (.8)	-3 (1.0)	14	12	5/30	9/34
3	32 (.1)	3 (1.0)	10	10	2/31	3/35
4	-32 (.1)	-3 (1.0)	9	9	0/31	3/35
5	10 (.9)	9 (1.0)	8	7	1/31	3/35
6	-10 (.9)	-9 (1.0)	12	13	1/31	5/35
7	16 (.2)	3 (1.0)	10	13	7/31	11/35
8	11 (.1)	1 (1.0)	8	14	1/31	7/35
9	14 (.15)	2 (1.0)	10	9	2/31	3/34
10	28 (.15)	4 (1.0)	8	8	1/31	3/35

* Conditionalised data were those trials where participants' experienced distribution was ±10% of the objective distribution (see text for more details). The denominator changes across problems due to lost data (see Footnote 3). In total, 7.4% (23/309) and 16.9% (59/348) of trials contributed to the conditionalised data in the Random Experience and Pseudo-random Experience groups, respectively.

As shown in the rightmost of Figure 2, the percentage of trials in which participants selected the predicted option is remarkably similar across the Description and the Random and Pseudo-random Experience subset data: 57.2%, 56.5%, and 54.2% respectively (Fischer's Exact Test; all pairwise *p*'s > .05, one-tailed). The odds ratios were all trivially small. Additionally, there was no difference in preference for the predicted option between the Description group and the average of the two Experience groups (57.2% vs. 54.9%, respectively; $p > 1$, one-tailed). In this latter comparison, our power to detect a difference of the size generally reported in the literature (i.e., odds-ratio of greater than 2.5) was approximately 97%[4].

Memory Effects

Following Hertwig et al. (2004), we looked for memory order effects. We illustrate the method with the example of a participant sampling from the Problem 1 who observed the following outcomes 4,4,4,3,0,4,3,3,4,3 before deciding to play from the safe option. First, we separated out the samples from each option (e.g. 4,4,4,0,4,4 and 3,3,3,3). Second, we grouped the first and second half of each option's sampling sequence together (e.g., 4,4,4,3,3 and 0,4,4,3,3[5]). Third, for each half of the samples, we computed each option's average payoff (e.g., in the first half, the average is 4 for the risky option and 3 for the safe option whereas in the second half, the average is 2.7 for the risky option and 3 for the safe option). Fourth, we predicted choice based on which option had the higher average payoff (e.g., risky option is predicted to be preferred when considering only the first half of samples, but the safe option is predicted when considering only the second half of samples).

[4] Calculated with G*Power3 (Erdfelder, Faul, & Buchner, 1996) under the "Exact" test family for the "Proportions: Inequality, two independent groups" statistical test and the following input parameters: tails = 1, odds ratio = 2.5, $\alpha = .05$, sample size group 1 = 360, sample size group 2 = 82.

[5] Where there were an odd number of samples, each half of the sample was allocated half of the middle number (and .5 was added to the denominator when the average was calculated on the next step).

Fifth, the predicted choice was compared to the actual choice made by the participant (participant in this case opted for the safer option, thus demonstrating a recency effect). In the subset of conditionalised data, we found a recency trend in the Random group (39% vs. 65%, $\chi^2 = 3.17$, $p = .077$) but no evidence in the Pseudo-random group (47% vs. 54% $\chi^2 = .543$, $p = .461$). Similar results were found when comparing just the first versus last ten samples. Thus, our manipulation also successfully reduced the impact of internal sampling bias.

Discussion

We observed no differences in preferences when conditionalising on the subset of data where the experienced distribution was approximately equivalent to the objective distribution presented to those in the description-based choice task. The conclusion that follows from this analysis is that the description-experience "gap" all but disappears when external and internal sampling biases are accounted for. However, this conclusion must be presented with some degree of caution because conditionalising on the data had two nontrivial consequences.

First, conditionalising necessitated discarding a large proportion of the data. Even in the Pseudo-random group, where we manipulated the sequence of outcomes, 83% of the trials were ignored. Of course, these data were ignored with good reason: they were the trials where the rare event had never been seen, or where the experienced outcome distribution was skewed and therefore did not accurately represent the true outcome distribution specified to those in the Description group. Certainly, we were surprised to find how difficult it was to drive experienced samples to closely represent the population distribution while permitting participants

to decide when to stop sampling. This problem is symptomatic of the free sampling paradigm in general.

Second, and specific to our method, the retained subset of data was not representative across participants or problems. Participants that sampled more frequently and problems with relatively less extreme outcome rarity were more highly represented (see Table 3). Thus, comparison of these subset data with those of the Description group data, which equally represented all problems and participants, is complicated. For example, a close examination of Tables 2 and 3 reveals that problems which found no gap to begin with were over-represented in the conditionalised subset of trials (e.g., Problem 7). Thus, to some extent, the outcome of our method of analysis depends on the problems used and the sampling motivation of participants.

Experiment 1 thus serves to highlight an important methodological point: even with such a seemingly simple paradigm there exist important subtleties that can nevertheless lead to non-trivial choice divergences (e.g., Fox & Hadar, 2006; Hadar & Fox, 2009). In light of these two complications, we carried out Experiment 2 to see if our conclusion held when a method that avoided these issues was used.

Experiment 2

In Experiment 2 we used a new variation of the sampling paradigm in which participants were exposed to a sample that was perfectly representative of the objective outcome distribution yet were still provided with moderate freedom to choose the number of samples. To reiterate, we contend that freedom to choose the length of the sample sequence is important for minimising attentional failures and internal sampling bias (see Kareev, 1995; 2000). Indeed, Rakow et al. (2008) found

that sampling behaviour is related to working memory capacity. Previous methods have either allowed participants to freely sample (e.g., Experiment 1), or obliged, typically, large samples (e.g., Ungemach, et al., 2009). As noted in Table 1, each method has its own advantages and disadvantages. Here, we find a compromise by obliging a small number of perfectly representative samples (a "block") while allowing participants the freedom to select the number of blocks of trials to observe.

Method

Participants

The participants were 36 first year psychology students from UNSW with median age of 18 years and a range of 18 to 21 years. Participation was in exchange for course credit and money dependent on choices.

Materials

*Choice Problems.*The choice problems were the same as those used in Experiment 1. Unfortunately, due to a programming error, the data for Problems 9 and 10 were lost.

Design and Procedure

Since the same problems as in Experiment 1 were used, we contrasted the existing Description group data from Experiment 1 with the new Experience group data.

Participants were asked to sample from the two alternative options in any order that they preferred. Unlike Experiment 1, an option became unresponsive after a "block" of samples had been observed. Each block of samples comprised of a randomly ordered sequence of outcomes that perfectly matched the true outcome distribution. For example, when the objective probability of the rare event was 20%

then a block consisted of ten samples and the rare event was randomly presented twice.

Once an option became unresponsive, it could not be sampled from again until a block of samples was also made from the alternative option. Participants could switch back and forth between the options freely until the options became unresponsive. The sequence of the outcomes was randomised for each block and each participant. After a block of samples had been taken from each option, the participant was given the choice to sample another block of trials from each option or to make a choice. This method ensured that all participants were exposed to a sample perfectly representative of the objective description, while maximising the amount of freedom they had to determine the size of their sample. When the participant opted to make a choice, they selected their preferred option to play from and the hidden result of that choice was added to their running total.

At the end of the experiment, participants were presented with a free response question that asked the participant to report the strategy or strategies that they used to make the choices throughout the experiment.

Results

Number of Blocks Sampled

Participants were free to choose the number of blocks of trials that they would sample. At least one extra block of trials was taken on 13.9% of occasions. The average number of blocks observed across all eight problems was 1.17 (*SD* = .46), which corresponds to an average of 23.3 (*SD* = 9.3) total samples. Inspection of individual data reveals that many participants elected to sample a second block of

trials on the first one or two problems that they encountered and then sampled only a single block of trials for the remaining problems.

Patterns of Choice

The percentage choosing the risky option in the Experience group of Experiment 2 is contrasted with the Description group of Experiment 1 in Table 2. At the level of individual problems, there was a reliable difference between groups only for Problem 2 ($\chi^2 = 5.57$, $p < .05$). As in Experiment 1, we re-categorized choices in terms of whether the predicted option was preferred. Averaging across problems, the predicted choice was selected on 54.9% of trials in the Experience group, which was not different from the 59.0% of trials in the Description group from Experiment 1 (Fischer's Exact Test; $p = .177$, one-tailed; see Figure 3). The odds of selecting the predicted option in the Description group were just 1.2 times the odds of selecting the predicted option in the Experience group. The power to detect a difference of the size generally reported in the literature (i.e., odds-ratio of greater than 2.5) was approximately 99%[6]. Thus, our data did not show a reliable description-experience choice gap.

Memory Effects

We found no evidence for a recency effect: there was no difference in choice prediction accuracy when based on the first half versus second half of observed outcomes (57% vs. 51%, respectively, $\chi^2 = 1.78$, $p = .18$). We also found no difference when comparing the prediction accuracy of the last versus first ten outcomes. Admittedly the blocked nature of the design, which helped to ensure that early and late trials were similar, made memory order effects more difficult to detect.

[6] Calculated with G*Power3 (Erdfelder, et al., 1996) under the "Exact" test family for the "Proportions: Inequality, two independent groups" statistical test and the following input parameters: tails = 1, odds ratio = 2.5, $\alpha = .05$, sample size group 1 = 288, sample size group 2 = 288.

Nevertheless, our manipulation successfully reduced internal sampling bias to the extent that that the impact of participants differentially weighting early or later observations was neutralised[7].

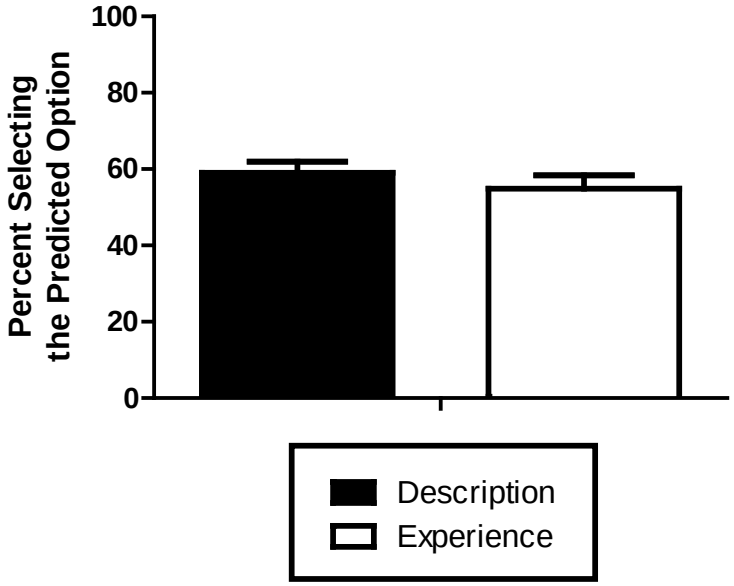


Figure 3. The percentage of participants selecting the predicted option, assuming rare events are overweighted, in the Description group (Experiment 1) and Experience group (Experiment 2) for Problems 1-8. Error bars indicate the standard error of the mean.

Discussion

Previous attempts to isolate factors contributing to the description-experience gap have run into difficulty because of the (1) comparison of non-equivalent problems caused by external and internal sampling biases, (2) comparison of trivial (but equivalent) problems due to yoking, and (3) asymmetrical elimination of large amounts of data to conditionalise samples that match the true distribution. The results of Experiment 2 indicate that when these issues are resolved, then the choice gap all but disappears[8].

[7] Our design does not rule out alternative forms of internal sampling bias such as the peak-end rule (Fredrickson & Kahneman, 1993).

[8] A reviewer noted an interesting trend in the data reported in Table 2. Specifically, in Experiment 2, the majority preference was for the option with the higher expected value in all six possible cases (i.e., Problems 1-6 since the EVs were the same for Problems 7 and 8). By contrast, in

General Discussion

Science moves forward through converging lines of evidence. Table 1 summarises the different lines that have been taken in the examination of the description-experience "gap". In the current set of experiments we add two additional lines: In Experiment 1, we conditionalised on the subset of data where the experienced distribution approximately matched the objective distribution. In spite of the methodological difficulties associated with eliminating external sampling bias while preserving sampling freedom, we did not observe a reliable choice gap. In Experiment 2, we again controlled for sampling biases by permitting participants to choose the number of perfectly representative blocks of trials to sample. Again, we did not observe a reliable choice gap. The lines of evidence, therefore, are converging on the conclusion that, in the context of pure exploration followed by a one-shot choice (i.e., the sampling paradigm), the gap between description and experience formats of choice is almost entirely due to external and internal sampling biases. It appears that people make equivalent choices when they use equivalent information to base their decision, regardless of presentation mode (e.g., Fox & Hadar, 2006; Rakow et al., 2008).

Our conclusion opposes the majority view drawn from previous studies (e.g., Hau, et al., 2010; see column 5 of Table 1)[9]. It is therefore important to highlight that

---

Experiment 1, the majority preference was for the option with the higher expected value in just two of the possible eight cases for *both* Description and Experience groups. This trend suggests that the most likely circumstances under which individuals choose options with the higher EV may be when they experience a sample that is perfectly representative of the population.

[9] Our conclusion also appears to conflict with the results of other work from our lab, where we found that "*the choice gap ... remained even when accounting for .. judgment distortion and the effects of [external] sampling bias*" (Camilleri & Newell, 2009, p. 518). The experiment in that paper looked at the role of representation in choice, and required participants to make a probability estimate for each outcome. This additional task contributed to participants sampling considerably more than in the current free sampling paradigm used in Experiment 1 (median = 26 vs. 10, respectively). Interestingly, participants' judgments did not predict their choice. One interpretation of this finding is that additional samples were taken to construct an accurate probability judgment and were not used as the basis for

our experiments represent the only attempt to directly target the influence of both external and internal sampling biases. Thus, the current set of experiments represents possibly the fairest comparison between experience and description choice formats to date. The result of this fair test is that the belief in a profound difference in choice preferences between description- and experience-based choice may be overstated.

Relation to Other Experience-based Choice Paradigms

We are careful to limit our conclusion to experience choice tasks in which purely explorative sampling is followed by a one-shot choice (i.e., sampling paradigm). There are other "experience" choice tasks that produce a Description-Experience gap that does not appear to be entirely explicable in terms of external and internal sampling biases. For example, in the "feedback" paradigm a large number of repeated, consequential choices are made between options (e.g., Barron & Erev, 2003). Therefore, participants are faced with a tension between exploring the options and exploiting the one they believe to be most favourable. Although there seems to be a correspondence in the preferences observed between the feedback and the sampling paradigms (Erev et al., 2010), these preferences do not appear to be driven by identical factors (Camilleri & Newell, 2011). Future studies must look to examine the range and relative contribution of these factors, over and above external and internal sampling biases. This contrast does, however, highlight how in moving forward we must abandon the propensity to simply label choice as either "description" or "experience" (cf. Hau et al., 2010; Rakow & Newell, 2010).

choice. In other words, although large samples reduced external sampling bias, it is possible that the observed choice gap was driven by a large amount of internal sampling bias.

Implications for Models of Risky Choice

The implication of our conclusion is that established description models of risky choice may be sufficient to account for the outcome of experience-based choices (with modification to account for sampling biases; e.g., Fox & Hadar, 2006). Nevertheless, established description-based models will never be able to provide a complete understanding of the processes underlying experience-based choices because they lack "modules" for cognitions unique to experience-based choice, including search and stopping strategies in sampling, and the roles of memory and learning.

Indeed, there is gathering evidence that decision-makers use a number of different strategies when making an experience-based choice. For example, it has been observed that sampling strategy has an impact on choice: Decision-makers who switch between options relatively often tend to prefer options that do better most of the time. In contrast, decision-makers who switch between options relatively seldom tend to prefer options that do better in the long run (Hills & Hertwig, 2010). Moreover, a recent model prediction competition declared the "ensemble model" winner of the experience (sampling) paradigm competition (Erev, et al., 2010). The ensemble model is interesting in that it inherently accounts for different choice strategies by assuming that each choice is made based on one of four equally likely rules (two versions of the natural mean heuristic, Stochastic Cumulative Prospect Theory, and a stochastic version of the Priority Heuristic).

As a preliminary exploration into the variety of search policies used in experience-based choice, we asked participants in Experiment 2 to write down in a free-response format the strategy or strategies that they used to make choices during the task. In general, participants produced fairly detailed explanations (mean

response length = 66 words). Examining these responses reveals a large variety of identifiable strategies (see Table 4). The most commonly reported strategy was one consistent with the natural mean heuristic, which simply tallies up the outcomes for each option and selects the option with the highest mean value (Hertwig & Pleskac, 2008). Other responses were consistent with various other strategies including risk aversion, risk seeking, Prospect Theory (Kahneman & Tversky, 1979), and an amended version of the Priority Heuristic in which the participant first compares the probability of the minimum outcomes and then proceeds to compare the magnitude of the outcomes (Brandstatter, Gigerenzer, & Hertwig, 2006; Erev, et al., 2010). Some participants reported using multiple strategies, both simultaneously and consecutively as the experiment progressed.

Our very preliminary excursion into the recounted strategy employed by our participants suggests, in line with the ensemble model, that multiple rules can be engaged depending on the specific strategy adopted by the decision-maker. It may be the case that different presentation formats encourage different strategies or rules to be preferred. Consistent with this hypothesis is the finding that experience-based choices can be made much more similar to description-based choices by explicitly presenting the possible outcomes (Erev, Glozman, & Hertwig, 2008). Presentation of possible outcomes may cause rules typically engaged by description-based choice to become more preferred in the experience format (Hertwig & Erev, 2009). One way for future work to examine this issue in more detail would be to use more sophisticated techniques (than free report) for investigating participants' strategies in both the sampling (e.g., Hills & Hertwig, 2010) and description (e.g., Johnson, Schulte-Mecklenbeck, & Willemsen, 2008) paradigms.

Table 4

Examples of Different Choice Strategies Reported by Participants in Free Responses made during Experiment 2

Strategy type	Example response
Natural mean heuristic (Hertwig & Pleskac, 2008)	*"I added the values as the game went along, and whichever had [the] better value (most positive, least negative) was the one I chose".*
Risk aversion	*"I generally took the option that would most likely give me a payout, even if it was small".*
Risk seeking	*"If the amount of points offered was above 10, I decided to choose that box regardless of its limited probability of paying out"*
Prospect Theory (Kahneman & Tversky, 1979)	*"If both choices were positive then I would go for the one with the highest probability of occurring. E.g. if 4 is certain, then I would go for that one rather than the other option of (32 or 0) where 0 had a higher probability of occurring. If the any of the choices were negative, then I would choose the one where 0 was more likely to occur".*
Amended Priority Heuristic (Brandstatter, et al., 2006; Erev, et al., 2010)	*If the machine had a value that appeared less than 2/5 times, then I would select the one that had a fixed value 100% of the time. However if the fixed value was low compared to the potential value that could have been obtained from the other, then I would have selected the other.*
Unique responses	*"I counted the number of times a number appeared on one machine, before a 0 appeared. If, say, a 3 appeared 5 times before a 0 on one machine and a 2 appeared every time on the other machine, I compared the totals (3x5 =15 compared to 2x6=12). I then picked the higher number".*
Multiple simultaneous strategies	*"Most of the time I counted which of the two slots would give me the most, or lose the least and selected the most ideal one. A couple of times I chose the one which displayed mainly '0's just in case my assumption (that a '60' or something will only appear once) was wrong".*
Multiple consecutive strategies	*"To begin with, I figured it was better to go with the lower, but more consistent pay out machine, but after a while I began to calculate the points in my head and, (if my maths is correct) the higher, but inconsistent payouts were better in total".*

Theoretical and Practical Implications

Our results appear to present a clear challenge to the claim that people make different choices when equivalent information about small monetary gambles are presented via description or (non-consequential) samples in highly controlled laboratory settings. However, the more general issue of how experienced and described information affects decision-making remains an issue of major theoretical and applied significance. Indeed, many of the decisions we make outside the lab are based on observation and feedback from experience and in this less contrived environment sampling biases remain a fact of life. For example, March and Shapira (1987) reported in a number of discussions with business managers that "possible outcomes with very low probabilities seem to be ignored, regardless of their potential significance … [which has] the effect of leaving organizations persistently surprised by, and unprepared for, realized events that had, a priori, very low probabilities" (p. 1411). Other examples that have been discussed in light of the description-experience gap include the formation of social impressions (Denrell, 2005), tourist responses to terrorist attacks (Yechiam, Barron, & Erev, 2005), the use of safety devices (Yechiam, Erev, & Barron, 2006), the heeding of safety warnings (Barron, Leider, & Stack, 2008), and doctor-patient interactions (Li, Rakow, & Newell, 2009). These studies all highlight the practical significance of thinking in terms of the continuum of differences between description- and experience-based choices and provide fruitful departure points for future research (cf. Rakow & Newell, 2010).

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 89-195). New York: Academic Press.

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.

Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making, 4*, 447-460.

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

Camilleri, A. R., & Newell, B. R. (2009). The role of representation in experience-based choice. *Judgment and Decision Making, 4*, 518-529

Camilleri, A. R., & Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*, 377–384.

Dallah, D. (2008). *Speeding Ticket Facts.* Retrieved 23 December 2009, from http://www.trafficticketsecrets.com/speeding-ticket-facts.html

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112*, 951-978.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior *Research Methods Instruments & Computers, 28*, 1-11.

Erev, I., Ert, E., Roth, A. E., Haruvy, E. E., Herzog, S., Hau, R., et al. (2010). A choice prediction competition, for choices from experience and from description. *Journal of Behavioral Decision Making, 23*, 15-47.

Erev, I., Glozman, I., & Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk and Uncertainty, 36*, 153-177.

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.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality & Social Psychology, 65*, 45-55.

Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *Quarterly Journal of Economics, 30*, 605-639.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making, 4*, 317-325.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23*, 48 - 68.

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.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*, 517-523.

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Charter & M. Oaksford (Eds*.), The Probabilistic Mind: Prospects for Rational Models of Cognition* (pp. 209-236). Oxford, England: Oxford University Press.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*, 225-237.

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science, 21*, 1787–1792.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstatter, Gigerenzer, and Hertwig (2006). *Psychological Review, 115*, 263–273.

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

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263-269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*, 397-402.

Li, S. Y. W., Rakow, T. R., & Newell, B. R. (2009). Personal experience in doctor and patient decision making: from psychology to medicine. *Journal of Evaluation in Clinical Practice, 15*, 993-995.

March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management Science, 33*, 1404-1418.

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.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*, 1-14.

RTA. (2007). *Road Traffic Crashes in New South Wales*. Retrieved 23 December 2009, from http://www.rta.nsw.gov.au/roadsafety/downloads/accident_ statistics_dl4.html

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1-26.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted, when rare outcomes are experienced (rarely)? *Psychological Science, 20*, 473-479.

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.

Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks. *Journal of Conflict Resolution, 49*, 430-439.

Yechiam, E., Erev, I., & Barron, G. (2006). The effect of experience on using a safety device. *Safety Science, 44*, 515-522.

Chapter 4: The Role of Repeated Choices

When and Why Rare Events are Underweighted: A Direct Comparison of the Sampling, Partial Feedback, Full Feedback and Description Choice Paradigms

Adrian R. Camilleri[1]
Ben R. Newell[1]

[1] School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Two paradigms are commonly used to examine risky choice based on experiential sampling. The feedback paradigm involves a large number of repeated, consequential choices with feedback about the chosen (partial-feedback) or chosen and foregone (full-feedback) payoffs. The sampling paradigm invites cost-free samples before a single consequential choice. Despite procedural differences, choices in both experience-based paradigms suggest underweighting of rare events relative to their objective probability. This contrasts with the overweighting when choice options are described, thereby leading to a 'gap' between experience and description-based choice. Behavioural data and model-based analysis from an experiment comparing choices from description, sampling, and partial- and full-feedback paradigms replicated the 'gap', but also indicated significant differences between feedback and sampling paradigms. Our results suggest that mere sequential experience of outcomes is insufficient to produce reliable underweighting. We discuss when and why underweighting occurs, and implicate repeated, consequential choice as the critical factor.

When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms

Many decisions can be based on analogous information acquired in different formats. For example, the choice to bring an umbrella to work may depend on the weatherman's description or our own experience with similar looking skies. Recently there has been a shift in the decision-making literature away from the study of completely described choice problems to decisions based on accumulated experience. Intriguingly, these two modes of presentation lead to strikingly different patterns of choice. For example, Barron and Erev (2003) found that people presented with experience-based choices tended to behave as if they discounted, or *underweighted*, rare events relative to their objective probability. In stark contrast, those presented with description-based choices behaved as if they fixated on, or *overweighted*, rare events relative to their objective probability. This different pattern of choice as a function of presentation mode is termed the description-experience "gap" and has caused a flurry of investigation (Rakow & Newell, 2010). Central to the investigation are three different experience-based choice paradigms.

Paradigms for Investigating Experience-based Choice

In the "partial" feedback paradigm, participants choose between options for a specified, typically large, number of trials (e.g., 400 trials; Barron & Erev, 2003). Each trial is associated with feedback and financial consequence for the selected option. The "full" feedback version of the paradigm also presents feedback for forgone alternatives (e.g., Yechiam & Busemeyer, 2006). In the sampling paradigm, there is a distinct sampling phase and a choice phase (e.g., Hertwig, Barron, Weber, & Erev, 2004). During the sampling phase, the decision maker explores the options

without financial consequence. Samples are typically small, ranging from between a median of 11 and 33 (Hau, Pleskac, & Hertwig, 2010). At any time the decision-maker can quit exploration and move on to make a one-shot, financially consequential choice.

Are All Experience-based Choices the Same?

Despite a number of procedural differences, it appears that sampling and the two feedback paradigms produce largely equivalent patterns of choice. Hertwig et al. (2004) found a correlation of .93 between the choices made in the six problems they examined with the sampling paradigm and the same six problems that Barron and Erev (2003) had examined with the partial feedback paradigm. Likewise Erev et al. (2010) found a correlation of approximately .8 for 120 problems presented in both paradigms. The implication of these results is that similar mechanisms underlie choice behaviour in the three paradigms, namely, sequential, "direct experience of outcomes and their likelihoods – and not repeated [consequential] choices" (Hertwig et al., 2004, pg. 537).

Two recent reviews have appraised a number of potentially relevant factors that contribute to underweighting in experience-based choice (Hertwig & Erev, 2009; Rakow & Newell, 2010). Implicit in this discussion is that not all factors are common to the paradigms and thus they are "not redundant" (Hertwig & Erev, 2009, pg. 521). However there is little discussion about how and why the *differences* between paradigms can affect patterns of choice. Here, we clarify when and why different forms of experiential choice diverge by contrasting described risky choices with those based on experience in sampling, partial- and full-feedback paradigms.

Our primary interest was in comparing the three experiential paradigms in order to test the hypothesis that sequential experience to outcomes is sufficient to produce

reliable underweighting (Hertwig, et al., 2004) and whether, if observed, it occurs for the same reasons. Before presenting the experiment, we briefly review some of the mechanisms that appear to contribute *differentially* to underweighting in the sampling and feedback paradigms.

When and Why Underweighting?

Memory Order Effects. In the memory and belief updating literature, there is evidence that the order in which information is presented can influence how that information is weighted (Atkinson & Shiffrin, 1968; Hogarth & Einhorn, 1992). A common effect is recency: the tendency to rely more heavily on recently observed outcomes. Reliance on such functionally smaller samples, which often underrepresent rare events, can cause underweighting. Recency has been implicated primarily in the sampling paradigm, for example, Hertwig et al. (2004) found that the second half of sampled outcomes predicted choices better than the first half. However recency effects have not been found consistently (e.g., Hau, Pleskac, Kiefer, & Hertwig, 2008) and, thus, its importance as a contributory factor to underweighting remains open to question.

Sampling Bias. Samples of information can often be unrepresentative of the underlying outcome distribution. Indeed, Hertwig et al. (2004) found that such sampling bias was a primary driver of the description-experience gap in the sampling paradigm: 78% of participants chose at a point where the rare event had been observed less often than expected based on the objective probability. As a result of frugal search efforts, many had never even seen the rare event. Hertwig and Pleskac (2010) showed that small samples are more likely to produce biased samples that under-represent the rare event, and subsequently lead to choice behaviour that appears to underweight those rare events. Currently, debate continues as to whether

biased samples are the primary (or even sole) cause of underweighting in the sampling paradigm (Camilleri & Newell, 2009, 2010; Fox & Hadar, 2006; Hadar & Fox, 2009; Hau, et al., 2010; Hau, et al., 2008; Rakow, Demes, & Newell, 2008; Rakow & Newell, 2010; Ungemach, Chater, & Stewart, 2009).

Sampling bias, however, is believed to be largely irrelevant in the feedback paradigms because large numbers of trials ensure that the rare event is seen, particularly when feedback for foregone alternatives is also provided (Jessup, Bishara, & Busemeyer, 2008). Thus something beyond sampling bias must contribute to underweighting in the feedback paradigms (Hertwig & Erev, 2009).

Exploration-Exploitation Conflict. A potential mechanism, although only relevant to the partial feedback paradigm, is the conflict between the objectives of learning more about one's options ("explore") and also trying to maximise one's earnings across repeated consequential choices ("exploit"; Hertwig & Erev, 2009). However, most designs confound repeated consequential choice and the exploration-exploitation conflict; one exception is Jessup et al. (2008), who presented participants with explicitly described options, thus eliminating the need for exploration, before having them make repeated decisions under partial feedback. Model fits suggested that choices were most consistent with objective probability weighting. These results seem to implicate the exploration-exploitation conflict as a cause of underweighting. However, because participants' were given complete descriptions of the alternatives prior to choice, learning processes were confounded with initial tendencies (Erev & Haruvy, 2005). That is, descriptions prompt overweighting (Kahneman & Tversky, 1979) and, hence, initial preferences in the Jessup et al. study were atypical. Thus, the importance of the conflict as a cause of underweighting remains unresolved.

In summary, several factors appear to exert their influence in different ways across the three experienced-based choice paradigms. Moreover, the exact patterns of influence and the reasons for underweighting remain uncertain.

The Experiment

We compared choices in description, sampling, partial and full feedback paradigms to examine the robustness of and reasons for underweighting. To reduce the influence of information asymmetries due to sampling bias, participants in the experience groups made 100 samples. Thus any difference in preferences between the Sampling and two Feedback groups would demonstrate that these "experience-based" choice tasks are not equivalent and that previous observations of underweighting in each are likely to occur for different reasons.

Our secondary interest was in comparing the Partial Feedback and Full Feedback groups to examine the relative influence of repeated, consequential choice and the exploration-exploitation conflict to underweighting. The provision of feedback of foregone pay-offs eliminates the conflict in the Full Feedback group, yet preserves the task of making repeated, consequential choices. Thus any difference in the extent of underweighting between the groups would implicate the exploration-exploitation conflict as the more important factor. If there are no differences, then repeated, consequential choices would be left as the major reason for underweighting.

Method

Participants

One-hundred twenty undergraduate first year University of New South Wales psychology students (82 females; median age = 18; range 17-42) took part in exchange for course credit and payment contingent upon choices.

Materials

Decision Task. The decision task was a virtual money machine game. In the description-based paradigm, two alternative money machines were presented and labelled with the outcome payouts and their probabilities (e.g., 80% chance of 4, else 0). In all three experience-based paradigms the machines were unlabelled but were associated with a distribution of possible outcomes in accordance with the objective probabilities from which samples were randomly drawn. Allocation of safe and risky options to machines was counterbalanced and the order of the problems was randomized.

Choice Problems. Problems consisted of a risky option that probabilistically paid out a high or low outcome, and a safe option that always paid out a medium outcome (Table 1). There were two problems in the gain domain and two problems in the loss domain. This permitted examination of five distinct choice patterns: risk aversion, risk seeking, adherence to the expected value, underweighting of rare events, or overweighting of rare events.

Table 1

Problem Options and Possible Choice Strategies

Problem	Choice Options		Expected Choice Pattern Under Strategy				
	Safe	Risky	Risk aversion	Risk seeking	Adhere to EV	Underweight rare events	Overweight rare events
1	9(1.0)	10(.9)	Safe	Risky	-	Risky	Safe
2	-3(1.0)	-4(.8)	Safe	Risky	Safe	Safe	Risky
3	2(1.0)	14(.15)	Safe	Risky	Risky	Safe	Risky
4	-3(1.0)	-32(.1)	Safe	Risky	Safe	Risky	Safe

Design

A between-subjects design was used where the main independent variable was choice paradigm (Description, Sampling, Partial Feedback or Full Feedback)[1]. The Full-Feedback group was run subsequent to the three other groups. In the Description and Sampling groups, the dependent variable was the choice made (risky or safe). In the two Feedback groups, the dependent variable was the choice made on the last (i.e., 100th) trial[2].

Procedure

As shown in Figure 1, participants in the Description group viewed explicit descriptions of each machine's payoff and then selected their preferred machine. Participants in the three experience-based tasks were allocated 100 samples to use. In the Sampling group, the samples were purely for the purpose of exploration and feedback was limited to the selected option. After the 100 samples, the computer moved participants to the choice phase where they selected their preferred machine. In contrast, each of the 100 samples made by participants in the two feedback groups was a decision with financial consequence. An on-screen summary provided information about the last payoff and the cumulative points from all plays for the problem. Participants in the Partial Feedback group received outcome feedback only for the selected option. Participants in the Full Feedback group also received outcome feedback for the foregone, unselected option.

[1] We also had a second, instructional manipulation for participants in the Description and Sampling groups. The manipulation indicated that the selected machine would be played repeatedly 100 times and that the outcome received would be the average of those 100 plays. This manipulation had no effect in the Description ($\chi 2 = .107$, $p = .744$) or Sampling ($\chi 2 = .626$, $p = .429$) groups (possibly due to insufficient salience see, for example, Wedell & Böckenholt, 1990). As a result, we have collapsed across this manipulation.

[2] We elected to use a binary DV across all experimental groups. Importantly, the pattern of results is essentially the same regardless of whether a mean or modal DV was used in the feedback groups. Similarly, there was no difference when the DV was based on the last 1, last 50 or the entire sequence of 100 trials.

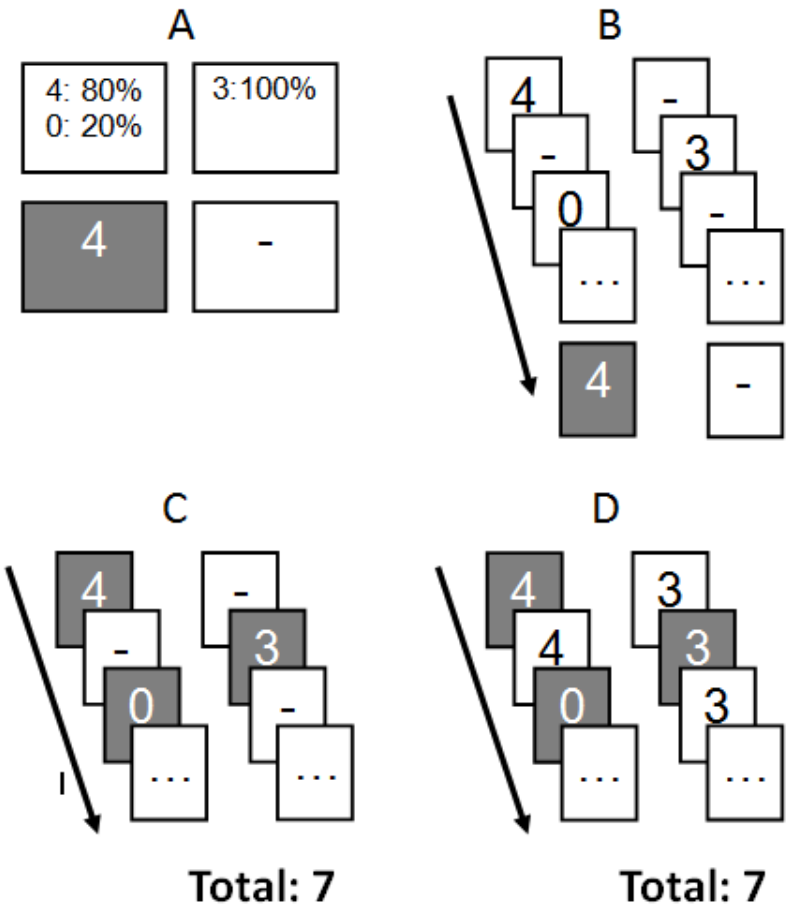


Figure 1. Depiction of the four different choice paradigms. Shaded rectangles represent consequential trials, that is, trials in which the outcome of the choice affected earnings. (A) Description group: explicitly stated outcomes and their probabilities followed by a one-shot choice. (B) Sampling group: initial sampling period of 100 trials followed by a one-shot choice. (C) Partial Feedback group: each of the 100 trials was consequential and the total earnings for the problem were always displayed. (D) Full Feedback group: identical to the Partial Feedback group with the addition of feedback for the foregone, or unselected, alternative.

Results and Discussion

Patterns of Choice

The percentage of participants who preferred the risky option for each of the four problems is displayed in Table 2. The most important comparisons in the current context are those between the Sampling and two Feedback groups. Although participants in each of these three groups sequentially played through 100 trials, that

is to say, were each making an "experience-based" choice, their final preferences were nevertheless very different.

Table 2

Percentage of Participants who Preferred the Risky Option

Choice Options		Percentage preferring the risky option			
Safe	Risky	Description (n = 40)	Sampling (n = 40)	Partial Feedback$^{\wedge}$ (n = 20)	Full Feedback$^{\wedge}$ (n = 20)
9(1.0)	10(.9)^{P}	15††‡	38‡	60	70
-3(1.0)^{P}	-4(.8)	58†‡	40†	15	20
2(1.0)^{P}	14(.15)	53†	38†	5‡	30
-3(1.0)	-32(.1)^{P}	45‡	48‡	65	80

^{P} Option predicted to be preferred if rare events are underweighted.

$^{\wedge}$ The DV was the choice made on the final (i.e., 100th) trial.

† Significantly different from Sampling group (χ^2<.05).

† Significantly different from Partial Feedback group (χ^2<.05).

‡ Significantly different from Full Feedback group (χ^2<.05).

To better gauge these differences, we re-mapped choices onto a single directional scale. Specifically, we focused on the option that would appear more attractive if rare events are underweighted – we called this the "predicted" option (see first two columns of Table 2). As shown in Figure 2, when averaging across problems and participants, there were significant differences in the number of predicted options preferred by those in the Sampling group and both the Partial Feedback (51.9% vs. 76.3%, χ^2 = 13.2, p < .001) and the Full Feedback groups (51.9% vs. 75%, χ^2 = 11.843, p < .01). When comparing the patterns of choice in Table 2 with the different choice strategies outlined in Table 1, we see that those in

the Sampling group displayed choice patterns most consistent with risk aversion3, whereas those in the two Feedback groups displayed choice patterns most consistent with underweighting. In fact, choices in the Partial and Full Feedback groups did not differ (76.3%, vs. 75%, χ^2 = .034, p >.1). These findings cast doubt on the assumption that the sampling and feedback tasks are equivalent paradigms of experience-based choice that produce similar choice preferences.

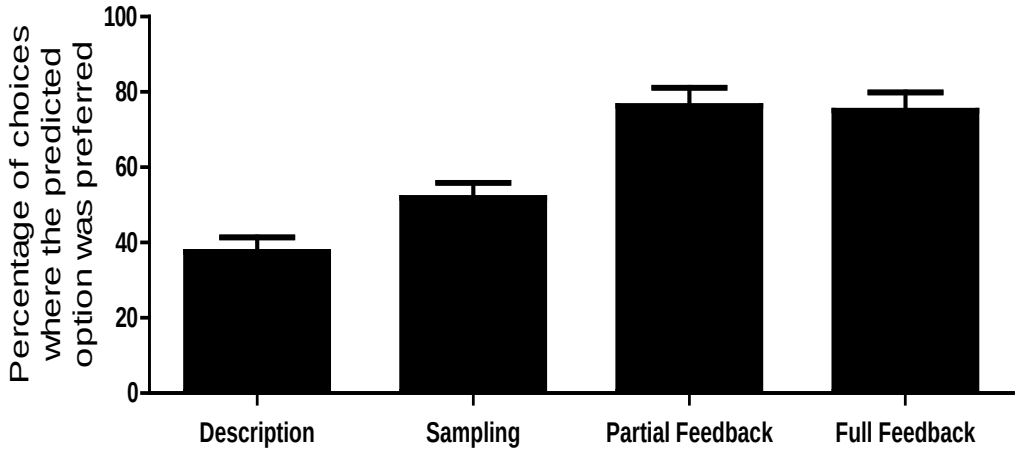


Figure 2. Percentage of participants who preferred the predicted option. The predicted option is the more attractive alternative when rare events are underweighted. Error bars indicate the SEM.

The data in Table 2 also show a Description-Experience "gap": the choice differences observed between the Description group and the three Experience groups are all in the expected direction assuming that rare outcomes receive less weight when experienced than when described. A comparison with Table 1 indicates that the Description group produces a choice pattern most consistent with overweighting. As shown in Figure 2, when averaging across problems, those in the Description group selected the predicted option less often than those in the Sampling (37.5% vs. 51.9%, χ^2 = 6.68, p < .05), Partial Feedback (37.5% vs. 76.3%, χ^2 = 32.04, p < .001) and Full Feedback groups (37.5% vs. 75%, χ^2 = 30.0, p < .001).

When Underweighting?

In order to obtain an improved understanding of when underweighting occurred, we fitted the data to Prospect Theory (PT; Kahneman & Tversky, 1979). PT is a highly successful model of description-based choice that operates by entering decision-makers' beliefs about outcomes and their probabilities into PT weighting and value functions to produce a weighted value for each option. PT has also been successfully applied to experience-based choice data (e.g., Hau et al., 2008; Ungemach et al., 2009). The probability weighting function contains a parameter whereby 1 indicates objective weighting of probabilities, <1 indicates overweighting and >1 indicates underweighting.

Rather than searching for the "best" fitting parameter, which can be problematic due to potential flat maxima and the two weighting functions trading off against one another, we tested the performance of PT across a broad range of parameter values (between 0 and 2 for both functions, in steps of .01). Following Erev et al. (2010), parameters were estimated across all choices and problems under the assumption of gain-loss symmetry (i.e., $\alpha=\beta$ and $\gamma=\delta$; see supplementary materials for more detail). Veridically experienced, rather than objective, probabilities were used when fitting the data.

The contour plots in Figure 3 show the proportion of correct predictions made by PT as a function of the 40,000 different value- and probability weighting-function parameter combinations. We constructed a scale to include 20 "bands", each .025 wide and starting from the lower limit. The regions with the best fit are represented by the darkest shading. As can be seen by the varied shading, some parameter combinations were more successful than others. The grey box behind the scale indicates the range of proportion of correct predictions. For example, in the Partial

Feedback Group (lower left panel), the highest proportion of correct predictions achieved was .75 and the lowest was .36.

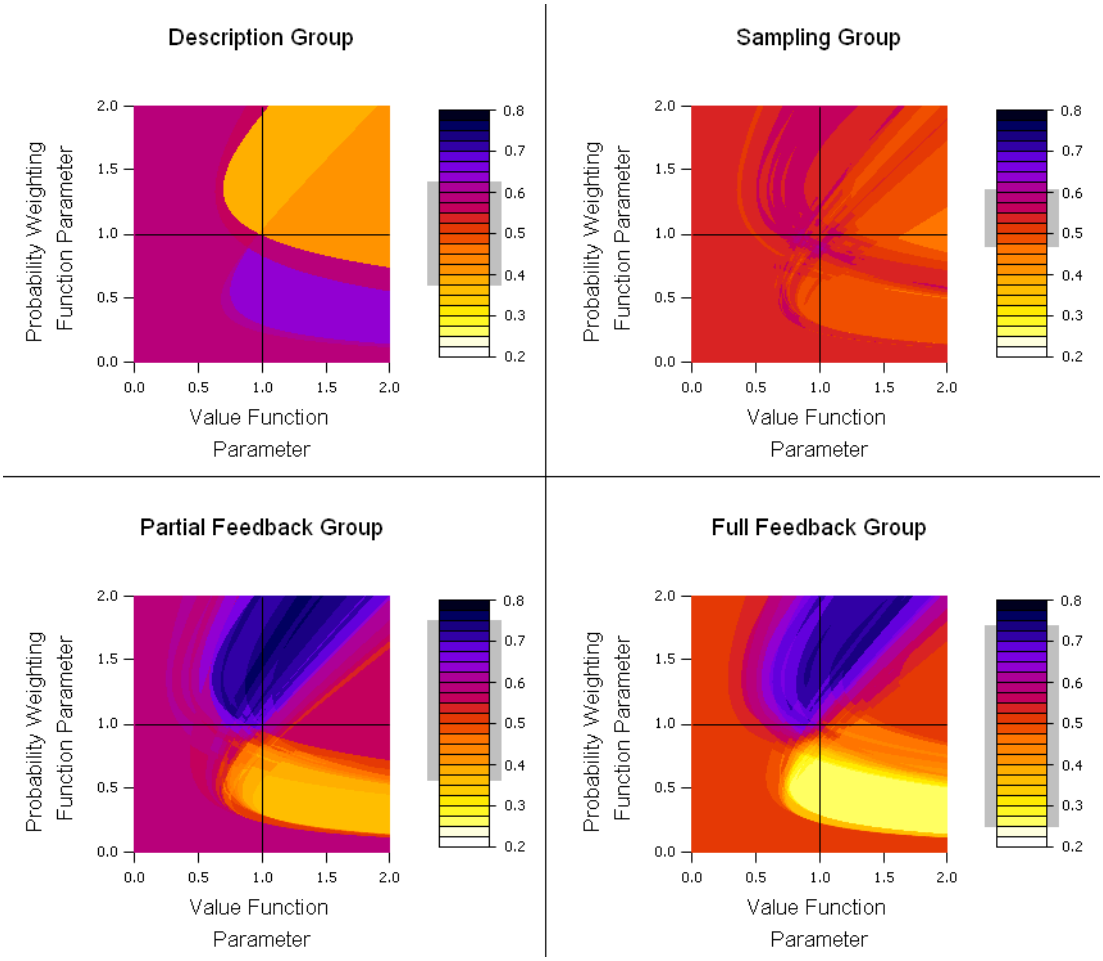


Figure 3. Contour plots showing the proportion of correct predictions when the data from the Description, Sampling, Partial Feedback and Full Feedback groups were fitted to PT. The proportion of correct predictions was calculated for each combination of value- and weighting-function parameters between 0 and 2, in steps of .01 based on veridically experienced probabilities. The regions with the darkest shading indicate the combinations providing the highest fit. The problem frame was ignored by assuming gain-loss symmetry (i.e., $\alpha=\beta$ and $\gamma=\delta$).

The regions with the best fit for the Description group are clearly for probability weighting function parameters below 1, implying overweighting of small probabilities. There were no clear distinct regions of best fit for the Sampling group; PT did equally well with probability weighting function parameters below, near and above 1. This finding mirrors that observed by Ungemach et al. (2009), who used a similar method and found that a "bias free" sampling group produced similar degrees

of fit across a wide range of weighting function parameters. In contrast, the regions with the best fit for the Partial and Full Feedback groups are very similar and are clearly produced by probability weighting function parameters above 1, implying underweighting of small probabilities. These findings thus cast doubt over the assertion that sequential experience of outcomes, the defining characteristic of "experience-based" choice, is sufficient to produce reliable underweighting.

The remaining difference in choices between the Sampling and Description groups seems to be due to a recency effect: choices made by those in the Sampling group were better predicted by outcomes observed in the last ten samples than the first ten samples (63.1% vs. 49.4%, $\chi^2 = 6.14$, $p < .05$). This effect, however, was not apparent when examined with the most commonly used method in the literature: comparing the relative predictive success of the first versus second half of the observations (e.g., Hertwig et al., 2004). This discrepancy may explain previous inconsistencies in the observation of recency effects between studies that report frugal sampling efforts and find recency (e.g., Hertwig, et al., 2004; Rakow, et al., 2008) and studies that report more extensive sampling efforts but do not find recency (e.g., Camilleri & Newell, 2009; Hau, et al., 2008; Ungemach, et al., 2009).

Why Underweighting only in the Feedback Groups?

When participants are allowed to sample freely (and hence frugally) clear underweighting is observed in model analyses like those reported above (Ungemach et al., 2009). However, the unrepresentative samples that arise from such free sampling are unlikely to have contributed to the disparity we observed between the Sampling and Feedback groups: the median difference between experienced and objective outcome distributions for those in the Sampling, Partial Feedback, and Full Feedback groups was just 2.2%, 3.0% and 2.0%, respectively.

Instead, the presence of repeated and consequential choice, unique to the two Feedback groups, appears to be the crucial element for underweighting to occur in the absence of sampling bias. The similar pattern of choices in the Partial and Full Feedback groups reinforces the key role for repeated, consequential choice and not the exploration-exploitation conflict given the absence of this tension in the Full Feedback group. One specific aspect of repeated, consequential choice that might lead to underweighting is that the feedback paradigms encourage choice inertia, that is, the tendency to repeat the last choice, irrespective of the obtained outcome (Erev & Haruvy, 2005). For example, it has been shown that the tendency to select the risky option is at least partly a function of how many times the risky option has been chosen on previous occasions (Barron, Leider, & Stack, 2008).

Are There 'Redundancies' Across the Experiential Paradigms?

Figure 4 plots the mean number of risky choices made across all trials for the three experience groups. As expected, the Sampling group shows relative indifference between the options across trials because all samples are inconsequential. In contrast, and consistent with underweighting, the two feedback groups show a preference towards one particular option: when the rare event is good, as in problems 2 and 3, preference is for the safe option; when the rare event is bad, as in problems 1 and 4, preference is for the risky option (with the exception of the Partial Feedback group in problem 4; see the next paragraph for an explanation). Although final preferences in the two Feedback groups were the same (Table 2), it is clear from Figure 4 that the pattern of choices across the 100 trials is not identical. One possible reason for this is the "hot stove" effect.

The hot stove effect describes how good outcomes increase the probability of repeating a choice whereas bad outcomes decrease the probability (Denrell & March,

2001). Relevant only in the context of the exploitation-exploration conflict, the hot stove can lead to risk aversion because risky options are more likely to produce bad outcomes and subsequent avoidance (Erev, et al., 2010). In partial support of this hypothesis, those in the Partial Feedback group made significantly fewer risky choices in Problem 4 than those in the Full Feedback group ($t_{(1,38)}$=2.65, p<.05). Additionally, there was an interaction in Problem 3 such that those in the Partial Feedback group made more risky choices in the first block than the last block whereas those in the Full Feedback group showed no difference ($F_{(1,38)}$=12.46, p<.01). Thus the 'hot stove' effect can contribute to the extent of underweighting by enhancing it when the rare event is good (e.g., Problem 3) but attenuating it when the rare event is bad (e.g., Problem 4; Fujikawa, 2009).

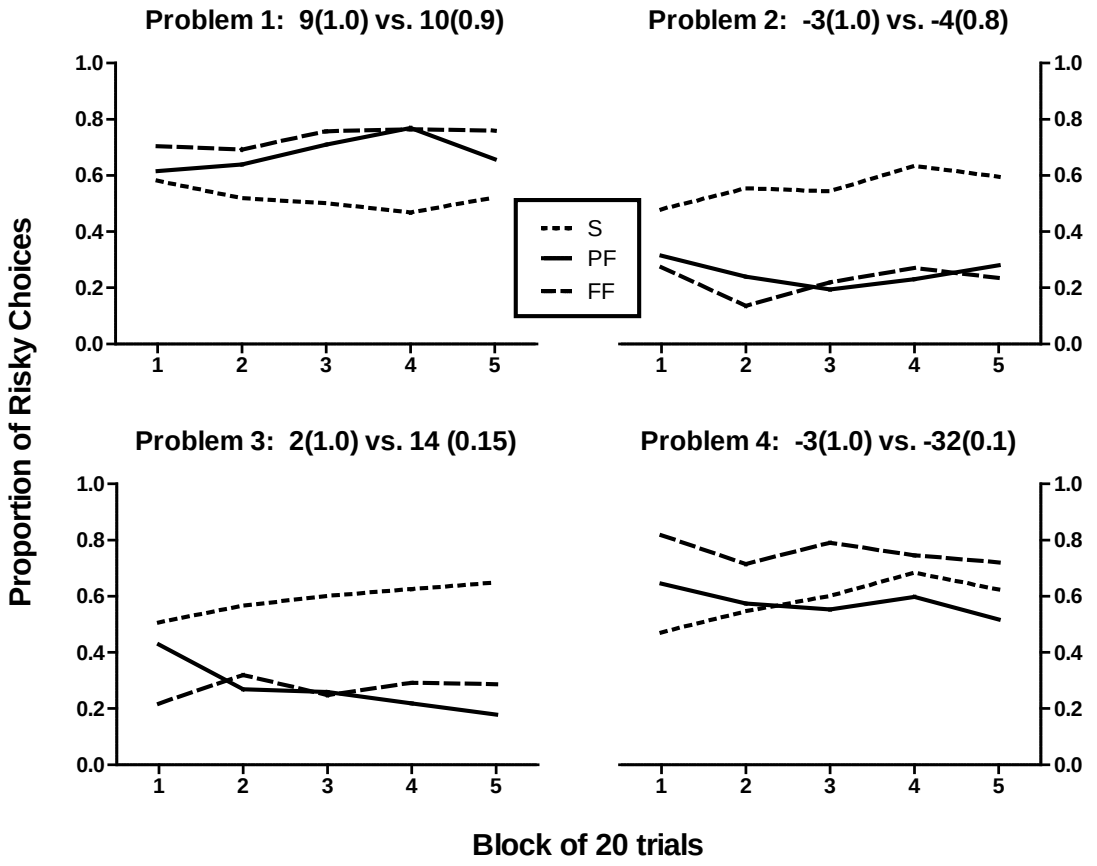


Figure 4. Proportion of risky choices in the Sampling (S; dotted line), Partial Feedback (PF; solid line) and Full Feedback (FF; dashed lines) groups for each problem in blocks of 20 trials.

These distinct patterns of choice across trials reinforce the notion that the two feedback paradigms are not 'redundant' (Hertwig & Erev, 2009). Although the exploration-exploitation conflict may not be the source of underweighting, (e.g., Hertwig et al., 2004) it can still influence choices through the hot stove effect, resulting in different patterns of responding in the two feedback paradigms.

Conclusion

The recent explosion of interest in comparing choices made from description and experience has been based on two methods for operationalising experience-based choice: sampling and feedback (e.g., Hertwig & Erev, 2009; Rakow & Newell, 2010). Despite assertions in the literature that choices under the two paradigms are highly consistent (Erev et al, 2010; Hertwig et al., 2004), very few studies have directly compared them in order to test this idea and also to examine whether similar mechanisms drive similar patterns of underweighting.

Our experiment demonstrates clearly that the paradigms differ in terms of (1) choices made (Figure 2), (2) the best-fitting probability weighting parameter values in PT (Figure 3), and (3) participants' sampling experience across trials (Figure 4). The pattern of results is consistent with clear underweighting of rare events relative to their objective probability in the two feedback paradigms, but not in the sampling paradigm. The remaining 'gap' between our Sampling 'bias-free' group and Description group appears to be due, in part, to a recency effect, which occurred in the absence of reliable underweighting. Debate about the size, robustness and cause of the sampling-description gap continues, and it is unlikely to have a single source (cf. Hadar & Fox, 2009; Hertwig & Erev, 2009; Rakow & Newell, 2010). Our chief concern, however, is not with this debate but with the assumption that

underweighting occurs in sampling and feedback paradigms for the same reasons (e.g., Hertwig et al., 2004). Our data speak clearly against this assumption.

The repeated, consequential choices unique to the feedback paradigm appear to be a much stronger driver of robust underweighting. Even in the absence of an exploitation-exploration conflict underweighting was observed, although when such conflict was present there was a tendency for choices to be biased away from risky options (i.e., the hot stove effect). These behaviours stand in contrast to choice overweighting when decision alternatives were explicitly described.

Our conclusions highlight that dichotomising choices as "experience-based" and "description-based" is too simplistic (Hau et al., 2010; Rakow & Newell, 2010); different kinds of experience can lead to very different patterns of choice and for different reasons. Accounting for these patterns should be core to the development of new theories and computational models of experience and description-based choice (cf. Erev et al., 2010).

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 89-195). New York: Academic Press.

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.

Camilleri, A. R., & Newell, B. R. (2009). *The role of representation in experience-based choice. Judgment and Decision Making, 4,* 518-529.

Camilleri, A. R., & Newell, B. R. (2011). Description- and experience-based choice: Does equivalent information equal equivalent choice? Acta Psychologica, 136, 276–284.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction - The hot stove effect. *Organization Science, 12*, 523–538.

Erev, I., Ert, E., Roth, A. E., Haruvy, E. E., Herzog, S., Hau, R., et al. (2010). A choice prediction competition, for choices from experience and from description. *Journal of Behavioral Decision Making, 23,* 15-47.

Erev, I., & Haruvy, E. (2005). Generality, repetition, and the role of descriptive learning models. *Journal of Mathematical Psychology*, *49*, 357–371.

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.

Fujikawa, T. (2009). On the relative importance of the hot stove effect and the tendency to rely on small samples. *Judgment and Decision Making, 4*, 429-435.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making, 4*, 317-325.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23*, 48-68.

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.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*, 517-523.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*, 225-237.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1-55.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from Prospect Theory in descriptive choice. *Psychological Science, 19*, 1015-1022.

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

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.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*, 1-14.

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.

Wedell, D. H., & Böckenholt, U. (1990). Moderation of preference reversals in the long run. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 429-438.

Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making, 19*, 1-16.

Supplementary Materials

Prospect Theory calculates the weighted value of each option and then chooses the most attractive alternative (Tversky & Kahneman, 1992). The expected value of each outcome, *j*, is given by:

$$E_j = w(p_j)\, v(x_j)$$

where $w(p_j)$ represents a weighting function for the outcome probability and $v(x_j)$ represents a weighting function for the outcome value. The probability weighting function $w(p_j)$ is given by:

$$w(p_j) = \begin{cases} \dfrac{p_j^{\gamma}}{\left(p_j^{\gamma} + (1 - p_j)^{\gamma}\right)^{1/\gamma}}, & if\ x \geq 0 \\ \dfrac{p_j^{\delta}}{\left(p_j^{\delta} + (1 - p_j)^{\delta}\right)^{1/\delta}}, & if\ x < 0 \end{cases}$$

The δ and γ are adjustable parameters that fit the shape of the function for gains and losses, respectively. Parameters below 1 overweight small probabilities and underweight large probabilities whereas parameters above 1 do the opposite. The value function $v(p_j)$ is given by:

$$v(x_j) = \begin{cases} x_j^{\alpha}, & if\ x_j \geq 0 \\ -\lambda\left(|x_j|^{\beta}\right), & if\ x_j < 0 \end{cases}$$

The α and β are adjustable parameters that fit the curvature for the gain and loss domain, respectively. The λ parameter (λ>1) scales loss aversion but is only relevant in mixed gambles and was therefore set to l in our analysis.

The Long and Short of It: Closing the Description-Experience "Gap" by Taking the Long Run View

Adrian R. Camilleri[1]
Ben R. Newell[1]

[1] School of Psychology, University of New South Wales, Sydney, Australia

Abstract

The demonstration that preferences differ when gambles are to be played once or multiple times is central to a long standing debate about the rationality of risky choice. We tested a novel prediction that choices made under single- and multiple-play conditions would be affected not only by imagined future prospects but also by the acquisition method of choice-relevant information (description vs. experience). In an experiment participants either read described gambles or made repeated choices in environments with probabilities and outcomes matched to those in the descriptions. Preferences for single- and multi-play choices of the described/experienced gambles were then elicited. Under single-play conditions different preferences following description and experience were observed, but this "gap" was almost closed under multi-play conditions. We conclude that biases caused by short horizon single-play frames – e.g., overweighting described rare events and over-reliance on recent samples of experience – are reduced in longer horizon multiple-play frames.

The long and short of it: Closing the description-experience "gap" by taking the long run view

Consider the following gamble: you have a 50% chance of winning $200 and a 50% chance of losing $100. Would you take the bet? How about if the gamble were played 100 times? Would you change your mind?

If you are like Nobel Prize winning economist Paul Samuelson's colleague then you'd refuse the single play bet but accept the 100 play option. According to the colleague, with 100 plays there is a virtual assurance "to come out ahead" (Samuelson, 1963, pp. 109). While it is true that if the bet were played 100 times the probability of losing is less than 1 in 2000, as Samuelson points out, you cannot ignore the possibility of losing $10,000. So what should you do?

This simple example ignited a long-standing debate over whether it is rational to behave differently to a single bet than to repeated play of the same bet (Lopes, 1981, 1996; Samuelson, 1963; Tversky & Bar-Hillel, 1983). Samuelson showed that it was irrational under expected utility theory to accept the repeated bet when the single bet was rejected under the wealth bands encompassed by the wager (i.e., -$10,000 to +$20,000). However, others have argued that it is entirely sensible to adhere to such behaviour if choice is based on achieving a certain aspiration; in such cases, choice may reasonably be based on the probability of coming out ahead (Lopes, 1981; Lopes & Oden, 1999).

Many important decisions require comparing long and short run horizons (e.g., retirement savings, healthy eating, and carbon tax legislation) and thus understanding how and why choices differ across time perspectives is crucial. We test the novel prediction that decisions about long and short run time horizons are influenced not

only by future prospects, but also by differences in the history of acquiring choice-relevant information.

Description, Experience, and "Rare" Events

Description-based choices are those in which the outcomes and their probabilities are provided in summary description form. For example, in making a decision about your retirement savings you might examine tables of data describing the performance of an investment strategy in terms of its returns. In contrast, experience-based choices are those in which the outcomes and their probabilities are initially unknown and must be inferred from samples of experience. For example, you might rely on your own remembered experience of previous returns delivered by different strategies.

An established body of research demonstrates that choices made on the basis of these two different types of information often systematically diverge (Gottlieb, Weiss, & Chapman, 2007; Hertwig, Barron, Weber, & Erev, 2004; Hertwig & Erev, 2009; Jessup, Bishara, & Busemeyer, 2008; Ungemach, Chater, & Stewart, 2009). A key finding is that when participants are given *descriptions* of problems they behave as if they *overweight* the probability of a rare[1] event occurring. In contrast, participants who learn about outcomes and probabilities via *experiential sampling* choose as if they *underweight* the probability of a rare event occurring.

Experience-based choice can be conceptualised as an "imagined" multi-play problem made real. Rather than having to consider the long run perspective and mentally simulate outcomes, the decision-maker is confronted with the outcomes across trials. How does this real experience impact decisions about future prospects?

[1] By convention, we defined a "rare event" as one that occurs 20% of the time, or less (Hertwig, et al., 2004).

To examine this question we compared single- and multi-play preferences for participants who had either read descriptions of problems, or had completed 40 trials of experience in choice environments with probabilities and outcomes matched to those provided in the descriptions. Single-play choice was operationalized in the description task as the one-shot choice made in a 2-alternative problem and in the experience task as the choice made on the final (i.e., 40th) trial. Multi-play choice was operationalized in both tasks as the allocation of 100 plays between alternatives made subsequent to the single play.

Hypotheses and Predictions

The standard finding in description-based choice is that participants are more likely to prefer the maximising option – that is, the option with the higher expected value (EV) – when choices are presented in the multi-play format. In other words, multi-play formats shift preferences more in line with normative predictions (see Wedell, 2011 for a review). However, a close analysis of the literature reveals somewhat different behaviour when "rare" events are considered. Almost every problem previously examined that contains a rare event has confounded the maximising option and the more uncertain option. In the one example where this confound did not occur, Wedell and Bockenholt (1990) presented participants with two problems containing a near sure option that had a higher EV than the alternative risky option (94% chance of 10 [EV=9.4] versus 19% chance of 42 [EV=8.0]). Surprisingly, the rate of maximisation actually decreased under multi-play conditions (i.e., more participants favoured the 19% chance of 42 in the multi-play).

This observation suggests an alternative interpretation of the data in contexts where a rare event is present. Rather than increasing the tendency to maximise, the multi-play format may instead reduce the tendency to overweight rare events, at least

in the description format. In problems like the one above where the safer option has the higher EV, a reduced tendency to overweight rare events produces less maximisation. Such a tendency would also produce preferences more consistent with those in the experience task (i.e., underweighting of rare events; Hertwig & Erev, 2009). However, this specific prediction cannot be tested with existing data because rare events have been present in both options when uncertainty and maximisation have been unconfounded (e.g., Wedell & Bockenholt, 1990).

We hypothesise that the multi-play format induces consideration of the long run and consequently minimises focus on the rare event in the description paradigm. Thus, in the description format, we expected that the multi-play preferences would be consistent with a reduced tendency to overweight rare events, regardless of EV. The left side of Figure 1 illustrates this hypothesis using one of the problems - Problem 2 – from our experiment. Problem 2 contrasts a safe 100% chance of 14 [EV = 14] with a risky 90% chance of 15 and 10% chance of 0 [EV = 13.5]. The rare event in this problem is 0. In the description format, we predict overweighting of the rare event, 0, which should lead to a preference for the safe option. If multi-play reduces overweighting of rare events, then we would expect the preference for the safe option to be weaker under multi-play conditions – despite the fact that it has the higher EV.

In contrast to the description paradigm, we have little precedence upon which to predict the impact of experience on choices in the multi-play question. A working hypothesis, however, focuses on the role of recency. In the experience task underweighting of rare events appears to be due to heavy reliance on recent outcomes (Lejarraga, Dutt, & Gonzalez, 2011), and elimination of this reliance can reduce underweighting of rare events (Camilleri & Newell, 2011a). We hypothesise

that the multi-play format induces consideration of the long run and consequently minimises reliance on recent outcomes in the experience paradigm. Thus, in the experience format, we expected that the multi-play preferences, which are made after 40 trials' worth of experience, would be consistent with a reduced tendency to underweight rare events, again regardless of EV. The right side of Figure 1 illustrates this hypothesis using Problem 2. We predict underweighting of the rare event, 0, which should lead to a preference for the risky option. If multi-play reduces underweighting of rare events, then we would expect the preference for the risky option to be weaker under multi-play conditions.

To summarise, our main experimental prediction was for an interaction effect such that the description-experience "gap" observed when people take the short-run perspective under single-play will be reduced or eliminated when long-run future prospects are considered under multi-play.

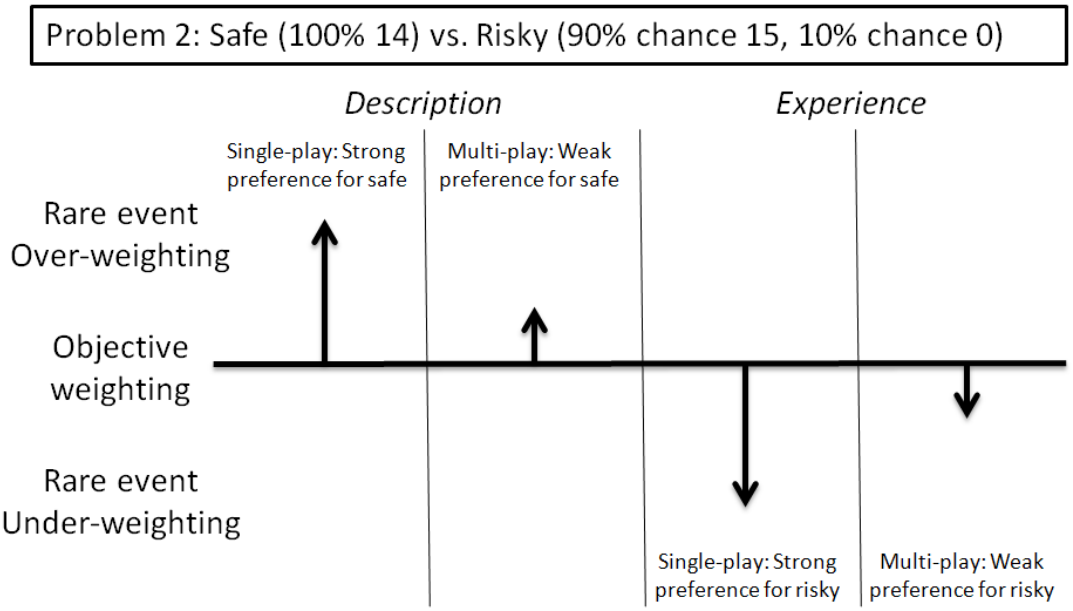


Figure 1. Experimental predictions with Problem 2 as an example.

One final important feature of our experiment concerns the elicitation of preferences in the multi-play question. The vast majority of previous studies have presented participants with a binary choice between options. That is, even in the multi-play condition, just one option is selected and then played repeatedly. In everyday practice, however, there are many decisions that permit the decision-maker to distribute preference across the available options. Moreover, the standard design leaves participants with no way to indicate indifference between the options, which is difficult to justify in two-alternative problems that possess very similar expected values (Regenwetter, Dana, & Davis-Stober, 2011). Therefore, in our experiment, we gave participants the freedom to allocate 100 plays across the two possible options in any distribution, including indifference (e.g., Bristow, 2011; Thaler, Tversky, Kahneman, & Schwartz, 1997).

Method

Participants

The participants were 109 online workers recruited from the Amazon's Mechanical Turk website. The average age was 29.7 years and 70% were male. Participants received $0.30 in exchange for completing the experiment.

Materials

Decision Task. The decision task was a virtual money machine game with two alternative options. In the description-based paradigm, the options were presented and labelled with the outcomes and their probabilities (e.g., "80% chance of 4, otherwise 0"). In the experience-based paradigm, the options were presented with only single-letter labels (e.g., "A"); however, each option was associated with a distribution of outcomes in accordance with the outcome distribution shown to those

playing the description version of the task. Participants were afforded 40 samples, during which time they were presented with a series of outcomes that were randomly arranged but together perfectly reflected the distribution. The outcome of the foregone alternative was also presented. This "full feedback" paradigm was selected in order to avoid the influence of the exploration-exploitation tension such as the hot stove effect (Camilleri & Newell, 2011b).

Choice Problems. The problems consisted of a risky option that probabilistically paid out a high or low outcome, and a safe option that always paid out a medium outcome. The four problems crossed choice domain (gains vs. losses) with the EV of the risky option (higher or lower; see Figure 1). There was also a fifth problem that served as a "catch" for participants that were not paying attention. The catch problem posed a choice between a sure 12 and a sure 11. Allocation of safe and risky options to left and right of screen was counterbalanced and problem order randomized.

Design

The study used a 2 x (2) x (4) mixed design. The between-subjects variable was the choice paradigm (Description vs. Experience). The within-subjects variables were number of plays (x1 vs. x100) and the problem types (Gain vs. Loss crossed with Risky option associated with higher vs. lower EV). The dependent variables were participant's single- and multi-play preferences. In the Description group, the single-play preference corresponded to the choice made (risky or safe). In the Experience group, the single-play preference corresponded to the choice made on the final trial (risky or safe). The multi-play preference in all groups corresponded to the number of plays allocated to the safe and risky options.

Procedure

The entire experiment took on average 7 minutes to complete, which produced an effective average hourly rate of approximately $2.50. This payment is above the average effective hourly rate on Mechanical Turk (Horton & Chilton, 2010).

Participants in the Description group were presented with descriptions of each alternative's payoff distribution and were asked to select their preferred option. Subsequently, two slider bars appeared beneath the options, each ranging from 0 and 100, and with a small textbox displaying the current value of the slider. The default starting position was at 50. Participants were asked to imagine they had 100 plays at the problem they had just seen and to adjust the sliders to indicate how they wanted to distribute their plays between the options. The two sliders were linked and forced to equal to 100; thus, adjusting one slider automatically moved the other.

Participants in the Experience group were informed that they would have "a dozen or so plays" in which to earn as many points as possible. In truth, they were allotted 40 plays. The outcome of each play was consequential and added to the continually displayed summary score, which revealed the cumulative points from all plays for the problem. After the 40 trials the slider bars appeared and the same procedure outlined above followed.

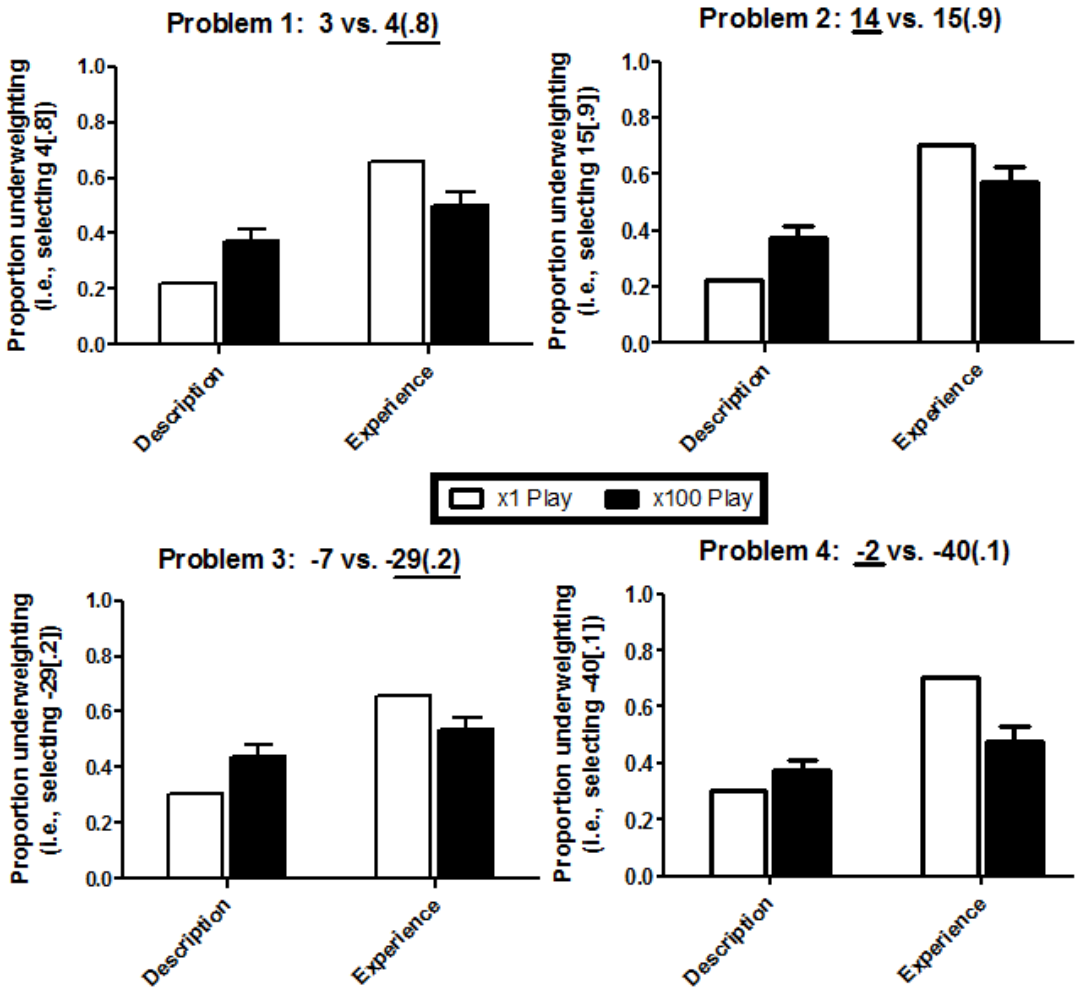
Results

After random allocation there were 55 participants in the Description group and 54 participants in the Experience group. Sixteen participants (15%; 6 in the Description group and 10 in the Experience group) preferred 11 over 12 in the catch question and were removed from subsequent analysis.

Figure 2 shows the proportion of participants preferring the option consistent with underweighting rare events (white bars) and the average proportion of 100 plays allocated to the option consistent with underweighting rare events (black bars). As predicted, and evident from Figure 2, for each of the problems the proportion of underweighting increased in the description group and decreased in the experience group as a function of taking the long-run view. (Note that our prediction for the Description group was *reduced* overweighting of rare events under multi-play conditions. However, as Figure 2 plots "proportion underweighting", our prediction was realized by observing a *greater* proportion of underweighting in the Description multi-play condition). As a result of these changes the description-experience gap that was present for single-play choices in all four problems (Problem 1: $\chi^2_{(1,\ N\ =\ 93)}$ = 17.86, $p < .001$; Problem 2: $\chi^2_{(1,\ N\ =\ 93)}$ = 21.57, $p < .001$; Problem 3: $\chi^2_{(1,\ N\ =\ 93)}$ = 11.58, $p = .001$; Problem 4: $\chi^2_{(1,\ N\ =\ 93)}$ = 14.72, $p < .001$; average D-E 'gap' in proportion underweighting = 41.7%) was reliable for multi-play choices in just one problem (Problem 1: $t_{(1,\ 91)} = 1.96$, $p = .053$; Problem 2: $t_{(1,\ 91)} = 3.04$, $p = .003$; Problem 3: $t_{(1,\ 91)} = 1.70$, $p = .092$; Problem 4: $t_{(1,\ 91)} = 1.66$, $p = .100$; average D-E gap = 13.5%)[2,3].

[2] To statistically assess the interactions via ANOVA is complicated because we had a combination of binary (single-play choice) and continuous (multi-play choice) dependent measures, thus the aggregated single- and multi-play choice data were not perfectly normally distributed, (Skewness = .172 [*SE* = .09], Kurtosis = -1.65 [*SE* = .179]), however, they were not beyond the bounds of acceptability (Leech, Barrett, & Morgan, 2005). Moreover, previous investigations have found the *F*-test to be remarkably robust to deviations from normality when sample sizes are moderately large as they were here (Glass, Peckham, & Sanders, 1972; Lindman, 1974). Thus, we also carried out a 2 (Description vs. Experience) x (2) (Single- vs. Multi-play) mixed ANOVA. As expected, there were significant interactions for all four problems (Problem 1: $F_{(1,\ 91)} = 11.49$, $p < .05$; Problem 2: $F_{(1,\ 91)} = 13.78$, $p < .05$; Problem 3: $F_{(1,\ 91)} = 8.31$, $p < .05$; Problem 4: $F_{(1,\ 91)} = 9.55$, $p < .05$).

[3] An indifference allocation of 50/50 was submitted on 58 (12.5%) of occasions: 32 times in the Description group and 26 times in the Experience group. The indifference allocation is tricky to interpret because 50/50 was also the default and could also therefore reflect participant laziness. Another mixed ANOVA, this time with all indifference allocations removed, found that the interactions remained for three of the problems (Problem 1: $F_{(1,\ 76)} = 5.07$, $p < .05$; Problem 2: $F_{(1,\ 80)} = 13.73$, $p < .05$; Problem 3: $F_{(1,\ 80)} = 3.49$, $p > .05$; Problem 4: $F_{(1,\ 80)} = 12.23$, $p < .05$). We also



Discussion

We investigated the impact of multiple-plays in description- and experience-based choice. As predicted, relative to single play preferences, the multi-play preferences were consistent with reduced *overweighting* of rare events in the description task and reduced *underweighting* of rare events in the experience task. Due to these preference shifts, the description-experience gap that was present under single-play conditions – which is consistent with a wealth of past research (see Hertwig & Erev, 2009) – was essentially eliminated under multi-play conditions.

analysed these same data with a repeated measures logistical regression using the SPSS Generalised Estimating Equations function (GEE) under the binary model. The GEE function required rounding the continuous scale allocation responses to binary preferences, which obviously eliminated much of the data variance. However, even under this analysis, an interaction remained for two of the problems (Problem 1: Wald Chi-Square = 4.25, p = .04; Wald Chi-Square = 2.25, p = .13; Problem 3: Wald Chi-Square = 5.72, p = .02; Problem 4: Wald Chi-Square = 2.89, p = .09).

Thus, at least part of the difference between the description and experience paradigms can be explained by biases caused by short horizon single play framing.

In the description format, single-play preferences were consistent with the idea that people overweight the impact of rare events (Kahneman & Tversky, 1979). Our finding that these preferences shifted when choices were framed as multi-play is also consistent with past literature (see Wedell, 2011, and references therein). However, almost all of the problems used in this literature have been in the form of our Problem 1, that is, in the gain domain and where the risky option is associated with the higher expected value. Thus, previous investigators have been inclined to conclude that multi-play leads to greater choice maximisation. However, the data from our Problems 2 and 4 are inconsistent with this conclusion. In these two problems the safe option was associated with the higher expected value and yet preferred *less* in the multi-play format. Our data therefore suggest that the multi-play format causes reduced overweighting of rare events in the description paradigm. We argue that this shift occurs because the multi-play format induces a long-run horizon that reduces weighting assigned to the rare event.

In the experience format, single-play preferences were consistent with the idea that people underweight the impact of rare events (Camilleri & Newell, 2011b; Hertwig & Erev, 2009). However, we found that these preferences also shifted when choices were framed as multi-play. Our data suggest that the multi-play format causes reduced underweighting of rare events in the experience paradigm. We argue that this shift occurs because the multi-play format induces a long-run horizon that reduces reliance on the more recent outcomes. The implication of this finding is that the sequential nature of the repeated decisions in the experience paradigm is central to the underweighting choice behaviour displayed in previous literature.

The results are also consistent with Lopes' (1981) suggestion that decision-makers are primarily concerned with achieving an outcome that exceeds their aspiration level. In our task, participants were able to allocate as many plays as they liked to the safe option to ensure that their aspiration was met, and were then free to allocate the remaining plays to whichever option appeared more favourable. When given this opportunity, 79% of participants adopted such a hedging strategy and allocated plays across both options. The allocation of plays to the safe option was very similar across problems, which reinforces our argument that EV was not the main driver of preference shifts associated with multi-play – rather, it was the favourability of the rare event.

Our observations have practical implications for choices that are made repeatedly including investment allocation, food selection, and energy use decisions. The results of this study suggest that each individual choice in a sequence of similar repeated choices will tend to be evaluated under short horizons, which can lead to sub-optimal outcomes in the long run (e.g., over-investment in bonds, over-indulgence in chocolate mousse cake, and climate change related disasters). Thankfully, at least part of this bias can be overcome by removing the sequential nature of the choices and requiring a single allocation choice that will have multiple realizations over a longer time horizon.

.

References

Bristow, R. E. (2011). *There is more to life than expected values: Results of free distributions of multiple-play gambles*. Master of Arts Masters thesis, Miami University, Ohio.

Camilleri, A. R., & Newell, B. R. (2011a). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276–284.

Camilleri, A. R., & Newell, B. R. (2011b). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*, 377–384.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. *Review of Educational Research, 42*, 237-288.

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18*, 240-246.

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.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*, 517-523.

Horton, J. J., & Chilton, L. (2010). *The labor economics of paid crowdsourcing*. Paper presented at the Proceedings of the 11th ACM Conference on Electronic Commerce.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from Prospect Theory in descriptive choice. *Psychological Science, 19*, 1015-1022.

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

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for Intermediate Statistics: Use and Implementation* (2 ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Lejarraga, T., Dutt, V., & Gonzalez, C. (2011). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making, 24.*

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs.* San Francisco: W. H. Freeman & Co. Hillsdale.

Lopes, L. L. (1981). Notes, comments, and new findings. Journal of Experimental Psychology: *Human Learning and Memory, 7*, 377-385.

Lopes, L. L. (1996). When time is of the essence: Averaging, aspiration, and the short run. *Organizational Behavior and Human Decision Processes, 65*, 179-189.

Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of Cumulative Prospect Theory and SP/A theory. *Journal of Mathematical Psychology, 43*, 286-313.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 1*, 42–56.

Samuelson, P. A. (1963). Risk and uncertainty: A fallacy of large numbers. *Scientia, 98*, 108-113.

Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking - An experimental test. *The Quarterly Journal of Economics, 112*, 375-405.

Tversky, A., & Bar-Hillel, M. (1983). Risk: The long and the short. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 9*, 713-717.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted, when rare outcomes are experienced (rarely)? *Psychological Science, 20*, 473-479.

Wedell, D. H. (2011). Evaluations of single- and repeated-play gambles. In J. J. Cochran (Ed.), *Wiley Encyclopedia of Operations Research and Management Science*: John Wiley & Sons, Inc.

Chapter 5: The Importance of Probabilistic Representation

The Role of Representation in Experience-based Choice

Adrian R. Camilleri[1]
Ben R. Newell[1]

[1] School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Recently it has been observed that different choices can be made about structurally identical risky decisions depending on whether information about outcomes and their probabilities is learned by description or from experience. Current evidence is equivocal with respect to whether this choice "gap" is entirely an artefact of biased samples. The current experiment investigates whether a representational bias exists at the point of encoding by examining choice in light of decision makers' mental representations of the alternatives, measured with both verbal and nonverbal judgment probes. We found that when estimates were gauged by the nonverbal probe, participants presented with information in description format had a greater tendency to overestimate rare events and underestimate common events. The choice gap, however, remained even when accounting for this judgment distortion and the effects of sampling bias. Indeed, participants' estimation of the outcome distribution did not mediate their subsequent choice. It appears that experience-based choices may derive from a process that does not explicitly use probability information.

The Role of Mental Representation in Experienced-based Choice

The Description-Experience Gap

In recent years a quickly growing literature has emerged contrasting two different formats of choice – description and experience – and the correspondence of decisions observed in each (Rakow & Newell, 2009). A decision from experience (DfE) is one where the possible outcomes and estimates of their probabilities are learned through integration of personal observation and feedback from the environment (Hertwig & Pleskac, 2008). A typical example might be the decision from where to buy your morning coffee as you make your way to work. By contrast, a decision from description (DfD) is one where all possible outcomes and their probabilities are explicitly laid out from the outset (Hertwig & Pleskac, 2008). A typical example might be the decision to bring an umbrella to work after hearing the morning weather forecast and the chance of precipitation.

Surprisingly, recent evidence has found that the decisions made under these two different formats of choice diverge. For example, Hertwig, Barron, Weber and Erev (2004) presented six binary, risky choice problems to participants in either described or experienced format. In the description format, outcomes and their probabilities were completely specified in the form: "Choose between (A) $3 for certain, or (B) $4 with a probability of 80%, otherwise zero". Participants playing this description-based choice task tended to make decisions consistent with prospect theory's four-fold pattern of choice – risk-aversion for gains and risk-seeking for losses when probabilities were moderate or high, but risk-seeking for gains and risk-aversion for losses when probabilities were small (Kahneman & Tversky, 1979). For example, 64% of participants preferred the certain $3 in the decision above. In the experience format, participants were initially unaware of the outcomes and their respective

probabilities and had to learn this information by sampling from two unlabelled buttons. Each sample presented a randomly selected outcome taken from an underlying outcome distribution with the same structure as the problems presented in the description format. Participants were free to sample as often and in any order that they liked until they were ready to select one option to play from for real. Strikingly, participants playing this experienced-based choice task tended to make decisions opposite to the four-fold pattern of choice. For example, only 12% of participants preferred the certain $3 in the decision above. This apparent Description-Experience 'gap' led some to call for the development of separate and distinct theories of risky choice (Hertwig et al., 2004; Weber, Shafir, & Blais, 2004). Fox and Hadar (2006), however, have argued that this conclusion is unwarranted in light of a reanalysis of the Hertwig et al. data. Specifically, they found that prospect theory could satisfactorily account for the patterns of choice when based on participants' experienced distribution of outcomes, which, due to sampling "errors", was often different to the objective distribution from which the sampled outcomes derived.

The crux of the debate centres on the relative importance of sampling bias. This issue has led investigators to employ a number of creative designs that have produced conflicting results (e.g., Camilleri & Newell, in prep.; Hadar & Fox, 2009; Hau, Pleskac, & Hertwig, 2009; Hau, Pleskac, Kiefer, & Hertwig, 2008; Rakow, Demes, & Newell, 2008; Ungemach, Chater, & Stewart, 2009). The purpose of this paper is to re-examine these discrepancies in light of how choice options are represented in the mind of the decision maker.

A Framework for Understanding the Description-Experience Gap

Figure 1 presents a simple framework of the steps involved in making a decision, which is based on the two-stage model of choice (Fox & Tversky, 1998).

At the stage of information acquisition, the decision-maker attempts to formulate a mental representation or impression of the outcome distributions for each alternative[1]. The two modes of information acquisition we are presently concerned with are description and experience.

Figure 1. A simple decision-making framework. Black chevrons represent external, observable events. Grey chevrons represent internal, mental events.

There are two primary accounts for the Description-Experience gap. According to the statistical, or information asymmetry, account, the gap reflects a population-sample difference due to sampling bias inherent to the sequential-sampling, experience-based choice paradigm (Hadar & Fox, 2009). Specifically, the information acquired, or utilised, by decision-makers through their sampling efforts is not equal to the underlying outcome distributions from which the samples derive. As a result of these unrepresentative samples, the experience-based decision maker's understanding of the outcome distribution is quantitatively different to the description-based decision maker's understanding of the outcome distribution. The fact that a Description-Experience gap occurs is therefore relatively trivial because the gambles that decision-makers are subjectively (as opposed to objectively) choosing between are different. Apples are being compared to pineapples. Thus, this account is primarily concerned with the level of information acquisition and the major prediction is that the gap should disappear when information acquired in both the DfD and DfE paradigms are equivalent.

[1] Not all choice frameworks require the formation of mental representations (e.g., Busemeyer & Townsend, 1993).

In contrast, according to the psychological account, the gap is something over and above mere sampling bias: it reflects different cognitive architecture at the level of choice. Description- and experience-based choices recruit different evaluative processes that operate according to different procedures. Thus, this account is primarily concerned with the level of choice and the major prediction is that the gap will remain even when information acquired in both the DfD and DfE paradigms is equivalent.

A number of methodologies have been used to account for sampling bias and therefore provide a test between the statistical and psychological accounts. Sampling bias has been eliminated by yoking described problems to experienced samples, (Rakow et al., 2008), conditionalising on the subset of data where the objective and experienced outcome distributions match (Camilleri & Newell, in prep.), and obliging participants to take representative samples (Hau et al., 2008; Ungemach et al., 2009). The first two of these studies found that elimination of sampling bias all but closed the gap. In contrast, the last two of these studies found that even after accounting for sampling bias there nevertheless remained a choice gap (see Hertwig & Erev, 2009, and Rakow & Newell, 2009, for good overviews). This mixed evidence has ensured that a level of controversy persists.

The Stage of Mental Representations

One way to reconcile these conflicting sets of observations is to reconsider the framework presented in Figure 1. The current methodologies accounting for sampling bias all attempt to equate information presented at the stage of information acquisition. That is, they all work to ensure that decision makers have been exposed to the same information. There are two reasons for suspecting that the information participants are exposed to may be unequal to the information participants actually

use to make their decisions. First, it is not clear that participants construct representations of outcome distributions from all of the information they are exposed to. In the free sampling paradigms, for example, participants may utilise a two-step sampling strategy in which they begin by obtaining a general overview of the outcomes of each alternative (e.g., the magnitudes) before moving on to a more formal investigation of the probability of each outcome occurring. Partial support for this claim comes from observations of recency, whereby the second half of sampled outcomes, as opposed to first half, better predicts choice (Hertwig et al., 2004, but see Hau et al., 2008). In the forced sampling paradigm, moreover, it seems doubtful that participants take into account, and linearly weight, information from up to 100 samples when forming a representation due to memory and/or attentional limitations (Kareev, 1995; 2000). Indeed, we suspect such limitations are responsible for the meagre amount of sampling typically observed in free sampling designs (e.g., a median of 15 samples in Hertwig et al., 2004).

Second, we know that when reasoning about uncertainty, mathematically equivalent (external) representations of probabilities are not necessarily computationally equivalent (Gigerenzer & Hoffrage, 1995; Hau, Pleskac, & Hertwig, 2010). For example, “80%” is mathematically equivalent to “8 out of 10”, yet these two pieces of information can be used in non-equivalent computational ways, leading to different decisions (see also the ratio bias effect; Bonner & Newell, 2008). Importantly then, it should not be assumed that what people are given (i.e., information contained in a description or aggregated from experience) is identical to what people take away. Viewing this point within the framework presented in Figure 1 implies that mathematically equivalent contingency descriptions and experienced contingencies could nevertheless be represented differently depending on whether

the information is acquired by description or experience. If true, the possibility then exists that even when sampling bias is objectively eliminated, there may still remain subjective differences in mental representations actually operated upon. And of course, it is these actually operated upon mental representations that we are most interested in.

A small number of studies have attempted to examine these mental representations (Barron & Yechiam, 2009; Hau et al., 2008; Ungemach et al., 2009). For example, Ungemach et al. (2009) asked participants to verbally report the frequency of rare event occurrences. Similarly, Hau et al. (2008) asked participants to verbally estimate the relative frequency (as either percentages or natural frequencies) of each outcome. The results of these studies are consistent and suggest that people are largely accurate and, if anything, overestimate small probabilities and underestimate large probabilities. The direction of these estimation errors would actually have the effect of reducing the size of the gap.

Based on this evidence, one might feel confident to conclude that the source of the gap is independent of distorted representations of the outcome distributions; instead, it must be due to sampling bias and/or inherent to the choice mechanism processes. This conclusion is perhaps premature for two reasons. First, there are concerns regarding the methodology used to measure the verbal representations. In the Hau et al. (2008) study 2, for example, participants were aware that, at least after the first problem, they would have to make relative frequency judgments. It is possible that participants' sampling efforts were then at least partially driven by their attempt to accurately learn the contingencies, and crucially, represent these contingencies in a verbal format. Ungemach et al. (2009) avoided this issue by presenting the judgment probe as a surprise. However, the probe comprised simply

of participants stating how frequently the rare outcome had been observed. This task is therefore quite distinct from participants appreciating the probability of the rare event being observed on the next sample, which, at the very least, additionally involves appreciation of the number of samples taken.

Second, there are concerns regarding the validity of the verbal judgment probe in the context of experienced-based choice. In the DfE task, the decision maker's only goal is to decide which of the options is "better". Presumably, decision makers could use a "satisficing" heuristic and attempt to make this decision with minimal computational effort (Simon, 1990; Todd & Gigerenzer, 2000). Therefore, in terms of mental representations, the minimalist requirement in this task is to form some sort of impression as to which option is "better", irrespective of the magnitude of that superiority or the specific probabilities of each outcome. Therefore, in the experienced-based choice task, there is no inherent need to formulate a propositional statement about the probability of each outcome (as is presented in the description-based choice task). Given evidence that humans possess a nonverbal numerical representation system (Dehaene, Dehaene-Lambertz, & Cohen, 1998), it may be that a nonverbal assessment probe is better able to capture the summary impression because it makes no reference to explicitly described verbal probabilities.

Pursuing this logic, Gottlieb, Weiss and Chapman (2007) used both a verbal or nonverbal assessment tool to probe decision makers' mental representation of outcome distributions in DfD and DfE (forced sampling) paradigms. The verbal probe asked participants to complete the sentence "__% of cards were worth __ points". The nonverbal probe consisted of a large grid composed of 1600 squares whose density could be adjusted by pressing on the up and down arrow keys of a normal keyboard. Participants were asked to adjust the density of the grid to match

their belief as to the relative frequency of each option. Interestingly, there was a disparity in judgment accuracy depending on whether judgments were probed verbally or nonverbally. Similar to past studies, when probed verbally, participants' judgment accuracy was best modelled by a linear function with fairly good accuracy regardless of mode of information acquisition. In contrast, when probed nonverbally, participants' judgment accuracy was best modelled by a second-order polynomial implying underestimation of large probabilities and overestimation of small probabilities. Importantly, there was an interaction suggesting that this distortion from perfect mapping was much stronger in the description than in the experience condition.

Two details are particularly intriguing about these findings. First, the second-order polynomial curves obtained with the nonverbal judgment probe were strikingly reminiscent of the probability-weighting function described by Prospect Theory (PT; Kahneman & Tversky, 1979). If PT is taken as a process model of choice, then the weighting function reflects the mental adjustment that decision makers apply to their calculation of expected utility for each option. However, these findings suggest that an alternative explanation is that probability information is distorted at the level of mental representation, and that this distortion may be observed only with a nonverbal judgment probe. Second, accuracy when probed nonverbally was worse for the description condition than in the experience condition. This difference is surprising because adjusting a grid's density to that of an explicit, known proportion would seem an easier task than adjusting to an imprecise, non-specified proportion gleaned from sequential sampling. The difference potentially implicates judgment distortions as contributing to the gap and, moreover, leads to suspicion that nonverbal probes of

mental representations may be a more sensitive form of mental representation assessment for experienced-based choice tasks.

The Current Experiment

Primary explanations for the Description-Experience choice gap have been statistical (the result of sample bias) and psychological (the result of a weighting bias at the time of choice). The current study examines whether the gap could also be a representational phenomenon, that is, the result of a distortion at the time of encoding. The specific aims of the current experiment were to test whether there exists a representational bias and whether, when controlling for sampling and any representational bias, there remains a choice gap. To examine these objectives we employed the free-sampling, money machine paradigm (Hertwig et al., 2004) in combination with both a verbal and nonverbal probe to assess participants' judgments of the outcome distributions (Gottlieb et al., 2007).

Method

Participants

The participants were 80 undergraduate first year University of New South Wales psychology students (48 females), with an average age of 19.5 years and a range of 18 to 36 years. Participation was in exchange for course credit, plus payment contingent upon choices.

Materials

Choice Problems. The eight choice problems used are shown in first three columns of Table 1. Each problem consisted of two options: an option that probabilistically paid out one of two values versus an alternative option that always paid out a single value. The expected value was always higher for the probabilistic

option. The problems were chosen to evenly split between the domains of gain and loss, and also to span a range of probabilistic rarity (5%, 10%, 15%, and 20%). The option predicted by Prospect Theory to be preferred was labelled the "favoured" option and the alternative option was labelled the "non-favoured" option (Kahneman & Tversky, 1979). Specifically, the favoured option was the option containing the rare event when the rare event was desirable (e.g., 14 is a desirable rare event in the option 14 [.15] and 0 [.85]), or the alternative option when the rare event was undesirable (e.g., 0 is an undesirable rare event in the option 4 [.8] and 0 [.2]).

Decision Task. The decision task was the free sampling "money machine" paradigm, similar to the one employed by Hertwig et al. (2004). In the description-based choice condition, two alternative money machines were presented on screen. Each machine was labelled with a description of how that machine allocated points. All of the safe option machines were labelled in the form "100% chance of x", where x represents the outcome. All of the risky option machines were labelled in the form "y% chance of x, else nothing", where y represents the probabilistic chance of a non-zero outcome, and x represents the outcome.

In the experience-based choice condition, the two alternative money machines were also presented on screen, but they were labelled only with the letters "A" and "B", respectively. Each of the machines was associated with a distribution of possible outcomes in accordance with the objective probabilities as shown in Table 1. Samples from each machine were non-random draws from the respective outcome distributions that were selected by an algorithm to maximally match the objective

probability with the participants' experienced distribution, thereby minimising sampling variability[2].

In both decision conditions, when the participant was ready to make their one-shot decision, they pressed on a "Play Gamble" button that allowed them to select the machine they preferred to play from. In all cases allocation of safe and risky options to the left and right machines was counterbalanced and the order of the problems was randomised.

Judgment Probes. Both the verbal and nonverbal judgment probes asked participants to first enter the number, and specific value, of each outcome paid out by each machine. Contingent on this response, participants were then subsequently asked to provide a probability estimate for each identified outcome. Thus, participants were not asked to make an estimate for an outcome they had not seen, and some participants did not make an estimate for an outcome they had seen (because they had not identified this outcome initially).

The verbal judgment probe asked participants to complete the sentence: "*x is paid out by the machine __% of the time*", where "x" refers to the outcome. In contrast, the nonverbal judgment probe presented a grid made up of 40x40 small squares, each containing the number "x", along with the instructions: "Adjust the frequency of x's in the grid to match the frequency of x paid out by the machine. You can adjust the density of the grid by pressing 'up' and/or 'down' on the keyboard until x fills the grid according to its frequency". The default grid showed

[2] On each sample, the participants' experienced distribution was compared to the objective distribution and the outcome that minimised this difference was presented. This algorithm produced repeating patterns of outcomes. For example, when the objective probability was 20%, the pattern of outcomes repeated itself in blocks of 5 outcomes. A typical approach to exploring the money machines in our data, based on the median values, was to sample from the risky option seven times, sample from the safe option twelve times, and then sample from the risky option eight times before making a final choice. Thus, the typical sequence of outcomes for a participant playing problem 1 would be something like 4, 4, 0, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 0, 4, 4, 4, 4, 0, 4, 4. It thus seems unlikely that participants in the current study were able to identify the repeating pattern.

50% of the squares, randomly dispersed (Figure 2). Each press of the key increased or decreased the frequency of squares by 1%, randomly over the grid. For the purposes of analysis, the visual display was converted into a percentage after the participant made his or her judgment.

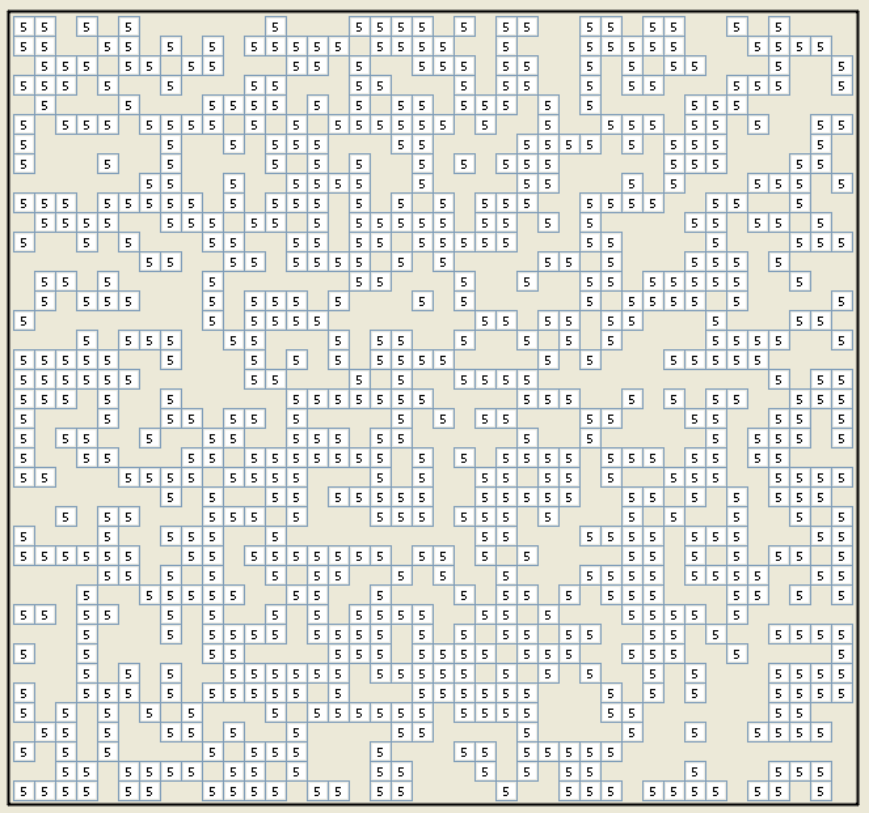


Figure 2. Screenshot of a default grid. The value in the box corresponds to the outcome value provided by the participant.

Design

The experiment was a 2 x 2 x 2 within-subjects design and counterbalanced such that participants completed one of the eight problems in each of the eight experimental cells. The three binary independent variables were presentation mode (description or experience), judgment probe type (percentage or grid), and judgment probe time (before or after choice). The two dependent variables were the choice made (favoured or non-favoured option) and the accuracy of judged outcome

probabilities (measured as the average absolute difference between experienced[3] and judged probabilities).

Procedure

An on-screen video tutorial explained that the experiment was about making decisions between different alternatives, that the objective of the game was to maximise the amount of points won, and that at the end of the experiment points would be converted into real money according to the conversion rate of 10 points = AUD$1. The tutorial combined written instructions with movements of a ghost player to demonstrate how to play the description- and experience-based decision tasks and correctly answer the verbal and nonverbal judgment probes. Participants were informed that they could sample from each option as often and in any order that they liked. Thus, participants could take samples ranging in size from one to many hundreds. Instructions for the grid probe were: "You will see small versions of the target value randomly superimposed on a square grid. You should adjust the density of the target value on the grid to match the frequency of the target value paid out by the machine". In order to reduce potential wealth effects, no feedback was given of the points that participants were awarded for their one-shot choice for each problem.

At the completion of the experiment a screen revealed the participant's 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.

[3] In the description condition, the "experienced" probabilities were the objective probabilities. In the experience condition, the "experienced" probabilities depended on what outcomes had actually observed.

Results

Judgment

Figure 3 plots judged probabilities against experienced probabilities separately for both presentation modes (description vs. experience) and both judgment probe types (percentage vs. grid)[4]. Inspection of the figure suggests that there is an interaction between presentation mode and judgment probe type. Specifically, it appears that the verbal percentage probe produced better calibrated judgments for those in the Description condition (i.e., estimates closer to the identity line), whereas the non-verbal grid probe produced better calibrated judgments for those in the Experience condition.

We tested this interaction using a mixed model (using the lmer function of R [Bates & Maechler, 2009; R Development Core Team, 2008], as described by Baayen, Davidson, & Bates, 2008, and Bates, 2005). This function is robust when designs are unbalanced, as is the case here as a result of omitted data. The dependent variable was a measure of judgment error: the absolute value of the difference between, on the one hand, the experienced probability of the common event, and, on the other, the normalized judged probability of the common event (i.e., the judged probability of the common event divided by the sum of that and the judged probability of the rare event – the two often did not add to 100). The main predictors were presentation mode, judgment probe type, and their interaction. Problem number (as a nominal variable or factor) was also included as a fixed effect; it accounted for significant variance, but judgment probe time (before vs. after choice) was excluded

[4] We collapsed across judgment probe time (before vs. after choice) because this manipulation had no effect. Eighty-one trials (12.6%) were excluded because estimates were unreasonable (the average absolute difference between experienced and judged probabilities was 40 or higher) or the participant failed to make an estimate.

because it was never significant in any analysis. Participant identity was included as a random effect. The interaction was significant at $p = .0042$ (as assessed by Markov Chain Monte Carlo sampling). Thus, the magnitude of the difference between participants' experienced probabilities and their judged probabilities varied depending on whether the information was acquired by description or experience. Examination of the fitted mean errors revealed that participants in the Description conditions were relatively more accurate with the percentage probe than the grid probe ($M = 0.98$ vs. 6.64, respectively) compared to participants in the Experience conditions ($M = 3.22$ vs. 5.70, respectively). Further inspection of the two bottom panels of Figure 3 suggests that there is a difference in the slopes of the regression lines between the Description and Experience conditions.

In order to make this directional inference, we regressed an error term (common event judged probability – common event experienced probability) on presentation mode (description vs. experience) for cases where the nonverbal grid judgment probe was used. After removing one outlier, the interaction was significant at $p = .0291$. A similar analysis for cases where the verbal percentage judgment probe was used was not significant. Thus, the tendency to overestimate rare events and underestimate common events was much stronger in the Description condition, but only when assessed with the nonverbal probe.

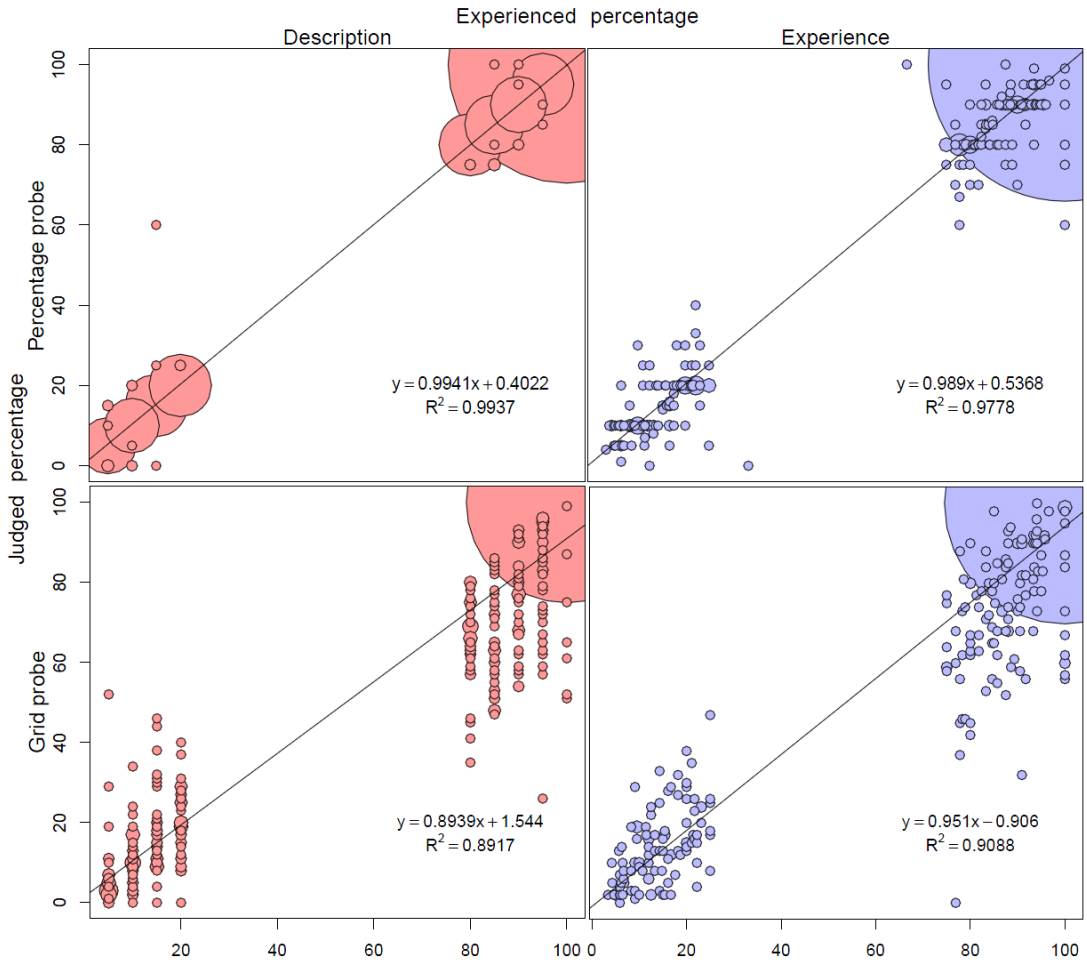


Figure 3. Experienced percentages plotted against judged percentages as a function of presentation mode (description on left panels, experience on right panels) and judgment probe type (verbal percentage in upper panels, nonverbal grid in lower panels). The size of the plotted circles relates the number of identical data points. The solid line depicts the least-square regression lines describing the relation between the experienced and judged probabilities.

Choice

The percentage of participants selecting the option predicted by Prospect Theory to be the favoured choice is displayed in Table 1. The difference between Description and Experience conditions falls in the expected direction for six of the eight problems[5]. Two of these differences were significant by individual chi-square tests (*p*'s < .05). Indeed, the odds of selecting the favoured option in the Description

[5] Although a within-subjects design, the comparisons were all between-subjects because participants made only one decision for each problem in either the description or experience choice format.

condition were more than 1.7 times the odds of selecting the favoured option in the Experience condition. Although indicative, and commonly used in the literature, this rough analysis fails to properly assess the role of presentation mode because it ignores the variance in participants' experience and judgments.

Table 1

Percentage Choosing the Option Predicted by Prospect Theory (Kahneman & Tversky, 1979) to be Favoured

Problem Number	Option		Percentage selecting the favoured option		
	Favoured	Non-favoured	Description	Experience	Gap
1	3 (1.0)	4 (.8)	68	54	14
2	-2 (1.0)	-50 (.05)	55	41	14
3	14 (1.0)	17 (.9)	71	42	29*
4	-3 (1.0)	-32 (.1)	47	49	-2
5	14 (.15)	1 (1.0)	57	49	8
6	-12 (.85)	-9 (1.0)	42	42	0
7	25 (.2)	4 (1.0)	51	33	18
8	-9 (.95)	-8 (1.0)	64	31	33*

* indicates significant difference between description and experience conditions.

To test the effect of presentation mode on choice, we used a logistic mixed model, with participant identity as a random effect, and including problem number as a fixed effect (as before). The dependent variable was whether or not the favoured option was selected. The main predictors were presentation mode, judgment probe type, experienced probability and normalized judged probability (as used before). Of these predictors, the only significant effects were of presentation mode (coefficient -.627, z = -3.43, asymptotic p = .0006) and experienced probability (coefficient -.071,

z = -2.38, p = .0172). The odds of selecting the favoured option in the Description condition were more than 1.8 times the odds of selecting the favoured option in the Experience condition. Importantly, the effect of normalized judgment was not significant (z = -.90). Thus, the effect of presentation mode on choice is apparently not mediated by its effect on judgment.

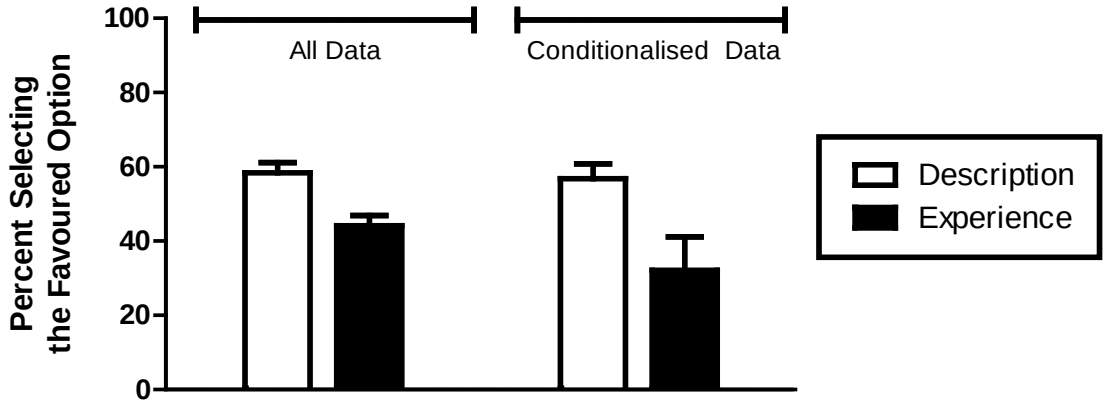


Figure 4. The percentage of participants selecting the favoured option in the Description and Experience conditions. The conditionalised data were those trials where the participants' experienced and (normalised) judged rare event probabilities were both within 10% of the objective rare event probability (see footnote 6). Error bars indicate the standard error of the mean.

In order to show this result graphically, we conditionalised on the subset of data where participants' experienced and judged distributions were approximately equal to the objective distribution[6]. The subset of data comprised of just 28 experience- and 153 description-based decision trials. Thus, the subset did not equally represent all participants, problems and conditions, and, hence, inferential statistics were not conducted. Nevertheless, the retained data do serve to visually represent the major finding of our regression analysis. Namely, as shown in Figure 4, even within the

[6] Specifically, we retained only those trials in which the experienced and (normalised) judged rare event probabilities were both within 10% of the objective rare event probability. For example, in Problem 1, where the objective probability for the rare event is .2, we retained only those trials where the experienced and judged probability for the rare event were both between .18 and .22 (i.e., "within 10%" of .2 = ±.02).

subset of data without sampling or judgment errors, there remains a gap between description- and experienced-based choices.

Discussion

Judgment

The current study attempted to take a representational perspective in explaining the observation of a gap between description- and experienced-based patterns of choice. The first aim was to examine whether there exists representational bias, that is, an encoding distortion of the outcome distribution prior to choice. To that end, we asked participants to judge each problem's outcome distribution using either a verbal or nonverbal probe.

When participants made their judgment using a nonverbal probe – adjusting the density of a large grid to correspond to the relative probability of each outcome – absolute judgment accuracy in the Description and Experience conditions was approximately equivalent. This result is particularly surprising because it implies that decision makers are equally able to nonverbally represent a non-explicit, gist impression constructed from sequential sampling and a numerical percentage explicitly presented. Of course, this is not to say that judgments were particularly accurate: they were not; participants in both groups displayed a tendency to underestimate common events and overestimate rare events. This observation replicates Gottlieb et al.'s (2007) intriguing finding that percentages are distorted when transformed into nonverbal estimates. The current study extends this observation to a free sampling design where participants decided the size of their samples. Admittedly, it is possible that at least some of this bias is due to an anchoring effect at the probe density starting point (50%). What is perhaps more

interesting, and not explainable in terms of anchoring, is the fact that the distortion, this tendency to underestimate common events and overestimate rare events, was much greater for those in the Description conditions than those in the Experience conditions.

When participants made their judgment using a verbal probe – entering a number to correspond to the relative probability of each outcome – absolute judgment accuracy was greater in the Description conditions. Contrary to some previous research, there was little evidence that participants overestimated small probabilities and underestimated large probabilities (Barron & Yechiam, 2009; Hau et al., 2008). In fact, accuracy in both conditions was fairly high, which replicates some other studies that have asked for probability judgments (Fox & Hadar, 2006; Gottlieb et al., 2007; Ungemach et al., 2009), and were superior to those achieved by participants making judgments via the nonverbal grid probe.

The greater absolute judgment accuracy observed when using the verbal probe may lead some to the conclusion that this type of probe should be preferred when assessing representations of outcome distributions We have three cautions. First, accuracy when using the verbal probe in the Description condition depended only on memory, not judgment, and is therefore inflated. Second, the nonverbal grid task was, on average, prone to greater variability because of the potential for super- or sub-additivity. Specifically, because one grid was presented for each outcome identified, participants' summed judgments for the outcome probabilities for each option often deviated from 100%. Super- and sub-additivity did not occur when using the percentage probe because participants could easily add up their estimates and ensure that they totalled 100%. Third, even if decision-makers can interpret and numerically report the content of their mental representations when explicitly probed

by a verbal probe, if this is not the actual representation and information used to make the decision, then such (albeit accurate) information is non-diagnostic in the pursuit of understanding experience-based choice.

What then are we to conclude about nonverbal judgment probes? Despite producing less accurate results overall, they uniquely discriminate between description- and experience-based formats of information acquisition. Nonverbal judgment probes may therefore permit greater sensitivity to presentation mode when gauging mental representations. Potentially, this is because representations of outcome distributions are themselves nonverbal (Dehaene et al., 1998).

Choice

The second aim was to examine whether representational biases constitute, in addition to sampling bias, a major cause of the choice gap between description and experience choice formats. As described above, there does appear to be a representational bias, at least when probed nonverbally, and this bias is stronger when information is acquired by description. Assuming choices are made based on these differentially distorted outcome distributions, representational biases may be sufficient to cause subsequent differences in choice.

To begin, we again found a disparity in the patterns of choice made to identical problems depending on whether they were presented by description or experience (Hertwig et al., 2004; Weber et al., 2004). The size of the gap observed in our data, 14.4 percentage points, is relatively small when compared to previous free sampling DfE paradigm studies (e.g., 36 percentage points in Hertwig et al., 2004). This is probably due to the relatively large amount of samples taken by our participants (median of 28, compared to 15 in Hertwig, et al., 2004) coupled with our

manipulation for sample outcomes to track the objective probabilities as closely as possible (see Footnote 2).

Even after accounting for sampling bias and judgment distortions, however, the mode by which information was acquired – by description or from experience – remained significant. The differential distortions observed in judged outcome distributions across presentation mode did not mediate the choice gap between description- and experience-based choices. Importantly then, the choice gap appears to be being driven by something over and above both sampling bias and judgment distortions. This finding supports the work of those that have obliged participants to sample until they have observed outcomes matching exactly or nearly exactly the objective outcome distribution (Hau et al., 2008; Jessup, Bishara, & Busemeyer, 2009; Ungemach et al., 2009).

Implications

How can we explain the remarkable conclusion that participants' own estimation of the outcome distribution does not mediate their subsequent choice? It may be the case that choices are made separately from judgment of the outcome distributions. Recently it has been noted that in many situations, both inside and the lab and out, people's choice behaviour is at odds with their judgment (Barron & Yechiam, 2009). For example, immediately following a suicide bombing, people believe the risk decreases but at the same time exhibit more cautious behaviour. Thus, choice may not be made using representations of the outcome distributions at all. Decision field theory, for example, models choice processes as the gradual change of preference between options and makes no reference to a mental representation of each option's outcome distribution (Busemeyer & Townsend, 1993). This conclusion has implications for the development of models of choice. Specifically, our results

suggest that models that incorporate two stages, one at the level of representation and one at the level of choice, may be unnecessary when it comes to predicting experienced-based choice. For example, one of the leading two-stage choice models – cumulative prospect theory (Fox & Tversky, 1998; Tversky & Fox, 1995) – fares no better at explaining our data when based on judged, compared to experienced, outcome distributions (Appendix A). This result echoes the finding of Hau et al. (footnote 2, 2008). Our conclusion also seems to be consistent with the findings from a recent choice prediction competition. Whereas all models submitted to predict description-based choices assumed that outcomes were weighted by probabilities, the majority of models submitted to predict experience-based choices were such that "the concept 'probability' did not play an important role" (Erev et al., 2009).

With regard to the two primary choice gap explanations – statistical or psychological– the current data lend support to the latter account. That is, that there exist true differences in the choice mechanics used to make experience-based decisions that are over and above the effects of biased samples and judgment errors. What else could be driving the gap? Hertwig et al. (2004) demonstrated that recency, the tendency to rely more heavily on more recently observed outcomes, was another influence on experienced-based choice and hence the gap. In our data, however, we observed no difference in success when predicting choice from the mean value of the first versus second half of observed outcomes (56.2% versus 60.5%, respectively, $t(560) = -1.026$, n.s.). Our interpretation is that the gap derives from a probabilistic focus in the description format and a non-probabilistic focus in the experience format. Indeed, Rottenstreinch and Kivetz (2006) argue that non-probabilistic thinking is more likely in situations where people partially control events and when there is relatively low salience of probabilistic cues. If Rottenstreinch and Kivetz's

interpretation is correct, the experience format in which probabilities are never explicitly mentioned is more likely to yield non-probabilistic thinking than the description format in which probabilities are clearly presented. Moreover, evidence from outside the lab also suggests that executives' decision-making rarely explicitly considers outcome probability (Jeske & Werner, 2008). We feel that the distinction between a probabilistic and non-probabilistic focus during choice is an interesting one for further research to pursue.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effect modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making, 4*, 447-460.

Bates, D. (2005). Fitting linear mixed models in R. *R news, 5*, 27-30. http://journal.r-project.org.

Bates, D. & Maechler, M. (2009). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-32, http://CRAN.R-project.org/package=lme4.

Bonner, C., & Newell, B. R. (2008). How to make a risk seem riskier: The ratio bias versus construal level theory. *Judgment and Decision Making, 3*, 411-416.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A dynamic cognition approach to decision making. *Psychological Review, 100*, 432-459.

Camilleri, A. R., & Newell, B. R. (in preparation). The description-experience 'gap': Psychological or statistical phenomenon? *The University of New South Wales.*

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21*, 355-361.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S., Hau, R., Hertwig, R., Stewart, T., & Lebiere, C. (2010). A choice prediction competition, for choices from experience and from description. *Journal of Behavioral Decision Making, 23*, 15-47.

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.

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science, 44*, 879-895.

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

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18*, 240-246.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making, 4*, 317–325.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities? *Journal of Behavioral Decision Making, 23*, 48-68.

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.

Hertwig, R., & Erev, I. (2010). The Description–Experience Gap in Risky Choice. *Trends in Cognitive Sciences, 13*, 517-523.

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Charter & M. Oaksford (Eds.), *The Probabilistic Mind:*

Prospects for Rational Models of Cognition (pp. 209-236). Oxford, England: Oxford University Press.

Jeske, K.-J., & Werner, U. (2008). Impacts on decision making of executives - Probabilities versus outcomes. *Journal of Neuroscience, Psychology, and Economics - Research, 1*, 49-65.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2009). Feedback produces divergence from Prospect Theory in descriptive choice. *Psychological Science, 19*, 1015-1022.

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

Kahneman, D., & Tversky, A. (2000). *Choices, Values, and Frames*. Cambridge University Press: New York.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263-269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*, 397-402.

R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*, 1-14.

Rottenstreich, Y., & Kivetz, R. (2006). On decision making without likelihood judgment. *Organizational Behavior and Human Decision Processes, 101*, 74–88.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–19.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavior and Brain Sciences, 23*, 727-780.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269–283.

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.

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.

Appendix A: Modelling the Data with Cumulative Prospect Theory

One of the most successful models in the area of description-based choice is prospect theory (Tversky & Kahneman, 1992). According to prospect theory, decision makers calculate a 'value' for each alternative by multiplying the utility value of the outcome by a decision weight. Crucially, the value and decision weight functions are nonlinear. For example, the decision weight function implies that people overweight low percentages and underweight moderate and high percentages. Particularly germane in light of the current experiment, the two-stage model of cumulative prospect theory (CPT) applies the decision weight function adjustment to the decision maker's judged outcome percentage, as opposed to the objective or experienced percentage (Fox & Tversky, 1998; Tversky & Fox, 1995).

We used CPT to predict choices as a function of objective, experienced, as well as raw judged probabilities for the non-zero event (see Hau et al., 2008 for details). We fitted the data to two versions of cumulative PT: one based on parameters estimated from description-based choice (Tversky & Kahneman, 1992), and one based on parameters estimated from experience-based choice (Hau et al., 2008).

As shown in Table 2, each version of the PT model did relatively poorly. Unsurprisingly, description-based choices were better predicted with parameters estimated from described choices tasks, and experienced-based choices were better predicted with parameters estimated from experienced choices. In the Experience conditions, on average, there was little difference in prediction accuracy between judged or experienced percentages, but both did better than objective percentages.

Table 2

Percentage of Choices Correctly Predicted by Cumulative Prospect Theory when Fitted with Parameters Estimated for Description- (Tversky & Kahneman, 1992) and Experience-based Choice (Hau et al., 2008)

Fitted with Tversky and Kahneman (1992) estimated parameters

Judgment Probe Condition	% Choices Correctly Predicted				
	Description		Experience		
	Objective Percentages	Judged Percentages	Objective Percentages	Experienced Percentages	Judged Percentages
Percentage	56	57	43	56	59
Grid	59	56	43	54	53

Fitted with Hau et al. (2008) estimated parameters

Judgment Probe Condition	% Choices Correctly Predicted				
	Description		Experience		
	Objective Percentages	Judged Percentages	Objective Percentages	Experienced Percentages	Judged Percentages
Percentage	49	52	52	58	61
Grid	40	54	52	56	53

Note: Objective percentages refer to the underlying problem outcome distribution. Experienced percentages refer to the outcome distribution observed during sampling. Judged percentages refer to the estimated outcome distribution.

The Relevance of a Probabilistic Mindset in Risky Choice

Adrian R. Camilleri[1]
Ben R. Newell[1]

[1] School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Choice preferences can shift depending on whether outcome and probability information about the options are provided in a description or learned from the experience of sampling. We explored whether this description-experience "gap" could be explained as a difference in probabilistic mindset, that is, the explicit consideration of probability information in the former but not the latter. We replicated the gap but found little evidence to support our main hypothesis. Nevertheless, the data inspired a number of interesting proposals regarding experimental design, preference for probability information, sampling strategies, optimal presentation format, and the probability judgment probe.

The Relevance of a Probabilistic Mindset in Risky Choice

Individuals, businesses, and governments are continually challenged by the prospect of making decisions in the face of uncertainty. For example, Google's acquisition of the mobile start-up company Android in 2005 was considered a risky move because, at the time, the smartphone industry was dominated by the battle between the iPhone and BlackBerry and few could see room for a new challenger. However, just five years on, Android is now the leading smartphone operating system in the U.S. by market share (Whitney, 2010) and has been deemed by Google as their best acquisition ever.

It is interesting to consider what mindset the Google leadership team adopted when they decided to acquire Android. The choice may have been predominately "description-based", that is, rooted in hard numbers of estimated financial outcomes and their likelihoods. In contrast, the choice may have been predominately "experience-based", that is, rooted in instinct sharpened by the practice of having acquired dozens of other companies. The question is more than academic in light of a growing body of evidence showing that choice differences occur between identical decisions depending on whether choice-relevant information is acquired from a description or garnered from experience (Rakow & Newell, 2010).

Description- vs. Experience-based Choice

Hertwig, Barron, Weber and Erev (2004) contrasted these two risky choice formats by presenting decision-makers with the same problem in either the description or the experience format. Those in the description group were explicitly told the potential outcomes and their probabilities. For example, Problem 1 was a choice between a "100% chance of 3" and an "80% chance of 4, else 0". In contrast,

those in the experience group were not explicitly told anything but were instead allowed to repeatedly sample outcomes, with replacement, from a distribution that matched the description given to those in the other group.

Choice preferences were clearly influenced by presentation format. For example, in Problem 1, just 36% of participants selected the risky option when the decision was made from description yet 88% preferred this option when the decision was made from experience. Such large differences have now been observed across many different problems examined in numerous studies (for a review, see Hertwig & Erev, 2009). The common finding is choice behavior consistent with overweighting of rare events when gambles are explicitly described but objective or underweighting of the rare events when gambles are learned from sequential feedback (Camilleri & Newell, 2011a).

Some researchers have argued that the gap is largely the result of external and internal sampling biases present in the experience format (e.g., Camilleri & Newell, 2011b). External sampling biases occur when an observed sample of outcomes does not accurately reflect the true outcome distribution, which is common when participants take small samples (Hertwig & Pleskac, 2010). Internal sampling biases occur when a mental sub-sample of outcomes does not accurately reflect the observed outcome distribution, which is common when participants rely more heavily on recent observations (Hertwig et al., 2004).

In addition to these causes, there remains a strong belief that the gap is caused by yet additional factors (e.g., Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig & Erev, 2009; Ungemach, Chater, & Stewart, 2009). The factor investigated in the present study we term “probabilistic mindset” and refers to the explicit consideration of outcome distributions or probabilities during choice. Specifically, we examined

the possibility that the gap might partly be the result of a probabilistic mindset in the description format but a non-probabilistic mindset in the experience format.

Probability vs. Frequency Information

Most studies of description-based choices confer likelihood information through probabilities. An alternative that leaves explicit outcomes and their likelihoods is frequency information (e.g., "32 out of 40 occasions get 4"), which has been shown to produce behavior that is different than when probability information is presented (e.g., Slovic, Monahan & MacGregor, 2000). Cosmides & Tooby (1996) argue that evolution has shaped the mind to operate with frequency information and go on to demonstrate that this information format improves decision-making across a number of tasks, including Bayesian reasoning.

In the context of the risky choice, evidence for a frequency effect has been mixed. On the one hand, Gottlieb Weiss, and Chapman (2007) presented their participants with different risky problems in percentage and frequency formats and found that choices in the latter were closer to the choices made by participants who saw outcomes sequentially (i.e., experience-based). On the other hand, Rakow, Demes, and Newell (2008) found no differences between percentage and frequency formats. Thus, our first research question was whether probability and frequency formats produce preference differences in the context of risky choice.

Probabilistic vs. Non-probabilistic Mindset

Traditional accounts of description-based choice have placed the consideration of probability information – in our terms, a probabilistic mindset – at the fore. For example, in prospect theory, the "value" of an option is determined by summing the

product of the possible outcomes by their probabilities, with each being adjusted by different non-linear weighting functions (Kahneman & Tversky, 1979).

Accounts of experience-based choice are more diverse. One school of thought suggests that prospect theory, with its emphasis on explicit probability representation, can also successfully account for experience-based choices (Hau et al., 2008; Fox & Hadar, 2006). Indeed, participants can provide fairly accurate probability estimates for the outcomes they have observed (e.g., Ungemach et al., 2009).

However, probability estimates do not accurately predict choice, suggesting that participants might be able to provide precise estimates when explicitly probed, but refrain from using such information when making the decision itself (Camilleri & Newell, 2009). This hypothesis is consistent with recent other findings including the coexistence of overestimation and underweighting of rare events in situations outside of the lab. For example, immediately following a suicide bombing people believe the risk decreases but at the same time exhibit more cautious behavior (Barron & Yechiam, 2009).

An alternative perspective is that experience-based choices do not naturally produce a probabilistic mindset and, thus, are inexplicable by models that require explicit probability representation. Many decisions appear to be made without probabilistic representation, particularly when probabilistic cues are not made salient (Huber, Wider, & Huber, 1997; Rottenstreinch & Kivetz, 2006). Indeed, there are several successful models of choice that do not depend on the explicit representation of probability information (e.g., the natural mean heuristic; Hertwig & Pleskac, 2010). Thus, our second research question was whether the description-experience

choice "gap" can be at least partially explained as a difference in probabilistic representation.

The Experiment

We designed a between-subjects experiment that crossed information format with induced probabilistic mindset to produce four different groups (see Table 1).

Table 1

The experimental groups produced by crossing choice format with induced probabilistic representation

		Mindset	
		Probabilistic	Non-probabilistic
Format	Description	D-Probability	D-Frequency
	Experience	E-Appraise	E-Sample

To investigate our first question, we examined the choices made by participants who received likelihood information in either probability or frequency format. A difference in preferences between the D-Probability and D-Frequency groups would provide evidence consistent with a frequency effect. Specifically, we expected those in the D-Frequency group to more often select the objectively better option, that is, the option with the higher expected value (EV; calculated as the sum of each outcome multiplied by its probability).

To investigate our second question, we additionally examined the choices made by participants who received likelihood information through the experience of sequential sampling, either with (E-Appraise group) or without (E-Sample group) the added obligation to occasionally appraise outcome probabilities (see Method). A difference between the average of the two Probabilistic groups and the average of the

two Non-probabilistic groups would provide evidence consistent with the description-experience gap being at least partially caused by a difference in probabilistic mindset.

Method

Participants

The participants were 100 undergraduate UNSW students (63 females) with a median age of 19 years. Participation was in exchange for course credit plus payment contingent upon the outcome of one randomly selected choice.

Design

The experiment used a 2 (information format: description vs. experience) x 2 (probabilistic mindset: probabilistic vs. non-probabilistic; Table 1) between-subjects design. The dependent variable was the choice in each problem.

Participants in the two description groups were given all information regarding outcomes and their probabilities. Those in the D-Probability group were presented with the percentage chance of each outcome (e.g., “80% chance of 4”) whereas those in the D-Frequency group were presented with the outcome occurrence frequency in forty samples (e.g., “32 out of 40 occasions get 4”).

Participants in the two experience groups had to discover the possible outcomes and their likelihoods by sampling exactly forty times. Participants were given the outcome and probability of the safe option and thus had only to sample from the risky option (cf. Hau, Pleskac, & Hertwig, 2010). The sequence of outcomes was randomly ordered but perfectly matched the description given to participants in the description groups. Those in the E-Appraise group were asked after every ten samples to judge the probability of a zero outcome occurring on the next trial (all

risky options involved a zero outcome; see below). The intent here was to induce a probabilistic representation of the outcome likelihoods. Those in the Sampling group were not required to provide probability estimates, nor were probabilities ever explicitly mentioned. Following all forty samples participants in both the experience groups made a choice regarding which option was preferred.

Materials

Choice Problems. The four choice problems used were taken, with slight modification, from the set created by Hau et al. (2010). Each problem consisted of two options with similar expected values, with at most two outcomes per option. All problems were in the gain domain. The problems were specifically chosen to be able to discriminate between five different choice strategies: risk aversion, risk seeking, adherence to expected value (EV), underweighting of rare events, and overweighting of rare events (see Table 2).

Table 2

Choice option, expected choice pattern under certain strategies, and percentage selecting the risky option

Problem	Choice Options		Expected Choice Pattern Under Strategy				
	Safe	Risky	Risk aversion	Risk seeking	Adhere to EV	Underweight rare events	Overweight rare events
1	3(1.0)	4(.8)	Safe	Risky	Risky	Risky	Safe
2	14(1.0)	15(.9)	Safe	Risky	Safe	Risky	Safe
3	5(1.0)	24(.2)	Safe	Risky	Safe	Safe	Risky
4	3(1.0)	32(.1)	Safe	Risky	Risky	Safe	Risky

Procedure

As the opening scenario makes clear, real-world risky choices are always embedded within a context, which can often provide various grounds, beyond outcomes and their likelihoods, from which to base choice. Thus, each of the four problems was presented within the context of a scenario. Participants' were instructed that their overall task was to maximize the amount of points won from their decisions. Each scenario followed the same format: introduce context, decision problem, measure of success, safe option, and risky option. An example of one scenario inspired by the opening illustration was the following:

> *You are the CEO of a successful multinational computer corporation. One of the most important decisions you make each year is whether or not to acquire and integrate a smaller company into your corporation. Your measure of success is year-end profit. On the one hand, you know that if you do not acquire any other smaller companies, then you will make moderate profits. On the other hand, if you risk acquiring another company then you could make large profits.*

The options in the scenario were then presented (e.g., do or do not acquire a small company) along with information about the possible outcomes and likelihoods as expected from hypothetical previous occasions (e.g., "100% of the time an acquisition was not made, profit was 14"). The problems and scenarios were completely counter-balanced. Participants were not given feedback during the experiment. At the conclusion of each problem, participants typed a response detailing what their choice strategy was.

Results

Description- vs. Experience-based Choice

The percentage of risky choices is shown in Table 3. Since preferences are contingent on whether the rare event is desirable or not, averaging across problems tends to obscure interesting comparisons. Thus, we remapped choices onto a single directional scale by re-categorizing choices in terms of whether the "predicted" option was preferred. The predicted option is the alternative appearing favorable if rare events are overweighted. In practice, this required inverting the percentages reported in the rightmost columns of Table 3 for Problems 1 and 2.

Table 3

Percentage of participants selecting the risky option

Problem	Choice Options		% selecting the risky option			
	Safe	Risky	D-Probability	D-Frequency	E-Appraise	E-Sampling
1	3(1.0)	4(.8)	36	44	60	64
2	14(1.0)	15(.9)	32	36	56	72
3	5(1.0)	24(.2)	52	28	20	32
4	3(1.0)	32(.1)	44	52	48	40

The proportion of participants selecting the predicted option, averaged across problems, is shown in Figure 1. The predicted option was selected significantly more often by those in the two description groups (red bars) than those in the two experience groups (blue bars; 54% vs. 36%; $\chi 2(1) = 12.4$, $p < .001$). Interestingly, this difference was primarily driven by the large difference between the D-Probability and E-Sample groups ($\chi 2(1) = 10.6$, $p = .001$), as opposed to the small

difference between the D-Frequency and E-Appraise groups ($\chi2(1) = 2.9$, $p = .09$). Nevertheless, our data clearly replicated a description-experience choice gap.

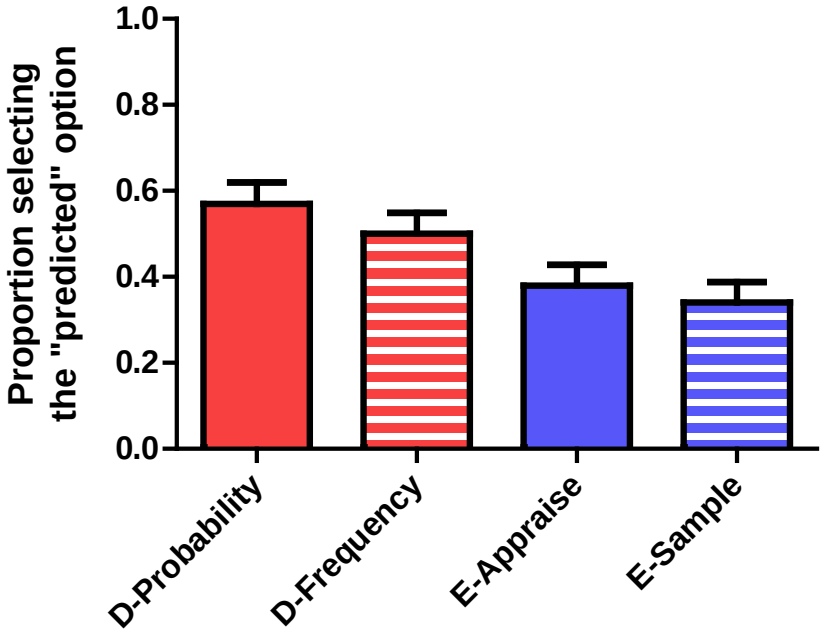


Figure 1: Proportion in each group selecting the option consistent with overweighting of rare events (i.e., the "predicted" option). Red bars represent description and blue bars represent experience. Filled bars represent probabilistic and lined bars represent non-probabilistic.

Probability vs. Frequency Information

Our first research question examined the possibility of a frequency effect in the context of risky choice. Consistent with our hypothesis, those in the D-Frequency group more often selected the option with the higher EV, however, this difference was not reliable (58% vs. 49%; $\chi2(1) = 1.6$, $p = .2$). Moreover, as evident in Figure 2, there was little difference in preference for the predicted option between the D-Probability group (filled red bar) and the D-Frequency group (lined red bar; 57% vs. 50%; $\chi2(1) = .9$, $p = .3$). Our power to detect a difference here with an odds-ratio of 2 was 77.5% (calculated with G*Power3; Erdfelder, Faul, & Buchner, 1996). Thus, our data did not show a clear frequency effect in the context of risky choice.

Probabilistic vs. Non-probabilistic Representation

Our second research question examined the possibility that different probabilistic mindsets may contribute to the gap. As is apparent from Figure 1, there was little difference between the two probabilistic groups (filled bars) and two non-probabilistic groups (lined bars) when averaging across problems (48% vs. 42%, respectively; $\chi 2(1) = 1.2$, $p = .3$). Our power to detect a difference here with an odds-ratio of 2 was 96.0% (Erdfelder et al., 1996).

We sorted participants' choice strategy explanations according to whether they included any of the following terms: chance, odds, percent, %, probability, expected value, likely, and likelihood. Responses that included these words were categorized as adopting a "probabilistic" mindset. In support of our manipulation, more responses were categorized as adopting a probabilistic mindset in the probabilistic groups than in the non-probabalistic groups (55% vs. 31%, respectively; $\chi 2(1) = 24.5$, $p < .001$). Specifically, in each group the proportions of responses categorized as adopting a probabilistic mindset were: D-Probability = 60%, E-Frequency = 35%, D-Appraise = 50%, and E-Sample = 26%. Reanalyzing the data using this classification to assign participant to levels of the independent variable did not change the results (44% vs. 45%, respectively; $\chi 2(1) = .01$, $p = .9$). Thus, our data did not provide any evidence for a probabilistic mindset effect.

Choice Strategies

We compared the choices made by each participant to the expected patterns under the strategies listed in Table 2. As shown in Figure 2, the vast majority of participants (58%) made choices that were inconsistent with any of the strategies. This was most true in the D-Probability group (80%) and least true in the E-Appraise group (36%).

Inspection of Figure 2 suggests a number of interesting, though highly provisional points. As expected, choices consistent with underweighting of rare events (left diagonal bars) were most common in the two the experience groups. Contrary to expectations, choices consistent with overweighting of rare events (right diagonal bars) were not at all common in the two description groups. Interestingly, a strategy that consistently selected the option with the higher expected value (black bars) was relatively more common in the D-Frequency (16%) and E-Appraise (20%) groups.

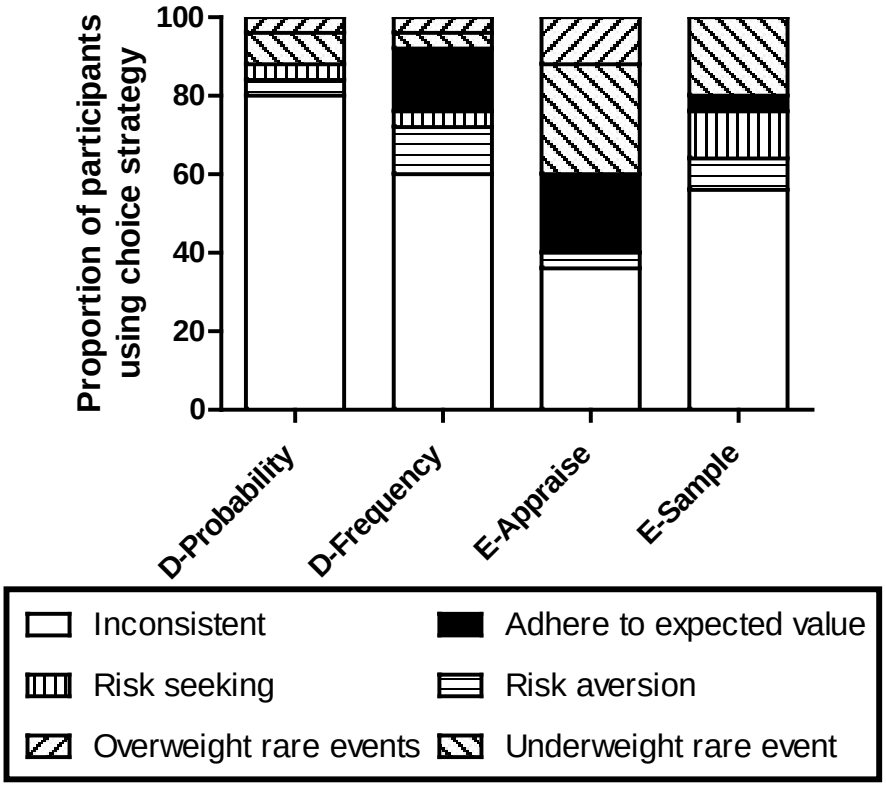


Figure 2: Proportion of participants whose four choices matched a specific choice strategy listed in Table 2.

Memory Order Effects

Following Hertwig et al. (2004), we compared participants' choices with those predicted based on both the first and second half of observed outcomes. For those in the E-Appraise group there was no difference in number of choices correctly

predicted when based on the first or the second half of observations (53% vs. 48%; $\chi2(1) = .5$, $p = .5$). In contrast, for those in the E-Sample group there was a weak primacy effect in that more choices were correctly predicted when based on the first rather than the second half of observations (52% vs. 40%; $\chi2(1) = 2.9$, $p = .09$).

Probability Judgments

The estimated probabilities of the zero outcome, made only by participants in the two probabilistic mindset groups, are plotted against the objective probabilities in Figure 3. For those in the E-Appraise group only the final estimate was used. In general, there was a tendency in both groups to overestimate rare events and underestimate common events. However, estimation error was significantly larger in the D-Probability group than in the E-Appraise group (27.1% vs. 19.1%, respectively; $F_{(1,1,98)} = 12.7$, $p = .054$), suggesting that participants in the experience condition were better calibrated and less susceptible to this judgment error.

A logistical regression with choice made (i.e., predicted option or not) as the dependent variable and presentation format, objective probability, and estimated probability as the independent variables found an effect only for presentation format (B = .79, Wald(1) = 7.4, $p = .007$). Thus, estimated probability was not a good predictor of choice.

Discussion

Consideration of our two research questions in light of the current dataset provide little evidence that the description-experience gap is driven by a probabilistic mindset in the former paradigm but not the latter. Nevertheless, we did make a number of interesting observations that provide valuable input to future work.

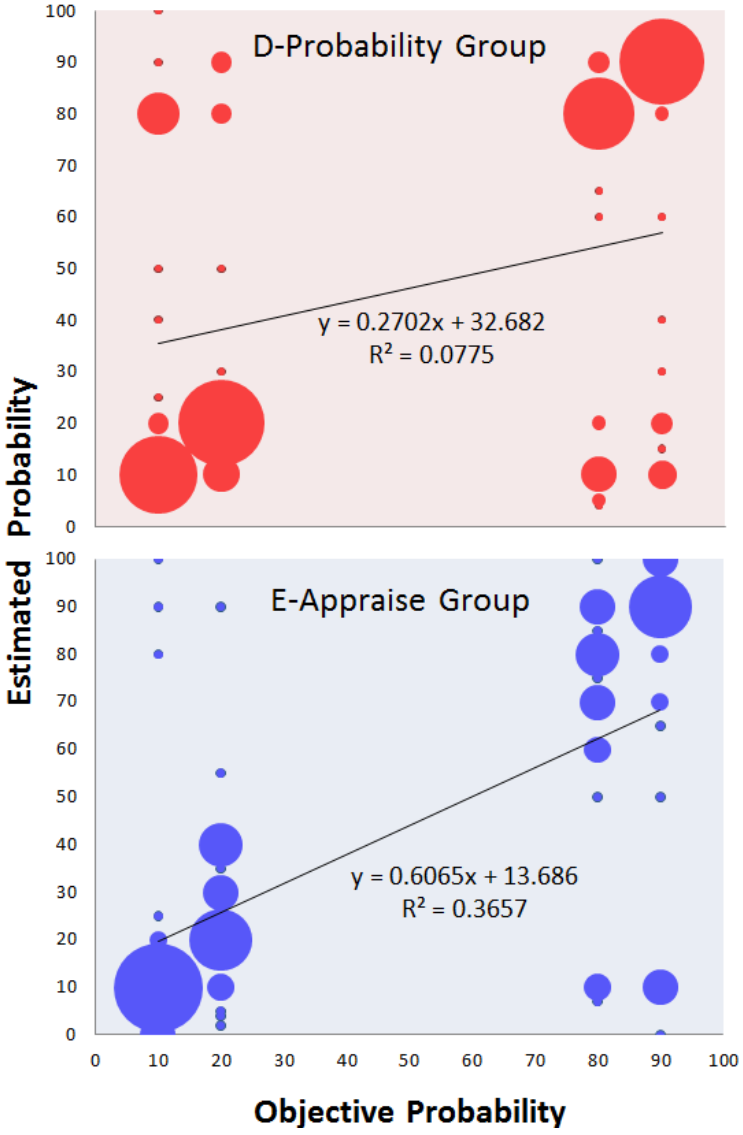


Figure 3: Estimated probability plotted against objective probability for the zero outcomes. The size of the circle indicates the number of identical data points. The solid lines depict the least-square linear regression lines.

With respect to our first research question, we found no difference in preference for the predicted option between those provided with likelihood information in probability format and those in frequency format. However, there was a tendency for participants presented with frequencies to more often adopt a maximization strategy. This finding is consistent with the argument that frequency information is more compatible with the intuitive, evolutionary-based cognitive

"algorithms" that have developed, which can produce better decision-making (Cosmides & Tooby, 1996).

The absence of a clear frequency effect in the current dataset is consistent with the observations made by Rakow et al. (2008) but inconsistent with those made by Gottleib et al. (2007). One potential reason for such inconsistency may be the different designs used: the two studies finding no effect used a between-subjects design whereas the one study finding an effect used a within-subjects design. As Kahneman (2003, pg. 477) notes, the latter "design provides an obvious cue that the experimenter considers every manipulated variable relevant". It is therefore recommended that future studies studying the frequency effect adopt a between-subjects design.

With respect to our second research question, we were able to find a clear description-experience gap even without the influence of external sampling biases (since experienced samples perfectly matched the described distribution). The persistence of the gap implies that it is caused by a number of different contributing factors (Hertwig & Erev, 2009).

Our dataset suggests that adoption of a probabilistic mindset – explicit consideration of outcome probabilities – is not one of these contributing factors. Participants in the description and experience conditions were not greatly influenced by inducing either a probabilistic or non-probabilistic mindset. This null effect is unlikely to be due to an ineffective manipulation, which appeared to be moderately successful when gauged by the content of free responses. However, we were surprised by how infrequently probabilistic terms were mentioned in free response strategy descriptions, especially for those cued with probability estimates. This tendency supports the argument that people are not naturally interested in probability

information (Huber et al., 1997). Future studies should continue to investigate the factors that cause people to prefer probability information (e.g., problem simplicity; Lejarraga, 2010).

There was a greater tendency for those in the E-Appraise group to adopt a maximization strategy (Figure 2). Indeed, the description-experience gap was not reliable when contrasting the E-Appraise group with the D-Frequency group. This observation is consistent with the argument that different information formats each come with a unique set of advantages and disadvantages such that the most effective mode of risk communication may be through multiple formats (Slovic et al., 2000). This strategy may induce "dialectical bootstrapping", that is, reasoning through the exchange of opposing ideas (Herzog & Hertwig, 2009). Future studies could examine whether prompting participants to consider the same information in multiple formats leads to greater maximization.

We detected a primacy effect in the E-Sampling group, indicating that earlier observations had a greater influence on choices than later observations. Since any subset of outcomes tends to under-represent rare events, this internal sampling bias reveals at least one cause of the description-experience gap in our data (Camilleri & Newell, 2011b). Note also that no memory effect was detected in the E-Appraise group where the gap was not reliable.

Primacy is a curious result in that it is opposite to the more common recency effect (e.g., Hertwig et al., 2004). Our hypothesis is that many participants adopt a two-stage sampling strategy whereby earlier samples are used to assess the potential outcomes and later samples are used to assess their likelihoods. Since we told participants what the safe outcome was, it is possible that they moved on to the second stage very quickly and subsequently became bored by the end of the task.

Presumably, those in the E-Appraise group were resistant because they were required to periodically make judgments and therefore remained alert throughout. To test this hypothesis, future studies could experiment with telling participants the number or value of possible outcomes (e.g., Hadar & Fox, 2009).

Some have argued that judgment error may also be implicated as a cause of the gap (e.g., Fox & Hadar, 2006). Consistent with this argument, we found that judgments tended to overestimate rare events and this overestimation was greater for those in the description condition. Worryingly, however, judgments were also incredibly inaccurate, particularly in the D-Probability group where participants had only to remember the recently presented probability. Moreover, and in line with Camilleri & Newell (2009), estimates themselves were unable to predict subsequent choices. These findings challenge the relevance of judgment biases to the choice gap discussion and question the very enterprise of explicitly probing decision makers for outcome probability estimates. Future studies pursuing this issue could experiment with less explicit probes (e.g., Gottlieb, et al. 2007).

References

Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making, 4,* 447-460.

Camilleri, A. R., & Newell, B. R. (2009). The role of representation in experience-based choice. *Judgment and Decision Making, 4,* 518-529.

Camilleri, A. R., & Newell, B. R. (2011a). When are why rare events are underweighted: A direct comparison of sampling-, feedback- and description-based choice. *Psychonomic Bulletin & Review.*

Camilleri, A. R., & Newell, B. R. (2011b). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica*.

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

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods Instruments & Computers, 28,* 1-11.

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.

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18,* 240-246.

Hadar, L. & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making, 4,* 317-325.

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.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23*, 48 - 68.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20,* 231–237.

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.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*, 517-523.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*, 225-237.

Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica, 95*, 15-29.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-91.

Lejarraga, T. (2010). When experience is better than description: time delays and complexity. *Journal of Behavioral Decision Making, 23*, 1, 100–116.

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.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*, 1-14.

Rottenstreich, Y., & Kivetz, R. (2006). On decision making without likelihood judgment. *Organizational Behavior and Human Decision Processes, 101*, 74–88.

Slovic, P., Monahan, J., & MacGregor, D. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior, 24*, 271-296.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted, when rare outcomes are experienced (rarely)? *Psychological Science, 20*, 473-479.

Whitney, L. (2010). *Android hits top spot in U.S. smartphone market*. Accessed 19th December 2010 from http://news.cnet.com/8301-1035_3-20012627-94.html.

Modeling Probability Estimates and Choice in Decisions from Experience

Guy Hawkins[1]
Adrian R. Camilleri[2]
Ben R. Newell[2]
Scott Brown[1]

[1] School of Psychology, University of Newcastle, Newcastle, Australia
[2] School of Psychology, University of New South Wales, Sydney, Australia

Abstract

In most everyday decisions, we learn about the outcomes of alternative courses of action through experience: a sampling process. Current models of these decisions from experience emphasize the importance of predicting choice at the expense of explaining how the sample outcomes are used to form a representation of the distribution of outcomes. Moreover, most current models fail to generalize even across quite similar forms of experience-based choice. We sought to overcome these limitations by changing the focus to the general class of instance (or exemplar) based models. We tested three instance-based models varying in complexity: the *k*-sampler, the Instance-Based Learning (IBL) model, and a new model, the Exemplar Confusion (Ex-CON) model. The models were evaluated against data collected from a new experiment and also the existing comprehensive Technion Prediction Tournament (TPT) data set. With a new experiment, we directly investigated a key element of decisions from experience: the estimation and representation of outcome probabilities. The IBL and Ex-CON models simultaneously predicted both probability estimates and choice behaviour in the new experiment, while the *k*-sampler could not account for probability estimates. All models performed well in predicting choice in the TPT. We conclude that two elements appear to be important in modeling experience-based choice: instance-based memory, and a reliance on recent samples caused by memory noise.

Modeling Probability Estimates and Choice in Decisions from Experience

An enduring assumption in models of human choice is that behavior can be described as if people multiply some function of the probability of an outcome by that outcome's value, and then maximize. This framework dates back to Bernoulli (1738/1967) and has undergone many modifications, particularly since Neumann and Morgenstern (1947) provided an axiomatization for rational choice, but the fundamental idea that people maximize expected utility remains in many successful models of choice. Modern examples include Prospect Theory (Kahneman & Tversky, 1979), Regret Theory (Loomes & Sugden, 1982), Cumulative Prospect Theory (Tversky & Kahneman, 1992), and Security-Potential/Aspiration theory (Lopes & Oden, 1999)[1]. Much of the research undertaken in developing and modifying these models has relied on "decisions from description", in which choices are made from explicitly stated outcomes and their associated probabilities (Barron & Erev, 2003; Rakow & Newell, 2010). This focus is understandable given that the primary interest of many models is in capturing the systematic cognitive distortions of utilities and outcome probabilities implied by people's choices (e.g., choices implying that people attribute more weight to low-probability events than their objective probability of occurrence warrants; Prelec, 1998).

However, this focus on decisions in which outcome and probability information is readily available neglects important aspects of the cognitive processes that must underlie many of the decisions we face in our daily lives – decisions for which probabilities and outcomes are not explicitly provided. In such "decisions

[1] For a model that does not adopt this framework see Brandstatter, Gigerenzer, & Hertwig (2006). Their Priority Heuristic eschews the notions of weighting and summing, but its success as a general model of choice has been challenged (see Birnbaum, 2008; Johnson, Schulte-Mecklenbeck, & Willemsen, 2008).

from experience" (Hertwig & Erev, 2009; Rakow & Newell, 2010), decision-makers must explore their environment to establish both the range of potential outcomes and the probability with which each occurs. Thus, to understand how decisions from experience are made it is crucial to examine how the probability distribution across potential outcomes is estimated, and represented in the mind of the decision-maker.

We suggest that a general class of exemplar models can not only predict people's choices in various experienced-based choice paradigms, but also describes the process by which people construct and represent the probability distribution upon which those choices are based. We first outline two common experience-based choice paradigms – the feedback and sampling paradigms – and briefly review some key findings. Next, we describe our modeling approach and highlight the critical elements of the class of models that we consider: *instance memory* and *noisy storage*. We then report data from a new experiment specifically designed to test variants of these models, followed by tests of the models against an important data set (the Technion Prediction Tournament data; Erev et al., 2010). Finally, we discuss some general conclusions and implications of our results for understanding the processes underlying experience-based choice.

Feedback- and Sampling-Based Decisions from Experience

Experience-based choices have primarily been studied using the *feedback* and *sampling* paradigms. In the feedback paradigm, the decision-maker is presented with (typically two) alternative options and encouraged to sample outcomes from each option in any order (e.g., Barron & Erev, 2003). The standard interface is a computer screen with two unlabelled buttons that can be clicked on. Each sample (or click) briefly reveals a single outcome, randomly selected with replacement, from a hidden outcome distribution associated with that option. As with economic choice

experiments in general, the distribution is usually very simple, comprising of just one or two outcomes. The decision-maker is encouraged to sample from both options to learn the outcomes available and also the probability with which each outcome occurs. Crucially, each choice is consequential; it has an associated payoff such that each sampled outcome adds to a running total that is constantly displayed to the decision-maker. The decision-maker is not informed how many samples will be granted but is encouraged to earn the highest score. The decision-maker is therefore faced with a tension between the objectives of learning more about the options (to "explore") while also trying to maximize earnings across an unknown number of repeated, consequential choices (to "exploit"). The decision-maker's preference is usually inferred as the most frequently selected option in the final block of trials.

In contrast, the typical sampling paradigm separates the goals of exploring and exploiting the options into two distinct phases (e.g., Hertwig, Barron, Weber, & Erev, 2004). During the sampling (or "exploration") phase, the decision-maker is encouraged to sample outcomes from each option in any order. Importantly, each sampled outcome during this phase is without financial consequence and purely for the purpose of learning the outcome distribution associated with the option. At any point during the sampling phase the decision-maker can elect to stop sampling and move on to the choice (or "exploitation") phase. During the choice phase, the decision-maker selects the option that they prefer with the goal of earning the highest score. The outcome of this single choice is added to a running tally that is hidden from the decision-maker until the end of the task.

A focus of recent literature has been on the extent to which the different ways of learning about probability distributions across outcomes leads to different patterns of choice. Early work examined differences between experience- and description-based

choices, and often revealed preference reversals in the two formats (e.g., Barron & Erev, 2003; Hertwig et al., 2004). Specifically, description-based choices implied *over*weighting of low-probability outcomes whereas experience-based choices implied *under*weighting of low-probability outcomes (Camilleri & Newell, 2011b; Ungemach, Chater, & Stewart, 2009). However, more recent work has highlighted that such differences can be reduced or eliminated, at least in the sampling paradigm, when samples are equated across formats (Rakow, Demes, & Newell, 2008) or forced to be representative of the underlying distributions (Camilleri & Newell, 2011a).

The observation that preferences are similar across description and experience paradigms when outcome distribution knowledge is controlled suggests that the real point of difference between these two formats might be the way distributional knowledge is estimated and represented, rather than how it is utilized to make decisions (Camilleri & Newell, 2011a). Indeed, the overall picture that emerges from this literature is that decisions from description and from the various forms of experience might best be viewed as spread along a continuum of uncertainty (Hau, Pleskac, & Hertwig, 2010; Rakow & Newell, 2010, cf. Knight, 1921). At one end of this continuum are the situations in which probabilities and outcomes are precisely specified (e.g., gambling on the toss of a fair coin), while at the other extreme are situations with so many unique characteristics that estimation alone must be relied upon (e.g. the probability that a new business venture will succeed). Decisions from experience inhabit the interesting middle ground of this continuum, where precise probabilities are harder to calculate, and potential outcomes more difficult to identify than for a simple coin-toss, but easier to ascertain than in one-off situations with unique features. Despite these differences in uncertainty about outcome distributions,

the process that is arguably common to all paradigms is the need to combine probability information with outcome values to make a choice.

Approaches to Modeling Experience-Based Choices

Several approaches have been taken to model experience-based choices. The most promising of these approaches were showcased in a recent choice prediction competition called the Technion Prediction Tournament (TPT; Erev et al., 2010). The organizers of the competition collected two comprehensive datasets encompassing a broad range of problems in the description, sampling, and feedback paradigms. The estimation data set was made public and researchers were invited to submit a model that was later tested against the unseen competition data set.

The submitted models can be categorized into four different approaches. First, Bernoullian-inspired models that combine weighted outcome and probability information. The most well-known example of this approach is cumulative prospect theory (CPT; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). A stochastic version of CPT (SCPT) won the TPT description competition. Second, choice heuristics that yield preference by relying on simple rules and limited information, often ignoring probability information. An ensemble model comprising of SCPT combined with three heuristics, including two versions of a natural mean heuristic that simply preferred the option with the highest observed mean, won the TPT sampling competition. Third, associative-learning models that eschew probability representation and instead rely on feedback and reinforcement of option attractiveness (Sutton & Barto, 1998). A normalized reinforcement learning model performed very well in the TPT feedback competition. Fourth, instance-based models that represent specific instances comprising of the situation, action, and result of a decision in memory, and then calculate value based on frequency and recency

(Anderson & Lebiere, 1998). An instance-based model built upon the ACT-R cognitive architecture with sequential dependencies and blended memory won the TPT feedback competition.

In the TPT, the sampling and feedback paradigms were procedurally similar and indeed produced preferences that were strongly correlated (e.g., r = .84 in the TPT data set; Erev et al., 2010). Nevertheless, the successful models in the sampling and feedback competitions were markedly different and no model performed well across the different experience-based choice paradigms. The difficulty in finding a model that can explain behavior in both tasks is surprising given that the paradigms are so similar, and it reflects a larger issue in cognitive science: the development of highly specific models that are only successful in the one task that they were designed to account for (Cassimatis, Bello, & Langley, 2010). A growing awareness of this issue has prompted the beginnings of a shift in thinking towards the development of more generalizable models that can be applied to a range of similar tasks (e.g., Anderson & Lebiere, 2003; Gonzalez & Dutt, in press). In line with this argument, we suggest that successful models of choice should generalize across different contexts that share many of the same underlying cognitive features. To achieve this goal, we turn to an approach that we believe is the most promising: instance-, or exemplar-based, models.

An Exemplar-Based Approach to Modeling Probability Estimation and Choice

The results of the TPT competition highlight the difficulty in discriminating between models using existing data and approaches. We propose that additional model constraint might come from two sources. First, we propose that a successful model should simultaneously account for observers' choices as well as their knowledge about outcome distributions. Second, a comprehensive model should

account for behavior in sampling and feedback paradigms with similar mechanisms and similar parameter settings. These constraints make sense if one assumes that decisions from description and experience inhabit different points on the continuum of uncertainty described above.

One of our aims is to demonstrate the value of modeling qualitatively similar tasks with qualitatively similar models. The "instance-based learning" (IBL; described below) model is a leading example of such an initiative that has been generalized to the feedback paradigm, a probability learning task, a repeated binary choice task within a changing environment (Lejarraga, Dutt, & Gonzalez, 2011) and, with some modification, to the sampling paradigm (Gonzalez & Dutt, in press) and even decisions from experience in market entry games (Gonzalez, Dutt, & Lejarraga, 2011). We adopt a simplified version of the IBL model's approach, and argue for a higher-level conclusion: that exemplar-based memory mechanisms provide general mechanisms for representing and learning about the outcome distributions of lotteries, as well as the final choice process. We hope that focusing on this conclusion might move the field ahead more quickly than continued attempts to select a single best model to account for certain data sets.

Exemplar (or instance – we use the terms interchangeably) models assume that observers record a memory trace each time they encounter a stimulus. This memory trace might include information about the stimulus, the corresponding feedback provided by the experimenter, and the general context in which the stimulus was encountered. Later, the properties of an unfamiliar stimulus can be predicted from the properties of related exemplars stored in memory. Exemplar models are not new in economic decision making – two leading exemplar models are the simple *k*-sampler model (Erev et al., 2010) and the IBL model (Gonzalez & Dutt, in press).

However, exemplar models have a much longer history in explaining many aspects of categorization behavior (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). They have also been extended to predict response times in decision-making tasks (Nosofsky & Palmeri, 1997), performance in absolute identication (Kent & Lamberts, 2005), and short-term memory (Nosofsky, Little, Donkin, & Fific, 2011). We argue that the storage of exemplars in memory provides a simple and psychologically plausible way to explain the estimation of probability information about gambles in risky choice problems. Such an argument is not without precedent, for example in studies of categorization. Estes, Campbell, Hastopoulos, and Hurwitz (1989) observed "base rate neglect", in which people acted as if they overestimated the likelihood of examples from rare categories, and underestimated the likelihood of examples from common categories. Exemplar-based models have proved successful in accounting for these effects, amongst others (e.g., Kruschke,1992; Nosofsky, Kruschke, & McKinley, 1992).

Evaluating exemplar-based models simultaneously against choice preferences and probability estimates requires data on both measures. A number of previous studies have asked decision-makers to estimate outcome probabilities, and found that estimates are either well calibrated (Fox & Hadar, 2006) or that rare events are overestimated (Camilleri & Newell, 2009; Gottlieb, Weiss, & Chapman, 2007; Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig, Pachur, & Kurzenhauser, 2005; Ungemach et al., 2009).. However, one limitation of these previous studies is that the elicited probability estimates have served only as inputs to models (e.g., CPT; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) rather than dependent variables used to constrain the model. A second limitation is that the assessed outcome distributions were very simple -- just two outcomes per option -- which

severely limited their potential to constrain any model, even if attempted (Reiger, 2003).

Our experiment addresses these limitations, providing data on both preferences and probability estimates. We used choices between pairs of the six lotteries described by Lopes and Oden (1999). Each of these lotteries contained five outcomes and had the same expected value but differed in the distribution of payoffs around the mean (including outcome variance, see Figure 1). We favored these five-outcome lotteries over the more standard two-outcome (safe vs. risky) lotteries typically used in the experience-based paradigms because they offered a richer set of probability estimation data, allowing decision-makers to provide ten estimations per gamble problem (two lotteries, each with five outcomes).

Method

Participants

The participants were 107 university undergraduate students from two Australian universities. The University of New South Wales participants were recruited from the first year psychology participant pool and were rewarded with course credit plus monetary payment contingent on the choices made during the experiment. In practice, all payments were rounded up to a value of $10. The University of Newcastle participants were recruited via the SONA voluntary register and noticeboards, and were incentivized with monetary or gift certificate payment, some of which was contingent on the choices made during the experiment ($10 -- $20). Data from two participants were excluded because of failure to follow instructions.

Design

Participants were randomly allocated to either the sampling task or the feedback task. Participants always made choices between two competing lotteries; we denote such a pair of lotteries as a "problem". The primary dependent variables were participants' preferences in each problem, and also their estimates of the probabilities of the outcomes from the lotteries in the problem. In the sampling group, lottery preference was operationalized as the one-shot choice made after learning about the lotteries in the free sampling phase. In the feedback group, lottery preference was operationalized as the deck selected most frequently in the final 50 samples.[2] From the six different lotteries defined by Lopes and Oden (1999) there were 15 possible pairings (ignoring order, and without identical choices), and participants in the sampling group played each of these 15 pairings once, in 15 problems (with the order randomized across participants). In order to equate the length of the experiment, participants in the feedback group played a random sample of just six of the fifteen problems. Data were excluded from participants who failed to sample from one of the lotteries during any problem. Some participants did not complete all problems during the one hour experiment and therefore not all problems have equal sample sizes. There were 40 participants in the sampling group who experienced a total of 528 problems, corresponding to 32-39 participants experiencing each problem. There were 67 participants in the feedback group that experienced a total of 386 problems, corresponding to 22-32 data points for each problem.

[2] Similar results were obtained when we used the mode across all 100 samples, as well as just the final sample. We preferred the mode of the last 50 samples because we assumed many trials near the beginning of the task would be for the purpose of exploration rather than an indication of preference. Additionally, we also assumed that participants would intermittently take reminder samples from their non-preferred option towards the latter part of each problem as study for the probability estimation task.

Materials

Each problem consisted of a choice between two lotteries. We used six lotteries in total, taken from Lopes and Oden (1999). Each lottery was associated with five possible outcomes that ranged from between $0 and $348. As shown in Figure 1, the outcome distribution for each lottery was unique but all had expected values close to $100. Although the lotteries were unlabelled during the experiment, for purpose of discussion we adopt the lottery names used by Lopes and Oden (1999): Riskless, Rectangular, Peaked, Bimodal, Short Shot or Long Shot, depending on the specific distribution of outcomes.

Figure 1: Visual representation of the outcome distribution for the six lotteries, adapted from Lopes and Oden (1999). Note that the labels were not presented to the participants.

Procedure

The experiment took on average one hour to complete. After signing consent forms, the participants were presented with instructions indicating that they would face a choice between two decks of cards that were displayed on screen. Participants were instructed to use their choices to earn as much money as possible. We

randomized the allocation of lotteries to the left and right decks of cards on screen, and the order of the problems. Both the sampling and feedback versions of the task presented participants with two unlabelled images of decks of cards. Each deck was associated with a distribution of outcomes from the lotteries of Lopes and Oden (1999; the names of the lotteries were not presented). When the participant clicked on a deck, an outcome was selected randomly from the associated lottery, with replacement, and displayed briefly as if the participant had turned over a playing card. We call the act of selecting a deck and observing an outcome a “sample”.

The sampling task began with an exploration phase in which each sample occurred without consequence, so the participant could learn about the lotteries. The participants were free to terminate the exploration phase at any time, after which they made a single and final choice indicating their preference. The outcome of this final choice was added to the participant's running total for the experiment. In the feedback task, each problem granted 100 samples before moving on to the next problem. Each of the 100 samples was consequential: the outcome of each sample was added to the participant's running score, which was constantly displayed on screen.

After making a choice (in the sampling paradigm) or completing 100 samples (in the feedback paradigm), participants were asked to estimate the probability of different types of cards that were in each deck, corresponding to the probability of the different outcomes in the lotteries. As shown in Figure 2, six different outcome values were presented beside adjustable sliders. Participants were required to move each of these sliders to a point between 0 and 100 indicating the estimated percentage. The default starting estimate was 0. One of the six outcomes was a “foil” that was not an outcome from the deck in question, but was selected from one of the

other decks. The foil card was included to identify participants who may not have been paying attention to the sampling process, yet learnt the structure of the probability estimate task. For instance, if a participant only observed a few samples then they may not have seen the five possible outcomes from the lottery. If this participant then learnt that there were always five outcomes in each lottery, based on the presence of five outcomes on the probability estimates screen, then they ought to always provide a non-zero probability rating for all five outcomes, a strategy circumvented by the use of a foil outcome. Participants were only allowed to continue to the next problem when the sum of the sliders for each deck equalled 100%. Also note that we specifically asked participants to estimate the probability of each of the different cards being observed on some future (hypothetical) sample -- this is different from asking for estimates of the sample frequencies of the outcomes they experienced. At the conclusion of the experiment participants were debriefed, thanked, and paid.

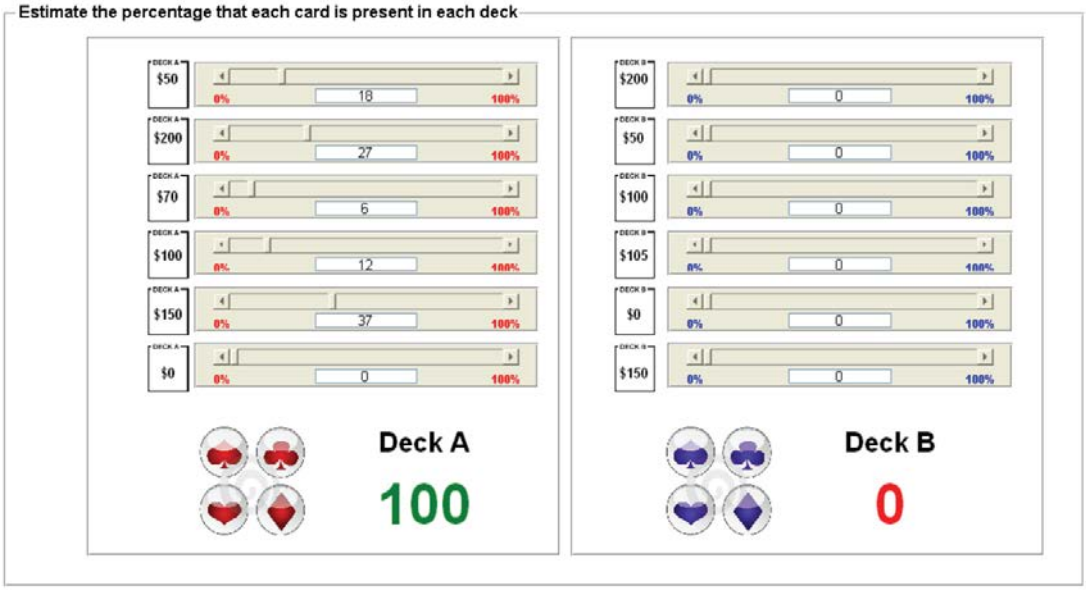


Figure 2: Screenshot of the estimation task after the sliders were adjusted for Deck A.

Results

Preferences

To gauge the extent to which the participants' actual samples from the outcome distributions differed from the population, or programmed, outcome distributions, we calculated the absolute difference between objective probabilities and the sample frequencies, averaged across outcomes. Participants' sampled frequencies were closer to the objective probabilities in the feedback than in the sampling group ($t(912) = 12.02$, $p < .001$). This makes sense given that the median number of samples taken in the sampling group was half the number of samples taken in the feedback group[3].

Figure 3 displays the percentage of participants preferring each lottery, averaged over problems in which the lottery was presented. Preferences were similar between the feedback and sampling groups except for the rectangular and bimodal lotteries, where participants in the feedback group showed a relatively stronger preference for the rectangular lottery ($\chi^2_{(1, N = 330)} = 4.89$, $p = .027$) but a relatively weaker preference for the bimodal lottery ($\chi^2_{(1, N = 319)} = 7.07$, $p = .008$)[4]. There was a strong, positive correlation between preferences in the feedback and sampling paradigms ($r = .64$, $p = .01$), which is consistent with previous studies where participants were free to choose their own stopping point in the sampling task (Erev et al., 2010). Preferences predominately favored the lotteries that minimized or eliminated the possibility of obtaining zero (i.e., the riskless, peaked, and shortshot decks), which is consistent

[3] The median number of samples was 51, which was still considerably higher than previous experiments with similar incentives (e.g., 15-19 as reported in Hau et al., 2010).

[4] Additionally, despite a number of important differences between our design and the one used by Lopes and Oden (1999), the lottery preferences were qualitatively similar between the studies. The major points of departure were that our participants showed greater overall indifference and also less preference for the riskless lottery.

with the idea of a negatively accelerated utility function for gains, as assumed by many theories of choice.

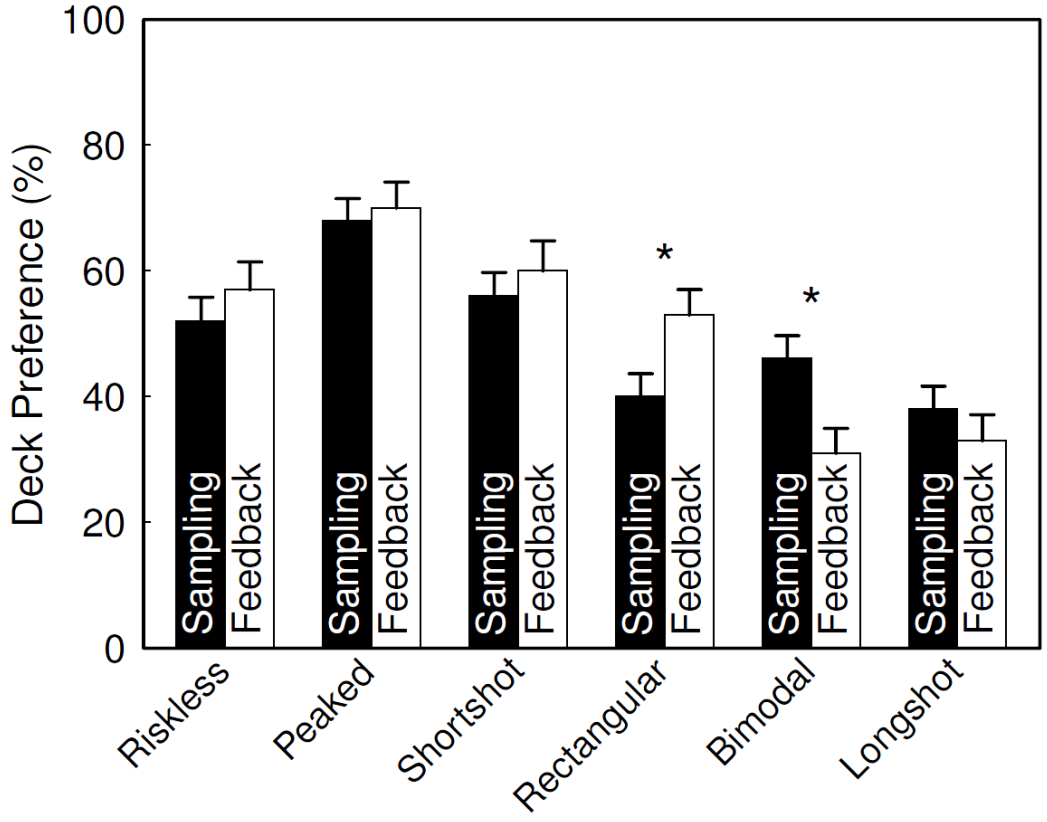


Figure 3: Percentage of participants preferring each lottery, averaged over problems.

Preferences between pairs of lotteries are displayed in Table 1. Each score indicates the average preference for the column-named lottery over the row-named lottery, and asterisks denote preferences in which a lottery was significantly preferred over indifference (i.e., 50-50) by a *z*-test. For example, in the first row of Table 1 the 79 value is asterisked, which indicates that a significant proportion of participants preferred the peaked lottery over the longshot lottery in the sampling task. Only a few such contrasts showed significant differences in preference between the groups. This is to be expected, given the lotteries all had identical expected values. The few contrasts that do show significant differences provide evidence of underweighting of rare events. For example, participants strongly favored the

lotteries where the rare event was highly undesirable (e.g., $0 the peaked and shortshot lotteries) over decks where the rare event was highly desirable (e.g., $348 in the longshot lottery).

Table 1

*Percentage of participants selecting the column-named lottery over the row-named lottery, separately for both experimental conditions. Abbreviations refer to deck type: RL = riskless, PK = peaked, SS = shortshot, RC = rectangular, BM = bimodal, LS = longshot. * p < .05 by (unadjusted) z-test*

Sampling						Feedback					
	RL	PK	SS	RC	BM		RL	PK	SS	RC	BM
LS	60	79*	75*	46	49	LS	62	72	88*	63	55
BM	55	69	59	39		BM	82*	70	88*	62	
RC	59	64	58			RC	59	79*	56		
SS	47	66				SS	44	70			
PK	39					PK	39				

Probability Estimates

We first confirmed that participants' estimates of the probability of foil outcomes (i.e., outcomes that were not part of the lottery) were accurate. Across participants and problems, foil cards were correctly assigned zero probability 62% of the time. On the remaining 38% of problems in which the foils were assigned some non-zero probability, this estimate was still small: 11% on average. Since the foils were mostly well identified by participants, we do not analyse those data further.

The probability distribution estimates showed that people were quite good at estimating the probabilities of the lottery outcomes. For example, Figure 4 graphs the median probability estimate assigned to outcomes against the actual sample

frequency of the outcome in the samples observed by participants (the whiskers show the 5th and 95th percentiles of these estimates, calculated across participants and problems). In both conditions, the median estimate assigned to outcomes increased almost always with increasing sample probability. However, there was a tendency to overestimate the probability of rare outcomes and underestimate the probability of frequent outcomes, shown by the inverted-S shapes in Figure 4. This pattern appeared in both sampling and feedback paradigms, and was also almost unchanged if we instead graphed the probability estimates against the *population* probability of the outcome (i.e., the proportion of times it would appear in the long run, defined by Lopes & Oden, 1999) rather than the sampled frequency of the outcome (i.e., the proportion of times it really did appear, in the samples observed by participants).

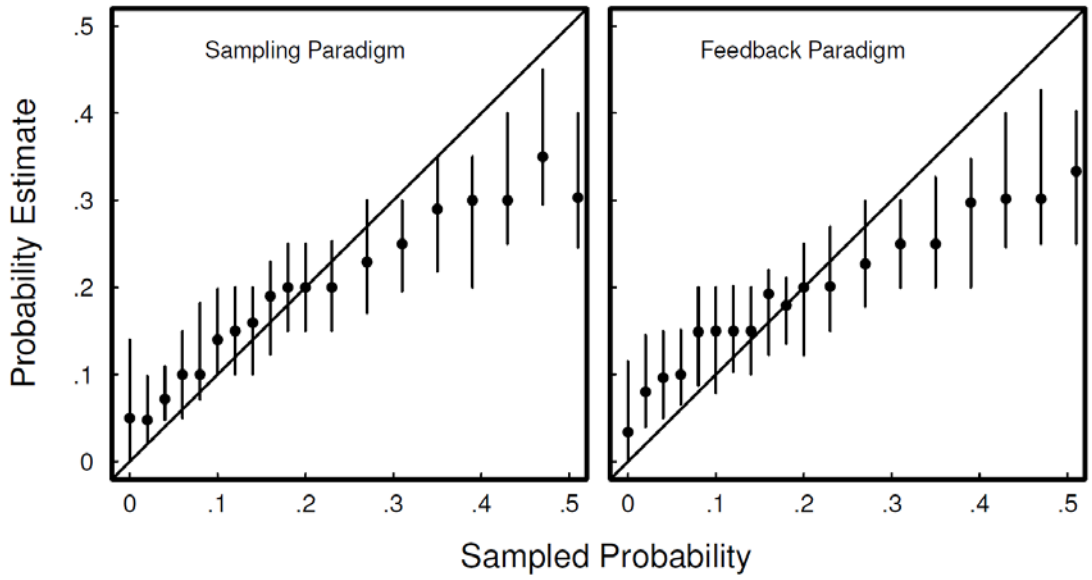


Figure 4: Probability estimates from participants (y-axes) against the sampled probability for the corresponding outcome (x -axis) in the sampling (left panel) and feedback (right panel) conditions. The circles are medians, the whiskers show 5th and 95th percentiles across participants. The diagonal y = x *line indicates the trend that probability estimates would follow if they were perfectly reflective of the sampled probabilities.*

We confirmed the statistical reliability of differences between the pattern of estimated versus sampled probabilities and the $y = x$ line (which would indicate

probability estimates that perfectly reflect the sampled probabilities) using the Wald-Wolfowitz (or "runs") test. This analysis examines the sign of successive residuals, on the assumption that -- under the null hypothesis of perfect calibration – residuals should be randomly distributed either side of zero. The Wald-Wolfowitz test indicated highly significant non-random scatter around the $y = x$ line in both paradigms, both p's < .001, reflecting the run of positive residuals for low sample probabilities and negative residuals for high sample probabilities.

A mundane explanation for the inverse-S shape of the probability estimates is a simple mixture of participants, some who were perfectly calibrated (no inverted-S shape) and others who were simply not engaged in our task. If a participant was not paying attention to the task, then their probability estimates would be unrelated to the observed sample probabilities, and their graphs would show horizontal lines at $y = .2$ (or $y = .17$, if the foil cards were considered). It is possible that a mixture of these two types of participants could lead to the inverted-S shapes, even if no individual participant displayed such a pattern. We tried to rule out this explanation by removing many participants to leave us with only the very best (those most likely to be engaged in the task), and evaluating whether the inverted-S shape was still present.

To this end we separately implemented two very strict inclusion criteria, hoping to keep only the best participants, and then re-examined the probability estimates. The first check excluded any participant who sampled fewer than 20 outcomes from either lottery on any game, which is considerably larger than the median number of samples typically observed across *both* lotteries in experience-based choice tasks (e.g., Hau et al., 2010). This criterion excluded many more people from the sampling task, leaving only 10 of 40 (25%) participants, compared to the feedback task, which

left 42 of 65 (65%) participants. Beginning again with all data, our second, and harsher, exclusion criterion removed any participant who estimated a non-zero probability for a foil card from either lottery for *any* problem, which left only 5 of 40 (13%) participants in the sampling condition and 10 of 65 (15%) in the feedback condition. For both of these strict exclusion criteria the inverse-S trend in probability estimates remained (graphs not shown), with the Wald-Wolfowitz test indicating highly significant departures from the $y = x$ line for both the sampling and feedback conditions for the two exclusion criteria, all p's < .001. Thus, it appears that even the best participants had a tendency to overestimate the probability of rare outcomes and underestimate the probability of frequent outcomes.

Simultaneously Accounting for Choices and Probability Estimates with Exemplar-Based Models

Exemplar (or instance) models assume that observers record a memory trace each time they encounter a stimulus, and later use these traces to make inferences about their experiences. In this section we describe three exemplar-based models that might simultaneously account for the choice preferences and probability estimates in our data. We then examine how well these models account for choice behaviour in simpler (more standard) problems, using the TPT model competition data from erev-ert-roth-etal_2010. One of our aims is to demonstrate that the details of specific models should not be the primary focus, but rather that exemplar-baed models in general provide a good foundation for economic decision making.

A Simple Model: k-*sampler*

Perhaps the simplest exemplar model is the primed sampler model described by Erev et al. (2010), which we refer to as a *k*-sampler. Our instantiation of the *k*-

sampler simply drew *k* of the samples shown to participants (with replacement), for each lottery, and then chose the lottery with the greater sample mean.

A Slightly More Complex Model: Exemplar Confusion

In view of shortcomings in the basic *k*-sampler (see the section on Probability Distribution Estimates below), we developed a related model that we call the exemplar confusion (Ex-CON) model. In this model, the *k*-sampler's limit on memory *capacity* (*k*) is replaced by a limit on memory *accuracy*. Memory accuracy, or forgetting, can be modeled in a number of ways, such as dropping exemplars from memory, or degrading the information content of exemplars. We instantiate memory imperfection with the latter process by including a sample-by-sample confusion process. Confusion in the Ex-CON occurs through interference -- that is, with the passing of events, rather than the passing of time alone -- an assumption which has precedence in the memory literature (e.g., Lewandowsky, Duncan, & Brown, 2004; Lewandowsky, Geiger, & Oberauer, 2008).

As in the *k*-sampler, we assume that there are two memory stores, which begin empty but this time have no limit in size; the two stores correspond to the two lotteries being evaluated. Each time a sample outcome is drawn from a lottery, a memory trace is added to the appropriate store. The only assumption we make about this memory trace is that it stores the outcome value for the observed sample. Each time a new sample is added to a store, the confusion process operates, which leads to a small chance of mixing up the exemplars. Each stored exemplar has a fixed probability (α) of having its outcome value confused – that is, substituted with the outcome value from another exemplar. If an exemplar's outcome value is confused, the new outcome value assigned to that exemplar is chosen uniformly from the list of

all labels in the store[5]. The α parameter governing the exemplar confusion process is the sole free parameter of the model to be estimated from data. After the sampling process for a problem has finished, the preference predicted by Ex-CON is determined by whichever set of exemplars has the highest average utility. The utility function we implement is one of diminishing marginal utility, specifically, we use the utility function for gains specified by Lopes and Orden (1999): $u(x) = x^{0.551}$. Thus, at the core of the Ex-CON is the expected utility theory assumption of multiplying some function of probability with an outcome value, and then maximizing.

The Ex-CON's assumption of a limitless store may seem psychologically implausible, however it can equally well be expressed in terms of a finite-memory system. In such an interpretation, rather than storing a new exemplar for each sample, the model simply updates the total count of each kind of exemplar, leading to a memory load of just a few numbers (however many outcomes exist in the lottery). Under this interpretation, the confusion process described above is instead expressed as a confusion between the sizes of the counts. Note that this reinterpretation makes no predictive difference from a modeling perspective (i.e., the model is mathematically equivalent), however, there are psychological implications. For instance, this re-description of the Ex-CON model casts it as a count-storing model rather than an exemplar-based model.

[5] The Ex-CON as implemented assumes that exemplars can only be confused with outcomes from the same lottery. An alternative model allows confusion of exemplar labels with outcomes from both lotteries. Note that this alternative version of the model only matters when different outcomes are contained in each lottery, which occurred in our experiment in less than half of the problems. Nonetheless, we investigated this alternative, and found that all model fits, parameter estimates and general conclusions were essentially unchanged, so we do not report it below.

A More Complex Model Still: Instance Based Learning

The IBL model has successfully explained choice behaviour in experience-based paradigms (Gonzalez & Dutt, in press; Lejarraga et al., 2011), as well as several other choice paradigms. Since the conclusion of the TPT (Erev et al., 2010), the IBL model has been shown to outperform all of the competition models in both the sampling (Gonzalez & Dutt, in press) and feedback (Lejarraga et al., 2011) paradigms. The IBL model goes further than the *k*-sampler or Ex-CON models by also describing the processes involved in choosing samples from the lotteries, and the corresponding tension between exploration and exploitation. This is accomplished by two extra model components: a "swapping" rule, which describes when to switch sampling from one lottery to the other; and a "stopping" rule, which describes when to stop sampling and make a final, consequential choice (in the sampling paradigm). The swapping and stopping rules used in the IBL are admirably simple and generally applicable – they could easily be exported to either the *k*-sampler or Ex-CON, for example.

To more precisely evaluate choice behavior, we evaluate the performance of the IBL model without the swapping and stopping rules, as far as possible. Instead, we provide the model with the exact sequences of outcomes sampled by our participants, as we did for the other two models (see next section for details). By doing this, we have effectively endowed all three models with omniscient swapping and stopping rules. For rigourous evaluation of the swapping and stopping rules in the IBL model, see Gonzalez and Dutt (in press), and Lejarraga et al. (2011). As an initial check that inclusion or exclusion of the stopping and swapping rules did not greatly alter a model's predictions for choice preferences, we re-ran our simulations of the Ex-CON model with these rules incorporated using the data from our experiment. The results

were very similar – but a little noisier – than without the rules, for both choice preferences and probability estimates.

After sampling the lotteries, the IBL makes a decision by selecting the lottery that has the highest blended value. The blended value is the total of all observed instances from that lottery, weighted by the probability of retrieving each instance. Retrieval probability depends on each outcome's frequency of occurrence and also the recency of those occurrences (for greater detail on IBL model equations see Gonzalez & Dutt, in press; Lejarraga et al., 2011). Our implementation of the IBL model contains two free parameters: the decay of instances over time (i.e., the instances of recency in sampling, *d*), and a noise parameter (σ). Removing the swapping and stopping rules from the IBL leaves all three models to be evaluated solely on their decision mechanisms: highest blended value in the IBL; highest average utility in the Ex-CON; and the highest sample mean in the *k*-sampler. To confirm that we implemented the IBL model exactly as described by its authors, we used our implementation of the model to re-create the model predictions graphed in Lejarraga et al.'s Figure 2. We confirmed that the predictions of our implementation agreed with theirs to within the limits of accuracy aorded by the graph. Note that even though we implemented the IBL model just as its authors intended, the data we report below regarding agreement between the model and data do not always agree with apparently corresponding figures reported by Gonzalez and Dutt and Lejarraga et al.. The differences are due to slight differences in the way the models were evaluated by Gonzalez and Dutt and Lejarraga et al. compared with how we evaluated the models – particularly choices about what should be the basic elements for correlation calculations, and whether the model inputs should be simulated outcomes from gambles or the actual outcomes experienced by participants. Readers

who wish to examine our implementation of the IBL model in more detail can find code for the model in the freely available R language (R Development Core Team, 2011) in the "publications" section of the website http://www.newcl.org/.

Model Implementation and Evaluation

Each model was provided with the same sequences of outcomes experienced by the participants, on a problem-by-problem basis. The sequence of outcomes for each problem and participant was shown to each model 100 times (i.e., there were 100 synthetic model participants for every real participant). After the sampling process had finished for each synthetic participant, the model preference was inferred differently for the sampling and feedback paradigms. In the sampling paradigm, the model choice was determined by the model's decision rule: highest sample mean (*k*-sampler); average utility (Ex-CON); or blended value (IBL model). In the feedback paradigm, after each sample outcome we used each model's decision rule to determine which lottery it would choose on the following sample, for all 100 samples in each problem, and then inferred the preference of the Monte-Carlo replicate as its modal choice over the last 50 trials, as with the human data[6]. We then calculated the modal preference of each model by averaging across the synthetic participants.

When evaluating the model's choice following each sample in the feedback paradigm, we incorporated the probability of inertia (pInertia) parameter of the IBL model in all three models. pInertia refers to the probability of simply repeating the previous choice regardless of the obtained outcome. This is not the same as including the swapping rule, because we still provided the models with the same sequence of outcomes as experienced by the participants. Including the inertia process influences

[6] As with the data, performance of the three models did not markedly differ when using the modal preference across all 100 samples, the last 50 samples, or the final sample in each problem.

the preference about the next lottery to sample following each outcome, and hence influences the decision rules that we aim to study. To ensure consistency in the evaluation of choice rules across models, we included the same inertia process in the *k*-sampler and Ex-CON models, with exactly the same fixed parameter value for all models. This parameter was fixed at pInertia = .13, which was just the median proportion of switches between lotteries while sampling, across problems and participants, in the feedback paradigm in data[7].

For each free parameter in each model, we calculated model predictions for 20 different values that spanned a feasible range for that parameter, determined from previous implementations of the models and also from conceptual constraints. For instance, in the *k*-sampler we calculated model predictions separately for $k = 1, 2, \ldots, 10$, and then log-spaced an additional 10 positive integers from $k = 11, \ldots, 50$. We chose the fine-grained grid for lower values of k since most people in experience-based choice generally take few samples from each deck. In the Ex-CON, we examined 20 log-spaced values for α in the interval [0, 0.1]. The upper end of this interval might appear small at first glance, but it implies a very high degree of confusion: if each sample leads to a 0.1 chance of confusion, the chance of accurately maintaining a memory trace for 10 samples is only $(1- .1)^{10} = .349$, and following 100 samples is $(1 - .1)^{100} < .001$. For the IBL model we simulated 20 log-spaced values for both the decay ($d = .1 - 10$) and noise ($\sigma = .1 – 1.5$) parameters, resulting in a total of 400 parameter combinations. These parameter ranges covered the best fitting values of previous implementations of the IBL model to similar

[7] In the Technion Prediction Tournament section below we show that fixing the value of pInertia as the median number of deck switches that occurred in data likely provides convergent estimates with optimizing the parameter value for goodness of fit.

experience-based choice data (e.g., Gonzalez & Dutt, in press; Lejarraga et al., 2011).

The goodness-of-fit of each model was assessed on the ability to predict two outcome measures in data: the proportion of times the modal preference of the model agreed with the participants' choices, on a trial-by-trial basis; and how well the model predicted participants' estimates of the outcome probabilities. To calculate the accuracy of choice predictions, for each game and each participant, we calculated the proportion of times the model successfully predicted the choice made by the participant when exposed to the same sequence of samples. We refer to this method as "trial-by-trial agreement". In contrast to trial-by-trial agreement, most studies generally report the average proportion of agreement (across problems) between the modal preference of the model and of the people (e.g., Erev et al., 2010),, which we refer to as "PAgree". Trial-by-trial agreement maintains more information than PAgree. For instance, suppose there was a choice problem between lotteries A and B for which 51% of participants preferred lottery A. The PAgree method would treat a model that perfectly agreed with the data (51% preference) as equal to a model that always preferred lottery A (100% preference), but the trial-by-trial measure captures the large difference between these models. Such considerations mean that trial-by-trial agreement is generally lower than observed when evaluating models using PAgree (compare our Table 2 with Erev et al.). Nevertheless, we confirmed that if we re-calculate our analyses using PAgree rather than the new trial-by-trial measure, the general results are unchanged – the detailed results are simply noisier.

The second outcome measure used to evaluate the models was how well they predicted the probability distribution estimates of the people. In the k-sampler and Ex-CON models, probability estimates were derived in the obvious manner, from the

frequency of each outcome in the stores[8]. The IBL model's probability estimates arose naturally from the model's architecture, by using the probability of retrieval for each outcome (for IBL model equations see Gonzalez & Dutt, in press; Lejarraga et al., 2011). We summarized the models' errors in predicting participants' probability estimates by calculating the sum of the squared deviations between participant and model probability estimates, across the sampled probability bins. This measures an average distance between the data and the model when represented in a plot such as in Figure 4.

When considered together, the two outcome measures provide constraints on the models that would not be provided by consideration of either outcome measure in isolation. As we will show, both the Ex-CON and IBL models can predict both outcome measures in isolation quite well, but predicting both outcomes simultaneously is more difficult. We estimated the best fitting parameters separately for the sampling and feedback paradigms. To re-assure the reader that our results are not due to poor choice of the best-fitting parameters, or unwise tradeoffs between mis-fit on the two outcome measures, we also provide profile plots of each model's predictions across a wide parameter range.

Agreement Between Data and Models

Before describing the best fitting parameter estimates, we briefly demonstrate the behavior of each model across its parameter range shown separately for both outcome measures. Figure 5 shows the trial-by-trial agreement for the *k*-sampler, Ex-CON and IBL models (left, middle and right columns, respectively), separately for the sampling (top row) and feedback (bottom row) conditions. Figure 6 uses the

[8] Note that the *k*-sampler and Ex-CON models as described have no way of giving a non-zero rating to a foil card.

same layout to show the sum-of-squares prediction error for the probability distribution estimates of the three models. For trial-by-trial agreement, larger values indicate better performance, but for probability estimates, smaller sum-squared error indicates better performance.

For each model there were parameter settings that predicted quite good trial-by-trial agreement with participants' choices (Figure 5). In the *k*-sampler this region was widespread, with the model predicting choice preference about as well no matter what parameter value is used, except for $k = 1$. In contrast, the middle column of Figure 5 shows that in both the sampling and feedback tasks, as α increased in the Ex-CON the trial-by-trial agreement with choice preference decreased, but this decrease was more marked in the feedback than the sampling condition. In the IBL model, trial-by-trial agreement was generally better with moderate levels of noise ($\sigma < .6$) and low decay in the sampling condition ($d \approx 1$) but high decay in the feedback paradigm ($d > 5$).

The probability estimates of the models were more sensitive to parameter settings than the predictions for choice preferences, except for the *k*-sampler which made very poor predictions across its entire parameter range (Figure 6). The Ex-CON's predictions for probability estimates agreed quite well with the data for any value of $\alpha > .02$, for both conditions. Qualitatively, when $\alpha = 0$ the Ex-CON model behaves like the *k*-sampler and predicts a simple $y = x$ line for probability estimates, but as α increases, the predicted probability estimates become inverse-S shaped (like the data) and then progressively flatter, as the exemplars become dominated by random noise. Similar to the Ex-CON, the IBL only predicted probability estimates that were close to data with high levels of noise ($\sigma > .8$) and less decay.

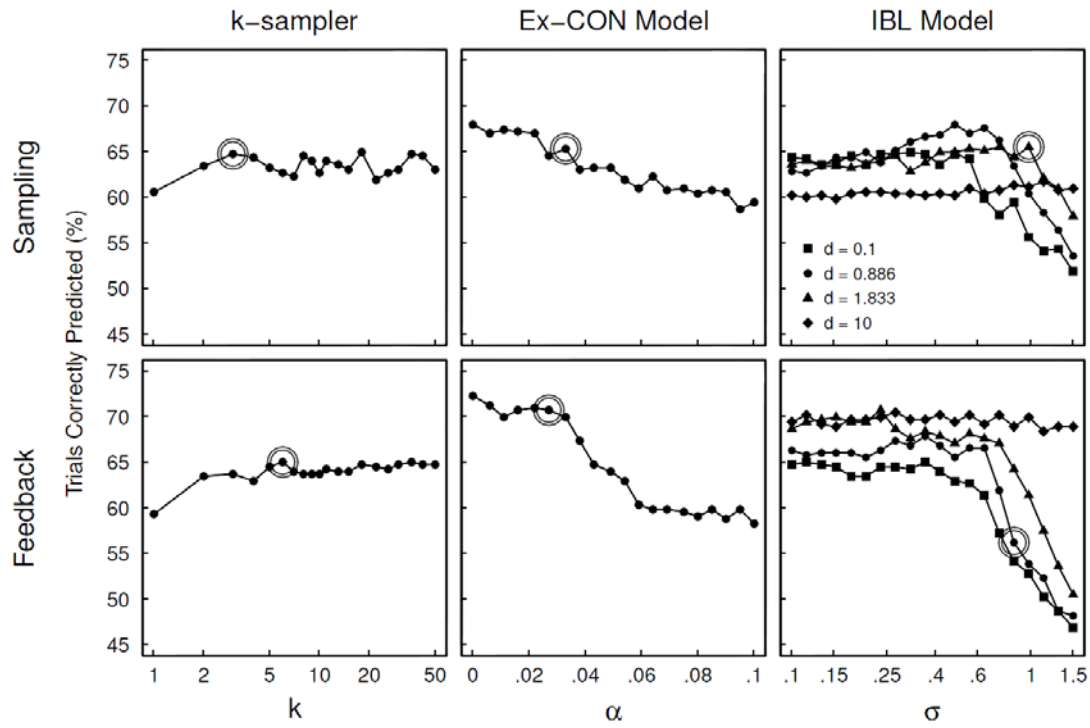


Figure 5: Percentage of choices correctly predicted on a trial-by-trial basis by the k-sampler, Ex-CON and IBL models (left, middle and right columns, respectively) across their respective parameter ranges in the sampling (top row) and feedback (bottom row) paradigms. For the IBL model (right column), we plot four representative values of decay from the 20 simulated values. A larger value indicates better performance. The double circle in each panel represents the parameter value we chose to maximize trial-by-trial agreement and minimize prediction error in estimating probability distributions (see Figure 6).

The profile plots in Figures 5 and 6 demonstrate the challenge in simultaneously predicting choice preferences and probability estimates. Trial-by-trial agreement with choice preferences is generally better when there is less noise in the system (i.e., low α and σ in the Ex-CON and IBL models, respectively). However, these same models best capture the inverted-S shape present in the probability estimate data when the system is subject to substantial levels of noise. We now describe in detail the best fitting parameter estimates for each outcome measure in each model.

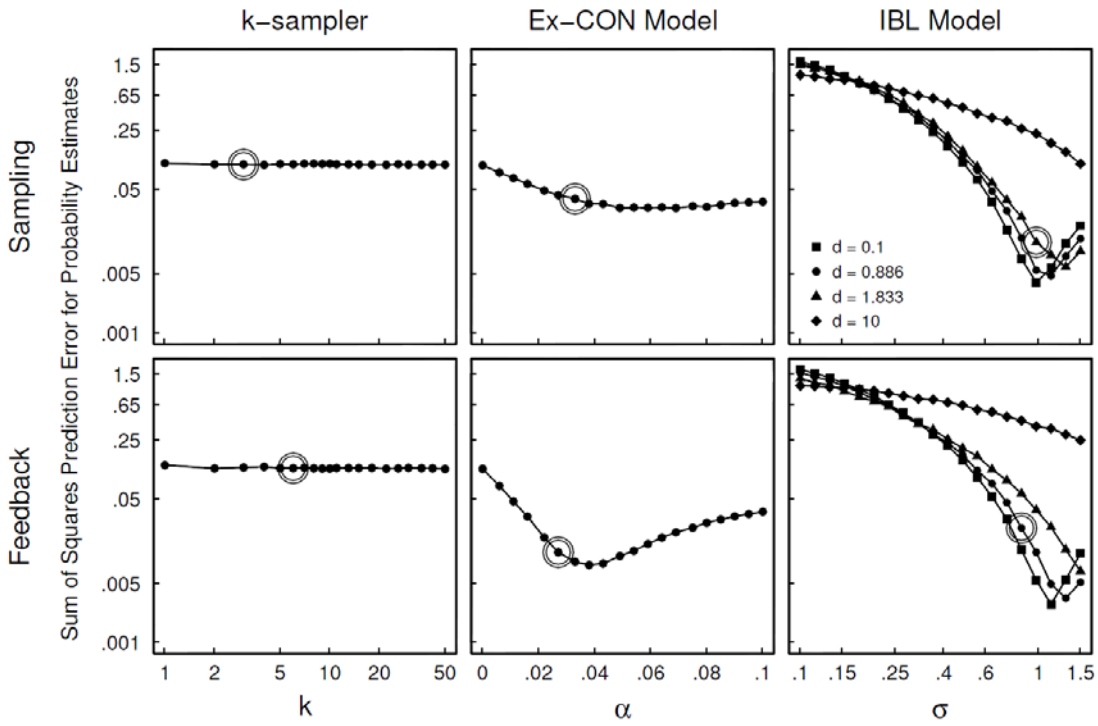


Figure 6: Sum-squared prediction error of the k*-sampler, Ex-CON and IBL models (left, middle and right columns, respectively) versus the probability estimates in data, across their parameter ranges in the sampling (top row) and feedback (bottom row) paradigms. The right column depicts the same four representative values of decay shown in Figure 5. A smaller value indicates better performance. The double circle in each panel represents the parameter value we chose to minimize prediction error in estimating probability distributions and maximize trial-by-trial agreement (see Figure 5).*

Choice Behavior

Table 2 shows the best fitting parameter estimates and trial-by-trial agreement for the models. Each model predicted approximately 65% trial-by-trial agreement in the sampling paradigm. In contrast, in the feedback paradigm the *k*-sampler again correctly predicted 65% of choices, with the Ex-CON model doing a little better, and the IBL doing a little worse.

Despite the small differences in trial-by-trial agreement across models, the parameter estimates from each model provide convergent evidence for the psychological constructs underlying the decision process. For instance, as more samples are observed, as in the feedback paradigm, the psychological representation of the outcome distributions becomes less noisy; that is, based on more stored

samples (larger *k*) or reduced noise (lower α and σ). For example, in the Ex-CON, after five samples an instance has an 84% probability of retaining its original label in the sampling paradigm (α = .033 and an 87% probability in the feedback paradigm (α = .027). After 10 samples, these probabilities drop to 71% and 76%, respectively, and continue to drop with each new sample.

Table 2

Parameter estimates for trial-by-trial agreement and probability estimates, and trial-by-trial agreement (as a percentage) for these parameter settings in the k*-sampler, Ex-CON and IBL models, as well as participants' probability estimates, for the sampling and feedback conditions*

		k -sampler	Ex-CON	IBL	Probability Estimates
Parameters	Sampling	*k* = 3	α = .033	*d* = 1.833, σ = .978	
	Feedback	*k* = 6	α = .027	*d* = .886, σ = .848	
Prediction Accuracy	Sampling	64.8	65.3	65.5	62.9
	Feedback	65.0	70.7	56.2	56.7

The decay parameter in the IBL model similarly suggests that instances are subject to more rapid decay in the sampling compared to feedback paradigm. The decay parameter indicates how long instances are likely to bear influence on the current choice. Unlike the Ex-CON model, the decay parameter of the IBL does not explicitly set forgetting probabilities but rather sets the influence that each instance has on memory, in the IBL's activation function taken from ACT-R. For the sampling paradigm, we estimated the decay parameter at *d* = 1.833, which indicates that after five samples the activity of an instance drops to just 5.2% of its initial influence, and to only 1.5% after 10 samples. In contrast, the lower decay value

estimated in the feedback paradigm (d = .886) indicates that an instance still has approximately 24% and 13% of its influence following 5 and 10 samples, respectively. While it is tempting to interpret these activation values as indicating more rapid forgetting than the corresponding recall probabilities from the Ex-CON model, it must be remembered that these are instance activations, and that the corresponding recall probabilities are calculated by further computation including a Luce choice rule.

The *k*-sampler, as *k* grows large, reduces to a simple heuristic model sometimes known as the "natural mean" heuristic, which is just the idea that the observer prefers the lottery whose sample outcomes were highest on average (Hertwig & Pleskac, 2008)[9]. We compared the three instance-based models against the natural mean heuristic, and found that it predicted choices quite well, but not as well as the models were able to. The natural mean heuristic gave trial-by-trial agreement of 62.9% and 64.5% in the sampling and feedback paradigms, respectively. As a further test, we used only the probability distribution estimates provided by participants, and inferred the expected value of the two decks in each problem from these estimates and their corresponding outcome values. Trial-by-trial agreement bewteen these preferences and the data are shown in the final column of Table 2. As with the natural mean heuristic, the choices predicted by simple weighting of outcomes using the probability estimates agreed with the data better than chance level, but not as well as the three instance-based models.

[9] The Ex-CON model would also be equivalent to the natural mean heuristic, when $\alpha = 0$, but only if we had assumed a linear utility function.

Probability Distribution Estimates

Although the *k*-sampler – as well as the natural mean heuristic and participant probability estimates – can predict participants' choice preferences at better than chance levels, they cannot predict peoples' probability estimates. The *k*-sampler predicts, on average, perfect probability estimates because the proportion of each outcome stored in the *k* samples will, on average, be the same as the sampled proportion of that outcome; this is just the law of large numbers. This prediction is shown by the diagonal light gray lines in Figure 7, which runs through the $y = x$ axis. These lines do not match the data, in which small probabilities were over-estimated and large probabilities were under-estimated.

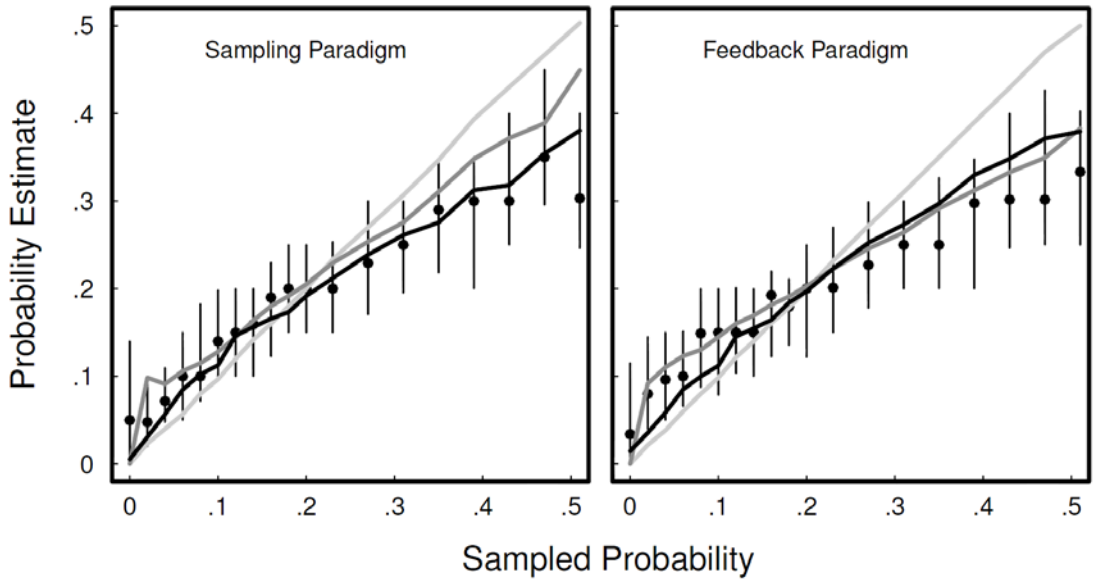


Figure 7: Probability estimates from participants (y-axes) against the sampled probability for the corresponding outcome (x-axis) in the sampling (left panel) and feedback (right panel) conditions. The circles are medians, the whiskers show 5th and 95th percentiles across participants. The light gray, dark gray, and black lines show median probability estimates of the k-sampler, Ex-CON and IBL models, respectively.

The dark gray and black lines in Figure 7 illustrate the Ex-CON and IBL models' predictions, respectively. In contrast to the *k*-sampler, both models capture the qualitative patterns in the probability estimates. Both models predict these data due to their noise processes -- the confusion process in the Ex-CON (described by

parameter α) and the additive noise process (parameter σ) in the IBL. The IBL model more closely captures probability estimates in the sampling paradigm than the Ex-CON model, but the reverse is true for the feedback paradigm.

Choices and Probability Estimates Together

Since participants simultaneously made choices and estimated probabilities, a successful model should be able to capture both measures under the same parameter settings. There is a sweet spot in the parameter settings for the Ex-CON and IBL models that allows a tradeoff in performance on both outcome measures. These parameter settings are the ones that have been illustrated in the model predictions drawn as lines in Figure 7 and indicated by the double circles in Figures 5 and 6.

The performance of the Ex-CON and IBL models can be substantially improved if the outcome measures are considered in isolation. For instance, both models approximate an exemplar model with perfect memory (like a *k*-sampler with $k = \infty$) under certain parameter settings: no confusion in the Ex-CON model ($\alpha = 0$) and low noise in the IBL model (low σ). Under these parameter settings the trial-by-trial agreement with data is very good: both models predict 68% agreement in the sampling task, and around 72% in the feedback task. However, for these same parameter values the models fail to accommodate the probability estimate data, making qualitatively incorrect predictions. The Ex-CON predicts perfect memory (i.e., like the light gray lines of the *k*-sampler in Figure 7) while the IBL model predicts a strong S-shape (not an inverted-S, opposite to data). Conversely, both models can predict probability estimates very similar to data with higher noise settings (Ex-CON, $\alpha \approx .04$-$.05$; IBL, $d < 1$, $\sigma = 1.0$-1.5), but for these parameter settings, choice prediction accuracy is not much better than chance performance.

The Technion Prediction Tournament

Several features of our experiment were different from typical experience-based choice tasks. To evaluate the performance of the models on standard lotteries (albeit, without constraint from probability estimates), we use the data from the TPT Erev et al., 2010). Erev et al. reported on a model competition that allowed quantitative comparison of different theoretical accounts of choice in three separate conditions: decisions from description, and decisions from experience either in feedback or sampling conditions. In each of the three conditions there were data from 60 choice problems between pairs of lotteries. In each problem, one lottery had a low probability of a high payoff outcome, otherwise a low payoff outcome, and the other had a certain, medium payoff outcome. Across choices, one third of gambles were gain-framed, one third were loss-framed, and the remaining third contained a mix of gain and loss outcomes.

The TPT used one data set for estimation of the models' parameters, and another data set for the competition. The estimation data were made freely available and researchers were to develop a model that could correctly predict as many choices as possible in these data. The competition organizers then used the models with the best fitting parameter settings from the estimation set to predict data from a new set of 60 binary choice problems (from an identical problem distribution as the original problems) in the competition set.

We evaluate the models as if they had been entered into the model competition, by first estimating their parameters using data from the estimation set, and then evaluating their predictive performance against the competition data set. As before, we evaluate the performance of the models in predicting choice behaviour using the fine-grained measure of trial-by-trial agreement between the models' modal choices

and participants' individual responses. This analysis is similar in spirit to that used in the TPT, but differs in detail. For ease of comparison, we also report Erev et al.'s (2010) PAgree, and we note that the results of the analyses are similar using either measure.

Entering the Models in the TPT

We simulated the process of entering the *k*-sampler, Ex-CON, and IBL models into the TPT. Each model was simulated in as similar manner as possible as when fitting our own data, but there were three differences. Firstly, the pInertia parameter was fixed, for each model, at the median proportion of sampling switching between lotteries that occurred in the feedback paradigm in the estimation data set (pInertia = .09). Interestingly, Gonzalez and Dutt (in press) estimated pInertia (via search and simulation) from the estimation data set in the IBL at this same value. The second change was that, because the TPT lotteries also included negative outcomes, we used both the positive and negative branches of the utility function estimated by Lopes and Oden (1999) for the Ex-CON:

$$f(n) = \begin{cases} x^{0.551}, & x \geq 0 \\ -(-x^{0.970}), & x < 0 \end{cases}$$

The third change was that we did not calculate predictions for probability estimates from the models, since there were no data for comparison.

We again explored the goodness-of-fit of the models against the estimation data set, across the feasible parameter ranges, using the profile plots shown in Figure 8. These plots also demonstrate that our method of calculating trial-by-trial agreement (black lines) is qualitatively consistent with Erev et al.'s (2010)'s PAgree (gray lines), although the trial-by-trial measure provides a slightly less optimistic view of the models. The *k*-sampler was slightly more sensitive to parameter settings in the

TPT estimation set than in our data, but still gave relatively constant trial-by-trial agreement when $k > 7$ in both sampling and feedback paradigms. Re-assuringly, when $k = 5$ in the sampling paradigm, the *k*-sampler gave a PAgree value of 90%, which is the same outcome Erev et al observed in their model competition (see their Table 3), suggesting that our implementation of their model and the PAgree method is the same. Such precise agreement is only to be expected in the sampling paradigm, because in the feedback paradigm the model's performance was influenced by the inertia process.

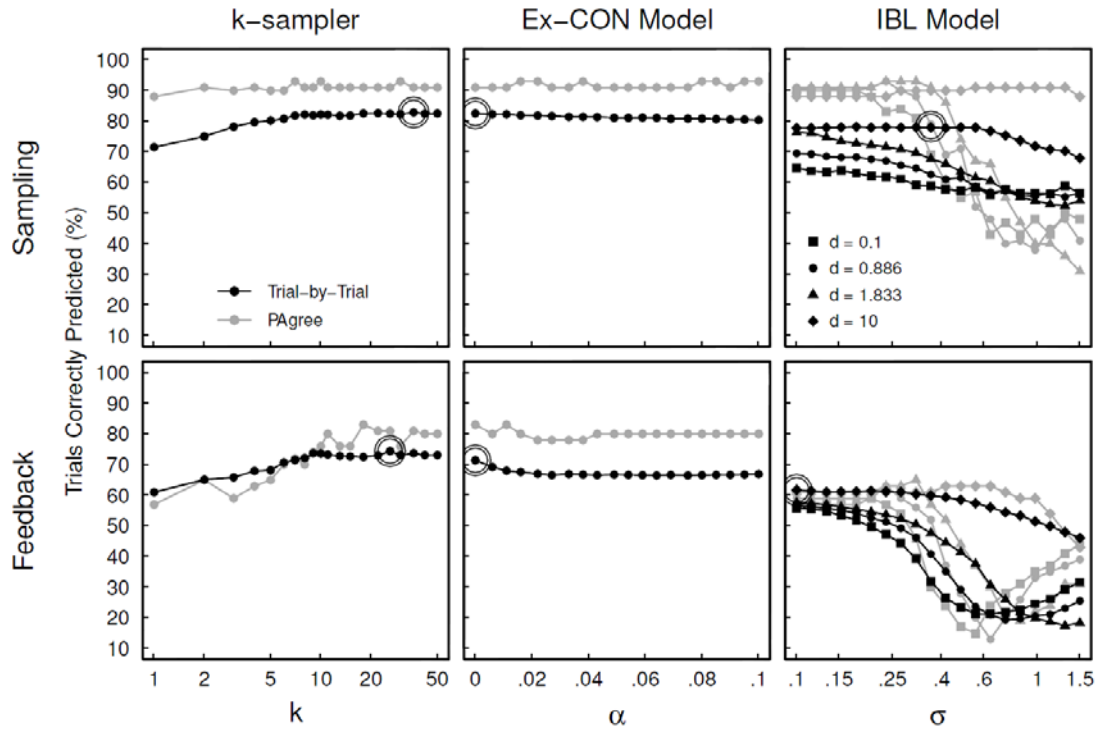


Figure 8: Percentage of choices correctly predicted by the k-sampler, Ex-CON and IBL models (left, middle, and right columns, respectively) across their respective parameter ranges in the sampling (top row) and feedback (bottom row) paradigms for the TPT estimation set. The right column depicts the same four representative values of decay shown in Figure 5 and 6. Trial-by-trial agreement is shown as black lines and PAgree as gray lines. The double circle in each panel represents the parameter estimate that maximized trial-by-trial agreement.

Predictions of the Ex-CON model agreed well with the data for almost all of its parameter range (middle column of Figure 8). The right column of Figure 8 shows that the IBL model was much more sensitive. For example, when there was strong

decay of instances over time, the noise parameter had relatively little influence on trial-by-trial agreement. This occurs because, if there is large decay, then an instance is forgotten very soon after it was observed, and before it can be influenced by recollection noise. In contrast, when decay is low there is a large and detrimental influence of noise on choice prediction accuracy, particularly in the feedback paradigm.

The best fitting parameter estimates, trial-by-trial agreement and PAgree (in parentheses) for each model are shown in Table 3. The performance of each model was impressive across both the estimation and competition data sets, although better in the sampling than feedback paradigm. Overall, the *k*-sampler gave the best trial-by-trial agreement and PAgree values, followed closely by the Ex-CON model and then the IBL. To compare the adequacy of our model fits with those reported for other models in the TPT we use PAgree values, since this was the common metric to both studies. According to PAgree, the models reported here performed about as well as, or better than, those reported in the TPT (Table 3, Erev et al., 2010) – except for the IBL model in the feedback paradigm. The IBL also performed a little more poorly on these data than previously reported by Gonzalez and Dutt (in press), but this difference is most likely due to our slightly different implementation of the model. In our implementation, we have focused tightly on the decision mechanisms of the models, at the expense of the swapping and stopping rules. Recall that we did not allow the models to choose their own balance between exploration and exploitation, but instead yoked them to observe exactly the same outcomes as experienced by the participants.

Table 3.

Percentage of participants' preferences correctly predicted on a game-by-game basis by the k -sampler, Ex-CON and IBL models for the estimation and competition data sets of the Technion Prediction Tournament, using the best fitting parameter values from the estimation data set (described in text). Numbers in parentheses represent PAgree values.

		k -sampler	Ex-CON	IBL
Parameters	Sampling	$k = 36$	$\alpha = 0$	$d = 7.848$, $\sigma = .361$
	Feedback	$k = 26$	$\alpha = 0$	$d = 10.0$, $\sigma = .1$
Estimation Data	Sampling	82.8 (91)	82.5 (91)	78.3 (90)
	Feedback	74.4 (81)	71.4 (83)	61.8 (61)
Competition Data	Sampling	84.0 (93)	84.2 (93)	71.8 (80)
	Feedback	75.0 (92)	73.8 (83)	66.2 (83)

Particularly noteworthy is that the three models reported here performed comparably to the winners and runners-up in both the sampling *and* feedback paradigms, in contrast to the TPT where each model was successful only in the sampling *or* feedback paradigm. The best fitting parameter estimates of the three models tend towards extremes in their parameter ranges, where the models approximate accounts of the choice process, as in the simpler ``natural mean'' heuristic (i.e., large k, $\alpha = 0$, low σ). This is a concern, particularly in the feedback paradigm where each trial involves many samples. However, as suggested in Figure 8, selecting a parameter setting – for any model – with a much higher noise value would only slightly lower the trial-by-trial agreement. For example, if $k = 7$ and $\alpha = .011$ in the sampling paradigm and $k = 10$ and $\alpha = .006$ in the feedback paradigm, trial-by-trial agreement does not substantially decrease, and also yields parameters which are more consistent with those estimated from our own data set. The

conclusion we draw is that without probability estimate data from the TPT the parameters of the models are under-constrained.

Conclusions

We argued that a general model of experience based choice should be able to account not only for peoples' choices in multiple paradigms, but also describe the process by which people construct and represent the probability distributions upon which those choices are based. A model should elucidate how people arrive at knowledge of the outcomes and their probabilities: the two elements typically assumed to be required for determining (rational) choice (e.g., Kahneman & Tversky, 1979; Lopes & Oden, 1999). Indeed, knowledge of the outcomes and their probabilities is the key difference between decisions from description and experience, and deserves investigation in its own right. We therefore conducted an experiment in which we simultaneously asked participants for their estimates of the outcome probabilities, and their lottery choice preferences.

We assumed that memories for the sample outcomes (instances) were used to represent the outcome distribution for each lottery. Whereas some models, such as CPT, take this outcome distribution representation as a model input, we sought to model the process by which this representation was constructed. To achieve this we focused on a general modeling framework that includes a memory for instances (or exemplars) as a basis for both probability estimation and decision-making, and a source of noise or confusion in the memories. Without these two elements, the models were unable to provide a comprehensive account of data from our experiment. For example, the *k*-sampler has a mechanism for storing instances but lacks a specific noise component, and therefore fails to accommodate peoples'

probability estimates, which showed reliable over-estimation of rare events and under-estimation of common events. In contrast, the Ex-CON and IBL models both contain instance level memory for observed outcomes as well as noise in the memory process, and it is the combination of these two elements that allowed both models to simultaneously predict choice and probability estimate data, under the same parameter settings.

To test the models' performance against existing data, we applied the models to data from a recent and wide-ranging examination of choice models, the Technion Prediction Tournament (TPT). These data confirmed that the models performed at least as well as their competitors, and also demonstrated the importance of the data on probability estimates in constraining the models. In the TPT the only data were participants' choices and with just this one outcome measure the Ex-CON and *k*-sampler models were insufficiently constrained – they produced equivalent performance across a large range of different parameter values (Figure 8).

One conclusion to draw from this work is that focusing on broad mechanisms, such as noisy, instance-based memory, may be more useful than adjudicating between specific instantiations of conceptually similar models. Furthermore, developing parsimonious models that generalize across similar paradigms (e.g., feedback and sampling) helps to accelerate theoretical progress (cf. Gonzalez & Dutt, in press). Although the Ex-CON and IBL model share the notion of noisy, instance based memory there remain illustrative differences between them. For example, the Ex-CON model uses a utility function as the basis of its decision-rule, whereas the IBL uses a blended value. The utility function we used was imported from Lopes and Oden (1999) and not estimated from the data that we collected. The fact that the Ex-CON model performed a little better than the IBL model across our data and the TPT

data when solely considered on the final choice rule suggests that considering outcome utility, rather than raw value, might be important. Moreover, demonstrating that the Ex-CON model was able to closely model the TPT data using such a choice rule is surprising in light of the observation made by the TPT competition organizers that the concept of probability did not play an important role in the models submitted to the feedback competition (Erev et al., 2010, p. 35). An alternative view is that the utility function in the Ex-CON model removes the focus of the model from exemplars and replaces it with the utility of the outcome distribution. In reality, both the Ex-CON and IBL models have a similar core structure of exemplars plus "something extra": for the Ex-CON this "something extra" is a utility function; for the IBL it is the additional decision processing. Our thesis is that exemplars and noise in recollection form an important part of successful models of economic choice, not that these are the only components of successful models.

By contrast, the IBL model captures an important aspect of the decision process that the Ex-CON model does not address, the process of exploration with swapping and stopping. A major contribution of the IBL model is its ability to model the swapping and stopping behavior that is observed as participants explore the outcome distributions of different lotteries (e.g., Gonzalez & Dutt, in press; Lejarraga et al., 2011). In our analysis we imbued the models with omniscient swap and stop knowledge in order to focus on representation construction and choice.

Of course, a complete model of experience-based choice will need to address all components of the decision-making process: how search is conducted, how discovered information is used to construct a representation, and how that representation is used to form a preference. Most models are only concerned with explaining the preference component. Here we have explicitly modeled how

discovered information is used to build up a representation and how that representation is used to form a preference. Future research could attempt to combine the swap and stop machinery of the IBL model with the current Ex-CON model to move toward a more complete model. For determining preferences, it would also be possible to introduce a more complicated decision-rule, such as that of CPT, with parameters estimated from the data. We did not adopt that approach here because we found a parsimonious account in the simple utility function already estimated by Lopes and Oden (1999). A future challenge for models such as these is an extension to description-based choices. This might be accomplished in many ways, but the most likely seems to be the assumption of a simulated sampling process, in which participants imagine the process of drawing samples from described problems. The difficulty of this approach lies in the precise details of the sampling, of course.

The notion that different cognitive mechanisms associated with search, representation formation, and preference, can be combined to form a more complete, general model of experience-based choice reinforces the idea of a continuum of uncertainty (Hau et al., 2010; Rakow & Newell, 2010). Under this interpretation, decision processes are common between description and the various experience-based formats of choice (i.e., a Bernoullian-inspired multiplication of outcome and probability), with the only difference arising from how the information is acquired and used to form a representation.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Camilleri, A. R., & Newell, B. R. (2009). The role of representation in experience--based choice. *Judgment and Decision Making, 4*, 518--529.

Camilleri, A. R., & Newell, B. R. (2011a). Description- and experience-based choice: Does equivalent information equal equivalent choice? *Acta Psychologica, 136*, 276-284.

Camilleri, A. R., & Newell, B. R. (2011b). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review, 18*, 377-384.

Cassimatis, N. L., Bello, P., & Langley, P. (2010). Ability, breadth, and parsimony in computational models of higher-order cognition. *Cognitive Science, 32*, 1304-1322.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., Hertwig, R., Stewart, T., West, R., & Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making, 23*, 15-47.

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.

Gonzalez, C., & Dutt, V. (in press). Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review.*

Gonzalez, C., Dutt, V., & Lejarraga, T. (2011). A loser can be a winner: Comparison of two instance-based learning models in a market entry competition. *Games, 2,* 136-162.

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18,* 240-246.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23,* 48-68.

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., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences, 13,* 517-523.

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

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In Chater, N. and Oaksford, M., editors, *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 209-236). Oxford, England: Oxford University Press.

Hintzman, D.L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

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

Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 289-305.

Lejarraga, T., Dutt, V., & Gonzalez, C. (2011). Instance--based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making, 24.*

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

Lopes, L. L., & Oden, G. C. (1999). The role of aspiration in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology, 43,* 286-313.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Nosofsky, R. M., Little, D.R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar--based categorization. *Psychological Review, 118,* 280-315.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded Classification. *Psychological Review,* 104, 266-300.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria. (ISBN 3-900051-00-3).

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*, 1-14.

Rakow, T. R., Demes, K., & 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.

Reiger, T. (2003). Constraining computational models of cognition. In Nadel, L., (Ed.), *Encyclopedia of Cognitive Science* (pp. 611-615). London: Macmillan.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

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.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Chapter 6: Conclusions

Nearly every decision we make occurs in the context of uncertainty. What career to pursue? Where to invest? Pete Sampras or Andre Agassi? Rarely are all the outcomes and their associated probabilities explicitly laid out before us. Instead, we must rely on our own personal experience gathered over time from similar situations. The aim of the work contained in the current thesis was to reveal the psychological mechanisms underlying such experience-based choices.

Reliance on personal experience often causes us to form a preference that is different to the one we would have formed if presented with the true outcome distributions associated with the alternative options. As revealed in Chapter 3, a key reason for this description-experience choice gap can be attributed to a reliance on inaccurate representations of the world. In most cases, our experiences are very limited and so decisions are made based on a relatively small sample of outcomes. A small sample of outcomes frequently misrepresents the true distribution of outcomes in the world, most often under-representing rare events. This external sampling bias is often combined with an internal sampling bias. The internal sampling bias can be most readily attributed to a noisy memory system that relies heavily on more recently sampled outcomes. Such reliance often compounds the under-representation of rare events in the sample relied upon to make a choice and produces preferences that are consistent with underweighting of rare events. The papers in Chapter 3 therefore demonstrate that the difference between description and experience choice formats can be eliminated when a representative sample is used as the basis of choice, at least when a single choice is made subsequent to learning about the options.

Of course, experience-based choices rarely occur in a vacuum after a lengthy period of costless sampling and reflection. Instead, we usually make experience-based choices on the fly and while simultaneously learning more about the outcome distributions associated with the alternative options. As revealed in Chapter 4, even when samples are perfectly representative of the world, preferences are still consistent with underweighting of rare events when each sampled outcome is consequential. The difference between these experiential tasks – that is, costless sampling followed by a choice (the sampling paradigm) and repeated consequential sampling (the feedback paradigm) – does not appear to be attributable to the tension between the goals of exploring and exploiting the options in the latter format because the difference remains even in the context of complete feedback. However, the difference is significantly reduced when repeated consequential choices are made in a single allocation rather than sequentially. The papers in Chapter 4 therefore demonstrate that the difference between description and experience choice formats is also attributable to the sequential nature of the experience-based choices.

Samples of outcomes acquired sequentially must be combined in some way to represent the outcome distribution. As revealed in Chapter 5, we tend to overestimate rare outcomes and underestimate more common outcomes when asked to explicitly report outcome distributions or to nonverbally represent them. Thus, we do not appear to perfectly weigh and combine sequentially observed outcomes. This inability does not appear to be due to a non-probabilistic focus in the experience format because inducing a probabilistic mindset has little effect. Instead, it appears that our judgment inaccuracies reflect the processes of a noisy memory system. This system is embodied in a new instance-based choice model: the exemplar confusion (Ex-CON) model. The Ex-CON is one of the few experience-based choice models

that can also explain how sequentially sampled outcomes are used to build up a representation of the outcome distribution that closely matches the distribution reported via probability estimates. The model also shows that obtained probability estimates are only useful in predicting choices when combined with a utility function implying diminishing marginal utility. The papers in Chapter 5 therefore demonstrate that explicit probability representation is an important feature of experience-based choice and that another key difference between description and experience-based choice is how probabilistic information is represented – in experience-based choice, this representation appears to be based on noisy, instance-based memory.

One of the main contributions of this work is support for the notion that experience-based choices lie along a continuum of uncertainty that is shared with description-based choices. There are two observations that support this continuum of uncertainty argument. First, when the unique features of experience-based choice are eliminated, then preferences become the same as those observed in the description format. The unique features of experience-based choices are the need to search the environment for information and the need to repeatedly integrate this information into a representation. These unique features give rise to the sources of difference between description and experience: sequential sampling of outcomes, acquisition of biased samples of information, and reliance on noisy memory. Crucially, when these differences are accounted for – by eliminating the sequential nature of the choice, by presenting representative samples, and by manipulating the sequence of outcomes to be cyclical – then choice differences disappear.

Second, the models that best account for experience- and description-based choices explicitly represents probability information and share a common choice mechanism. Based on the results from the Technion Prediction Tournament,

description-based choices are best modelled with a stochastic version of cumulative prospect theory (SCPT). The analysis carried out in the current thesis suggests that experience-based choices are best modelled with the Ex-CON model. The SCPT and Ex-CON models both explicitly represent probability information, combine this with outcome information, and then maximise utility as suggested by axioms of rationality dating back to the 17th century.

The notion of a continuum is in contrast to proposals suggesting that description- and experience-based choices are conceptually unique and therefore require fundamentally different theories choice. According to the current thesis then, models of choice that do not explicitly represent probability and combine it with outcome information – including choice heuristics and reinforcement models – fail to completely capture the psychological mechanisms involved in experience-based choice.

If decisions under uncertainty do lie along a common continuum, then the primary goal of future research is to produce a single, complete model of choice under uncertainty. Such a model would simultaneously account for experience- and description-based choices. The Ex-CON model demonstrates that different basic cognitive processes can be bolted together to produce complex processes like those that occur when making decisions under uncertainty. With this analogy as inspiration, a complete model of choice under uncertainty would be constructed from basic components that are combined and activated under different choice conditions. From the perspective of experience-based choice, more work is required to improve understanding of the search component, particularly the machinery behind stop-and-swap behaviour. From the perspective of description-based choice, more work is

required to improve understanding of how descriptions of probability are represented in the mind.

The insights provided in this thesis are not limited to the theoretical. Beyond the walls of the lab individuals, organizations, and governments continually rely on experience to guide decisions under uncertainty. The findings reported in the current thesis may help to explain why rare events such as the 1993 attack on New York's World Trade Center or the 1988 savings and loan crisis often fail to adequately alter behaviour or policy to prevent future tragedy. The findings also help to explain why different people may hold conflicting opinions about important social issues such as nuclear energy use or immunization despite having access to ostensibly equivalent information. Ultimately, the best choices will be made by those of us who recognize the limitations inherent in our information and memory capability, and actively try to mitigate these biasing powers.