Acta Psychologica 136 (2011) 276–284

Contents lists available at ScienceDirect

Acta Psychologica

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

Description- and experience-based choice: Does equivalent information equal equivalent choice?☆

Adrian R. Camilleri *, Ben R. Newell

School of Psychology, University of New South Wales, Sydney 2052, Australia

ARTICLE INFO

Article history:
Received 31 May 2010
Received in revised form 14 October 2010
Accepted 29 November 2010
Available online 24 December 2010

PsycINFO classification:
2340

Keywords:
Risky choice
Description–experience gap
Decisions from experience
Decisions from description
Sampling

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.

1. Introduction

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.

1.1. 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.

☆ This research was supported by an Australian Postgraduate Scholarship and a UNSW Research Excellence Award to the first author, and an Australian Research Council Discovery Project Grant (DP 0770292) to the second author. We thank Tim Rakow, Tim Pleskac, Amos Schurr, and Ralph Hertwig for the very helpful comments on earlier drafts of this manuscript.

* Corresponding author. Tel.: +612 9385 8395.
E-mail address: acamilleri@psy.unsw.edu.au (A.R. Camilleri).

0001-6918/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.11.007

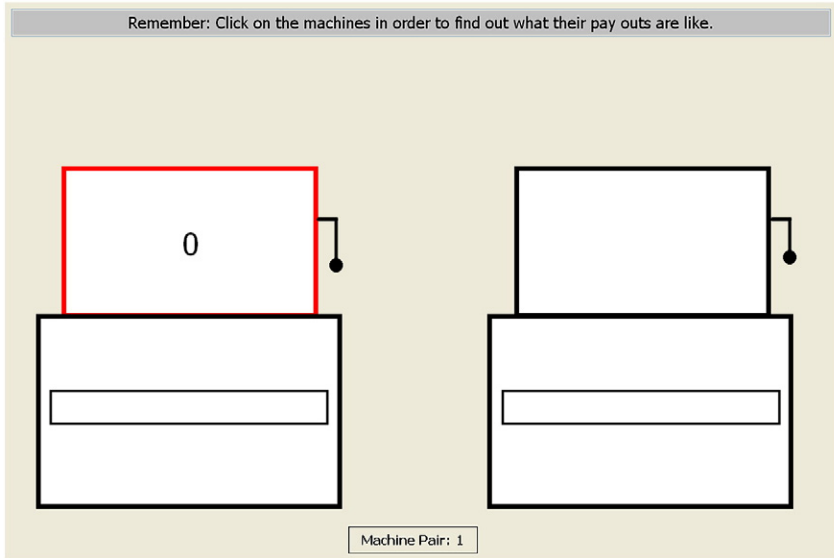

Fig. 1. Screenshot of the experience-based version of the task just as the left option had been selected (revealing a 0).

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., Fig. 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.

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.

1.2. Biased samples

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

1.2.1. 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 a car accident, many participants 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 the small samples, due to the skewed binomial distribution inherent in risky choice problems, resulted in fewer encounters with the rare event than is expected from the payoff distribution (i.e., the number of samples [N] × 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, .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).

1.2.2. Internal biases

Internal sampling biases occur when a mental sub-sample of outcomes does not accurately reflect the observed sample 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), perhaps due to inattention or memory overload (see Kareev, 1995, 2000), or may classify the event as irrelevant to the current decision (Gilboa & Schmeidler, 1995). Indeed, according to Kareev (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 the 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]

1.3. 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 –

[1] Not all studies have found a recency effect (e.g., Hau et al., 2008; Ungemach et al., 2009).

Table 1
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[a]	Internal sampling bias[b]	Results[c]	Notes
Hadar and 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 et al. (2010): Exp. 1. Camilleri and Newell (in press)	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.	–
Ungemach et al. (2009).	Obliged large samples and fixed the outcome pool.	None.	High – sample sequence randomly generated. Fixed large sample length.	Small gap.	–
Rakow et al. (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 reduced to trivial problems (Hau et al., 2010).
Hau et al. (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.

[a] External sampling bias refers to observation of a sample sequence that does not reflect the objective outcome distribution.
[b] Internal sampling bias refers to use of a mental subset of outcomes that does not reflect the observed outcome distribution.
[c] The "gap" refers to different patterns of choice observed as a function of whether options are presented as descriptions or learnt about from experience.

thereby reducing external sampling bias by ensuring that a highly representative sequence is presented (e.g., Hau, Pleskac, Kiefer, & Hertwig, 2008). A consequence of this manipulation, however, is to increase the 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 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.

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.

2. 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

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

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.

3. Method

3.1. 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.

3.2. Materials

3.2.1. 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 Fig. 1). Allocation of safe and risky options to the left and right machines was counterbalanced and the order of the problems was randomised.

3.2.2. 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.

Table 2
Percentage of participants choosing the risky option in Experiments 1 and 2.

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

[a] Indicates the predicted option, that is, the more favourable option if rare events are overweighted.
[b] Due to a programming error, the data for Problems 9 and 10 in Experiment 2 were lost.
[c] Predicted direction is that rare events have more of an impact on decisions when they are described than when they are experienced.
* Denotes significantly different from Description group by $\chi 2$ ($p < .05$).

3.3. 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, and 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, while 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.

3.4. 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$.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.

4. Results

4.1. 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,

[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.

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 ($\chi^2=4.02$, $p=.045$).

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 Fig. 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 ($\chi^2=37.1$, $p<.001$) and 29.9% of trials in the Pseudo-random Experience group ($\chi^2=53.7$, $p<.001$). 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 as a whole, our data replicate previous studies and demonstrate a description–experience "gap".

4.2. 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 $\pm10\%$ of the objective distribution. In order to maintain standardization across problems with rare events of differing rarity, data were categorized 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 $10\%=10\pm1\%$), 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 $15\%=15\pm1.5\%$), 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 $20\%=20\pm2\%$). 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.

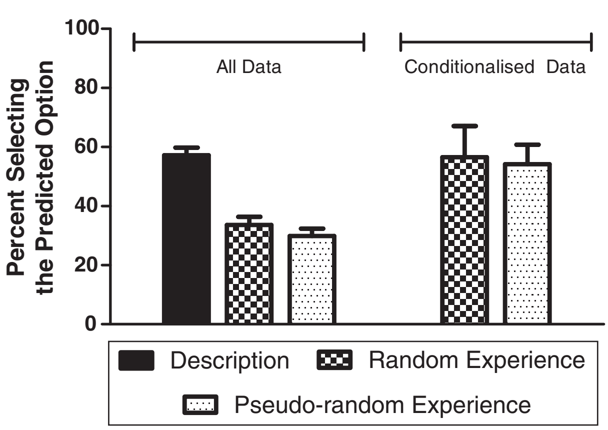

Fig. 2. The percentage of participants selecting the predicted option, assuming rare events are overweighted, in the Description and two Experience groups of Experiment 1. The conditionalised data were those trials where the participants' experienced distribution was within 10% of the rare event objective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

As shown in the rightmost of Fig. 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]

4.3. 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). 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%, $\chi2=3.17$, $p=.077$) but no evidence in the Pseudo-random group (47% vs. 54% $\chi2=.543$, $p=.461$). Similar results were found when comparing just the first vs. last ten samples. Thus, our manipulation also successfully reduced the impact of internal sampling bias.

5. 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

[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, α = .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).

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[a]	
	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

[a] 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.

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.

6. 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, Chater, & Stewart, 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.

7. Method

7.1. 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.

7.2. Materials

7.2.1. 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.

7.3. 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.

8. Results

8.1. 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

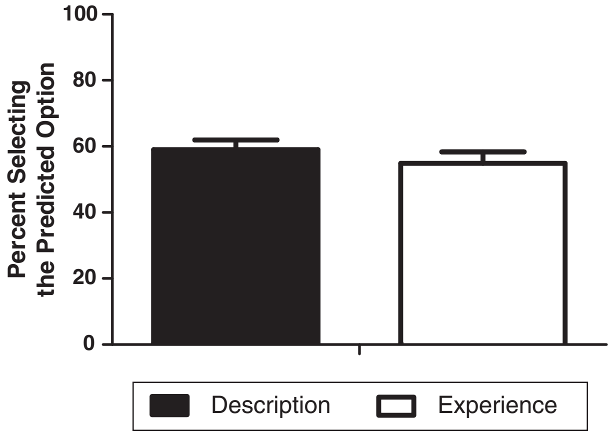

Fig. 3. The percentage of participants selecting the predicted option, assuming rare events are overweighted, in the Description group (Experiment 1) and Experience group (Experiment 2) for Problems 1–8. Error bars indicate the standard error of the mean.

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

8.2. 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 Fig. 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.

8.3. Memory effects

We found no evidence for a recency effect: there was no difference in choice prediction accuracy when based on the first half vs. 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 vs. 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. 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]

9. 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]

10. 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 that the gap is more than just sampling biases (e.g., Hau et al., 2010; see column 5 of Table 1).[9] It is therefore important to highlight that 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.

10.1. 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,

[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, α = .05, sample size group 1 = 288, sample size group 2 = 288.

[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, for 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). In contrast, for both Description and Experience groups in Experiment 1, the majority preference was for the option with the higher expected value in only two of the possible eight cases. This trend suggests that the most likely circumstances under which individuals will choose options with the highest 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 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.

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, in press). 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).

10.2. 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

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 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".

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.

10.3. 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, , 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. (in press). When are why rare events are underweighted: A direct comparison of sampling-, feedback- and description-based choice. Psychonomic Bulletin & Review. doi:10.3758/s13423-010-0040-2.

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.