

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

Adrian R. Camilleri · Ben R. Newell

Published online: 17 November 2010
© Psychonomic Society, Inc. 2010

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 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.

Keywords Sampling bias · Underweighting · Risky choice · Description-experience gap · Decisions from experience · Rare events

Description- versus experience-based choice

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

Electronic supplementary material The online version of this article (doi:10.3758/s13423-010-0040-2) contains supplementary material, which is available to authorized users.

A. R. Camilleri (✉) · B. R. Newell
School of Psychology, University of New South Wales,
Sydney 2052, Australia
e-mail: acamilleri@psy.unsw.edu.au

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 relatively 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., 2008, 2010; 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, & Bussemeyer, 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 experience-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 randomised.

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.

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 Fig. 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

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.

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

in the Full Feedback group also received outcome feedback for the foregone, unselected option.

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

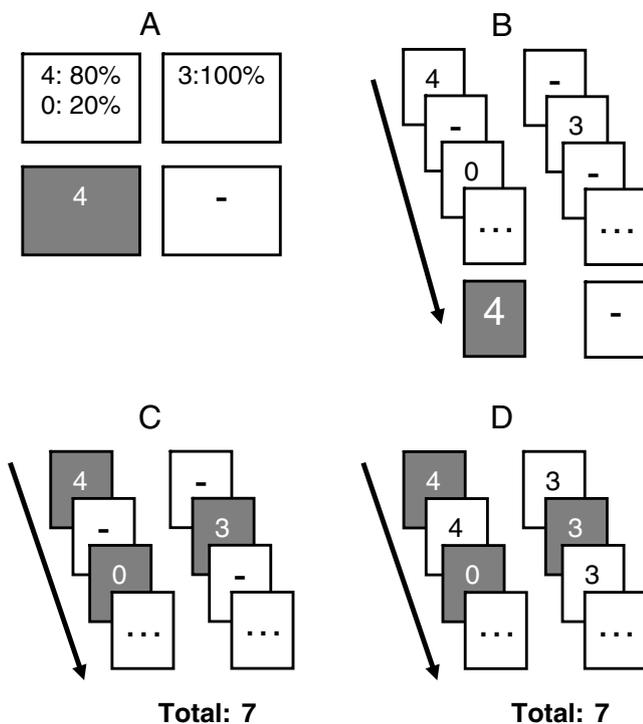


Fig. 1 Depiction of the four different choice paradigms. *Shaded rectangles* represent consequential trials, that is, trials in which the outcome of the choice affected earnings. (A) Description group: explicitly stated outcomes and their probabilities followed by a one-shot choice. (B) Sampling group: 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

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.

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 Fig. 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%, $\chi^2 = 13.2, p < .001$) and the Full Feedback groups (51.9% vs. 75%, $\chi^2 = 11.843, p < .01$). 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 aversion,³ 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%, $\chi^2 = .034, p > .1$). 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.

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 Fig. 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%, $\chi^2 = 6.68, p < .05$), Partial Feedback (37.5% vs. 76.3%, $\chi^2 = 32.04, p < .001$) and Full Feedback groups (37.5% vs. 75%, $\chi^2 = 30.0, p < .001$).

³ Note that risk aversion is unlikely to be attributable to the one-shot nature of the Sampling paradigm because allowing participants to receive the average of 100 plays had no effect (see Footnote 2).

Table 2 Percentage of participants who preferred the risky option

Choice options		Percentage preferring the risky option			
Safe	Risky	Description (<i>n</i> = 40)	Sampling (<i>n</i> = 40)	Partial feedback [^] (<i>n</i> = 20)	Full feedback [^] (<i>n</i> = 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

[^] The DV was the choice made on the final (i.e., 100th) trial

[†] Significantly different from Sampling group ($\chi^2 < .05$)

[†] Significantly different from Partial Feedback group ($\chi^2 < .05$)

[‡] Significantly different from Full Feedback group ($\chi^2 < .05$)

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 Fig. 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.

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.

Fig. 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

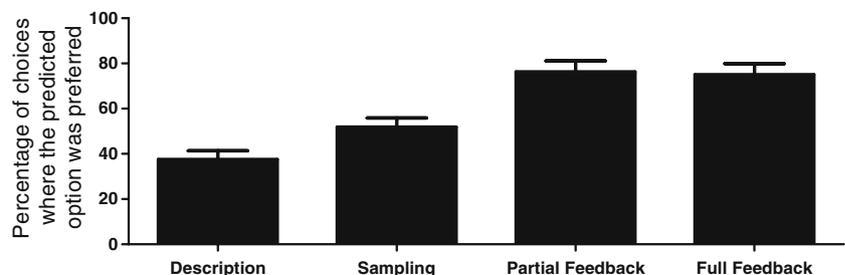
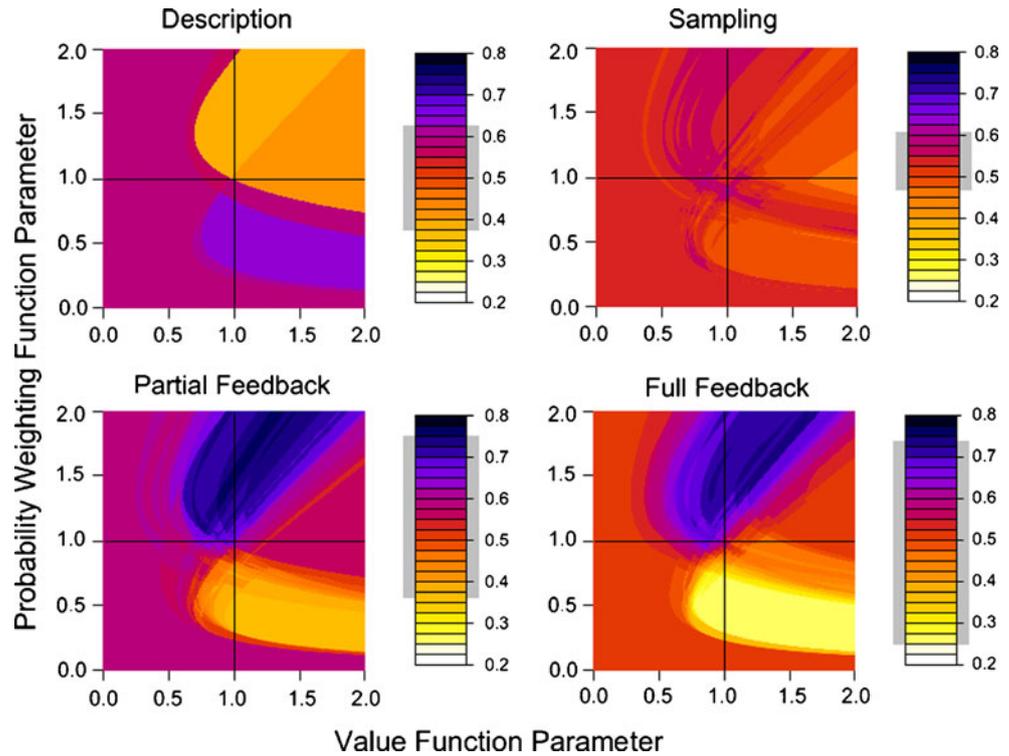


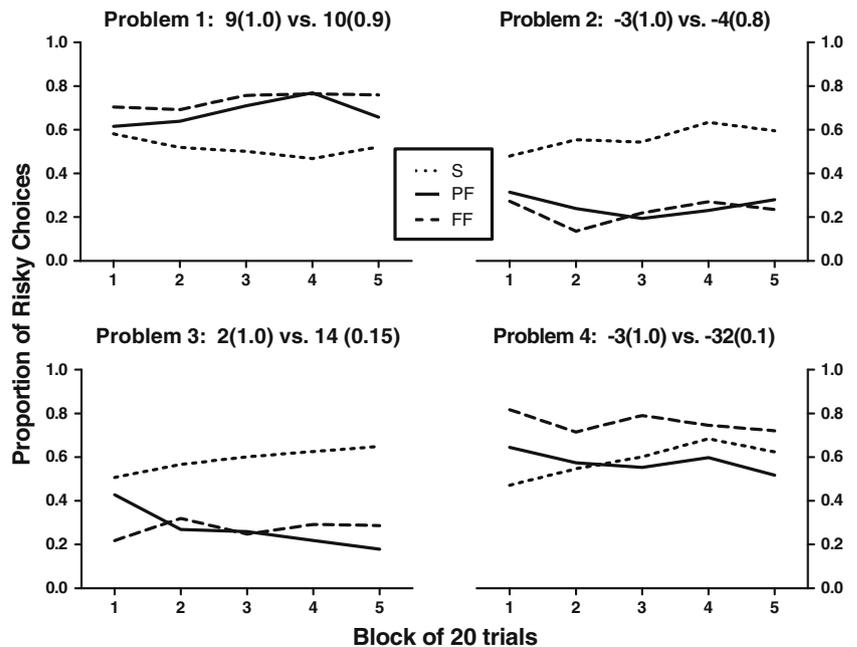
Fig. 3 Contour plots showing the proportion of correct predictions when the data from the Description, Sampling, Partial Feedback and Full Feedback groups were fitted to PT. The proportion of correct predictions was calculated for each combination of value- and weighting-function parameters between 0 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 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).

Fig. 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



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 Fig. 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).

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 (Fig. 2), (2) the best-fitting probability weighting parameter values in PT (Fig. 3), and (3) participants’ sampling experience across trials (Fig. 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).

Acknowledgement 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 are indebted to Michael Hill and Chris Tynan for all their help.

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. (2010). Description- and experience-based choice: Does equivalent information equal equivalent choice? Article submitted.
- Denrell, J., & March, J. G. (2001). Adaptation as information restriction - The hot stove effect. *Organization Science, 12*, 523–538.
- Erev, I., & Haruvy, E. (2005). Generality, repetition, and the role of descriptive learning models. *Journal of Mathematical Psychology, 49*, 357–371.
- 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.
- 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.
- 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.