The Relevance of a Probabilistic Mindset in Risky Choice

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

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

Abstract
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 task.

Keywords: decisions from experience; decisions from description; description-experience; probability; risky choice.

Introduction
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; Rottenstreich & 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>
<thead>
<tr>
<th>Table 1: The experimental groups produced by crossing choice format with induced probabilistic representation.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Format</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

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).

**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 the rightmost column of Table 2. 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 2 for Problems 1 and 2. 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 (\( \chi^2(1) = 2.9, p = .09 \)). Nevertheless, our data clearly replicated a description-experience choice gap.

#### 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%; \( \chi^2(1) = 1.6, p = .2 \)). 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 (labeled red bar; 57% vs. 50%; \( \chi^2(1) = .9, p = .3 \)). 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.

### Table 2: Choice option, expected choice pattern under certain strategies, and percentage selecting the risky option.

<table>
<thead>
<tr>
<th>Problem Number</th>
<th>Choice Options</th>
<th>Expected Choice Pattern Under Strategy</th>
<th>% selecting the risky option*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Safe Risky</td>
<td>Risk aversion Risk seeking Adhere to EV Underweight rare events Overweight rare events</td>
<td>P F A S</td>
</tr>
<tr>
<td>1</td>
<td>3(1.0) 4(8)</td>
<td>Safe Risky Risky Risky Risky Safe</td>
<td>36 44 60 64</td>
</tr>
<tr>
<td>2</td>
<td>14(1.0) 15(9)</td>
<td>Safe Risky Risky Safe Risky Safe</td>
<td>32 36 56 72</td>
</tr>
<tr>
<td>3</td>
<td>5(1.0) 24(2)</td>
<td>Safe Risky Risky Safe Risky Safe</td>
<td>52 28 20 32</td>
</tr>
<tr>
<td>4</td>
<td>3(1.0) 32(1)</td>
<td>Safe Risky Risky Safe Risky</td>
<td>44 52 48 40</td>
</tr>
</tbody>
</table>

Inconsistent Risk aversion Adhere to expected value Risk seeking Overweight rare events Underweight rare event

<table>
<thead>
<tr>
<th>Bar Color</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inconsistent</td>
</tr>
<tr>
<td></td>
<td>Adhere to expected value</td>
</tr>
<tr>
<td></td>
<td>Risk seeking</td>
</tr>
<tr>
<td></td>
<td>Overweight rare events</td>
</tr>
<tr>
<td></td>
<td>Underweight rare event</td>
</tr>
</tbody>
</table>

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%; \( \chi^2 (1) = .5, p = .5 \)). 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%; \( \chi^2 (1) = 2.9, p = .09 \)).

### 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; \( F_{1,19} = 12.7, p = .054 \)), 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.
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 Gottlieb 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.

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.
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