
ELSEVIER
Volume 126, Issue 1, January 2013
ISSN 0010-0277
pp. 1–134
COGNITION

Cognition 126 (2013) 54–71

Contents lists available at SciVerse ScienceDirect

Cognition

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

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

Adrian R. Camilleri [a,*], Ben R. Newell [b]

[a] Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708-0120, USA
[b] School of Psychology, University of New South Wales, Sydney, Australia

ARTICLE INFO

Article history:
Received 13 October 2011
Revised 30 August 2012
Accepted 3 September 2012
Available online 27 September 2012

Keywords:
Risky choice
Rare events
Description-experience gap
Repeated play
Multi play
Broad bracketing

ABSTRACT

Previous research has shown that many choice biases are attenuated when short-run decisions are reframed to the long run. However, this literature has been limited to description-based choice tasks in which possible outcomes and their probabilities are explicitly specified. A recent literature has emerged showing that many core results found using the description paradigm do not generalize to experience-based choice tasks in which possible outcomes and their probabilities are learned from sequential sampling. In the current study, we investigated whether this description-experience choice gap occurs in the long run. We examined description- and experience-based preferences under two traditional short run framed choice tasks (single-play, repeated-play) and also a long-run frame (multi-play). We found a reduction in the size of the description-experience gap in the long-run frame, which was attributable to greater choice maximizing in the description format and reduced overweighting of rare events in the experience format. We interpret these results as a "broad bracketing" effect: the long-run mindset attenuates short-run biases such as loss aversion and reliance on small samples.

1. Introduction

Choices are ubiquitous: They are ever-present at work, at home, at the mall, and at the track. Unsurprisingly, psychologists have devoted considerable time and effort attempting to understand and predict how people make decisions (e.g., Gigerenzer & Gaissmaier, 2011; Kahneman, 2003; Mellers, Schwartz, & Cooke, 1998; Shafir & LeBoeuf, 2002; Weber & Johnson, 2009). Because potential choices vary across so many dimensions, those who study decision-making often come to rely on a small set of choice paradigms designed to isolate and standardize variables across experiments. For example, the "heuristics and biases" program beginning in the late 1960s and early 1970s was extremely influential and it relied heavily upon "description-based" choice paradigms incorporating simple monetary gambles where outcomes and probabilities were explicitly stated (see Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982).

As a result of such standardization, many of the subsequently formed research questions have only been explored through the prevailing lens of the established description-based choice paradigm. For example, the question of whether decision-makers form similar preferences when a choice is framed in the short run (one outcome) or in the long run (many aggregated outcomes) has only been examined in the context of description-based paradigms (Wedell, 2011). We find this limitation worrying in light of a recent movement away from the description paradigm to an "experience" paradigm within which many of the conclusions made using the description paradigm fail to generalize: an apparent "description-experience gap" (Hertwig & Erev, 2009). The goal of this paper then is to examine an old research question with a new lens: is there a description-experience choice gap in the long run?

* Corresponding author.
E-mail address: adrian.camilleri@duke.edu (A.R. Camilleri).

0010-0277/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.09.001

1.1. Different formats of choice

Consider the following gamble: you have a 50% chance of winning $200 and a 50% chance of losing $100. Would you take this single bet? How about if the gamble were played 100 times in a row and you received the sum total of the 100 outcomes? Would you change your mind? If you are like Nobel Prize winning economist Paul Samuelson's colleague then you would refuse the single-play bet but accept the multi-play bet because, according to the colleague, with 100 plays there is a virtual assurance "to come out ahead" (Samuelson, 1963, p. 109). As Samuelson notes, though it is true that if the bet were played 100 times the probability of ending up with less than you started with is remote, one cannot ignore this very real possibility.[1]

This simple anecdote ignited a long-standing debate over whether it is rational to behave differently to a single bet than to multiple plays of the same bet (Lopes, 1981, 1996; Samuelson, 1963; Tversky & Bar-Hillel, 1983). Samuelson showed that it was irrational under expected utility theory to accept the repeated bet when the single bet was rejected under the wealth bands encompassed by the wager (i.e., winning all bets [+$20,000] to losing all bets [−$10,000]). However, others have argued that it is entirely sensible to adhere to such behavior if choice is based on achieving a certain aspiration; in such cases, choice may reasonably be based on the probability of coming out ahead (Lopes, 1981; Lopes & Oden, 1999). The literature contrasting single- vs. multi-play paradigms in the lab has found that reframing a single-play choice as a multi-play choice often causes decision-makers to seemingly shift from overweighting rare events to more appropriately weighting them and thereby more frequently adhering to normative standards such as value and utility maximization (see Wedell, 2011 for a review).

The literature examining single- vs. multi-play preferences in the lab has almost exclusively relied upon on a descriptive choice paradigm. Description-based choices are those in which the outcomes and their probabilities are provided in a summary description form. For example, in making a decision about your retirement savings, you might examine tables of data describing the performance of different investment strategies in terms of their respective returns. In recent years a new literature has emerged contrasting the descriptive choice paradigm with an experiential one. Experience-based choices are those in which the outcomes and their probabilities are initially unknown and must be inferred from repeated samples of outcomes over time. For example, in making a decision about your retirement savings, you might rely on your own remembered experience of previous returns delivered by implementing different strategies. The literature contrasting description- and experience-based choice paradigms in the lab has found that reframing a single-play, description-based choice as repeated-play, experience-based choice causes decision-makers to behave as if they overweight rare events in the descriptive paradigm but underweight them in the experiential one (under the assumption of a nonlinear utility function, see Camilleri & Newell, 2011b).[2] The implications of this tendency are that many of the conclusions made about human choice behavior, particularly how decisions are made in the context of rare events, do not generalize across choice paradigms: the description-experience choice gap (Hertwig & Erev, 2009; Rakow & Newell, 2010).

Table 1
The four choice paradigms produced by crossing the number of decisions and the number of outcomes.

Number of outcomes	Number of decisions	
	One	Many
One	Single-play	Cumulative-play
Many	Multi-play	Repeated-play

In the preceding paragraphs we have implicitly introduced two dimensions upon which description- and experience-based choices differ: the number of consequential decisions required and the number of consequential outcomes produced by those decisions. In Table 1 we explicitly cross these two dimensions and label the four unique types of choice paradigm that are produced as Single-play, Multi-play, Cumulative-play, and Repeated-play (Bristow, 2011).

In a *single*-play situation, the decision-maker has a single decision that produces a single outcome. For example, a gambler bets all of their wealth on black coming up on the very next spin. In a *multi*-play situation, the decision-maker has a single decision that produces many outcomes of a roulette wheel. For example, a gambler programs software to automatically bet some proportion of their wealth on black coming up on each of the next 100 spins. In a *cumulative*-play situation, the decision-maker has many decisions that produce a single outcome. For example, a gambler who needs to make a certain amount of cash to repay a loan shark bets some proportion of their wealth on red or black coming up for the next 100 spins and actively selects a color each time. In a *repeated*-play situation, the decision-maker has many decisions that produce many outcomes. For example, a gambler bets some proportion of their wealth on red or black coming up for 100 spins and actively selects a color each time.

Note that the cumulative- and repeated-play situations are very similar and are distinguishable in this case only by the outcome being sought: fending off loan sharks or making money. Another example may clarify the difference further. Consider two commission-based salesmen working across a period of 1 month: Both salesmen make many decisions about things like where to sell and what selling strategy to adopt, but one receives a commission for every sale made whereas the other receives a commission only if some monthly sales threshold is reached. The salesman receiving the threshold-based commission operates within

[1] In order to lose money across 100 plays of this particular gamble, one would need to be unlucky enough to lose at least 2 bets for every 1 that is won. The chances of this occurring is only approximately 0.0004, which can be calculated with cumulative binomial distribution function.

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

a cumulative-play situation because across the month many sales decisions cumulate to either earn the single commission or not. In contrast, the salesman receiving the per-sale commission is in a repeated-play situation because across the month many sales decisions produce many commissions. Although interesting in its own right, we ignore the cumulative-play paradigm in the current study because it is not essential to the current research question.

It should also be noted that the term "repeated-play" is used differently in the two literatures that we are bringing together. In the short vs. long run choice literature, the term refers to a situation that we define as "multi-play" whereas in the description vs. experience literature the term does indeed refer to the situation that we define as "repeated-play". By explicitly presenting the dimensions on which these choice tasks differ, we hope to clarify the distinction between multi- and repeated play in the literature going forward: specifically, multi-play requires the decision-maker to make a single decision which is then played out multiple times whereas repeated-play requires the decision-maker to make many individual decisions repeatedly across trials.

In appreciating the differences between the paradigms described in Table 1, note that the problem presented to Samuelson's colleague was a bet between a single-play paradigm (one consequential decision and one consequential outcome) and a multi-play paradigm (one consequential decision and many consequential outcomes). Note also that the description-experience choice gap is traditionally set up as a comparison between a single-play paradigm (one consequential decision and one consequential outcome, as in Samuelson's initial bet) and a repeated-play paradigm (many decisions and many outcomes). Presenting the different choice paradigms along the two dimensions as in Table 1 makes it apparent that contrasting the single-play-description paradigm with the repeated-play-experience paradigm varies three dimensions – the number of decisions, the number of outcomes, and the presentation format – and thus makes the source of the description-experience gap impossible to identify.

Recognizing the problems associated with comparing the single-play-description and repeated-play-experience paradigms, Hertwig, Barron, Weber, and Erev (2004) introduced a modified-experience-based choice task in which the decision-maker freely sampled outcomes from the alternative options before moving onto a decision stage where a single decision was made. Note that this "sampling" procedure neatly shifted the experience choice task from a repeated-play to a single-play task and therefore allowed a direct comparison with the traditional single-play-description task. The results from closely controlled lab studies employing the sampling task find that under these conditions the description-experience gap closes significantly (Camilleri & Newell, 2011a, 2011b; Hau, Pleskac, Kiefer, & Hertwig, 2008; Rakow, Demes, & Newell, 2008; Ungemach, Chater, & Stewart, 2009).

With similar intentions, Jessup, Bishara, and Busemeyer (2008) converted the traditional single-play-description paradigm into a repeated-play task by giving participants full descriptions as well as trial-by-trial feedback. A similar study was also conducted by Lejarraga and Gonzalez (2011). The investigators of both studies reported that participants' initial description-based preferences began to resemble experience-based preferences as more trial-by-trial feedback was received. The observations made in these two experiments and those with the sampling task described above suggest that the number of decisions and/or outcomes may indeed contribute to the choice gap; however, there are four issues that limit insight into the underlying mechanics that we describe in the next section and subsequently address with an experiment.

1.2. Limitations of existing literature

One limitation associated with simply observing a reduced choice gap when contrasting single-play or repeated-play versions of the description and experience paradigms is that it fails to reveal which of these two variables – number of decisions or number of outcomes – is responsible for the gap in the first place. For that reason, in the current experiment we collected experience-based choice preferences in the repeated-play, single-play, and multi-play paradigms, thus permitting an evaluation of the causality associated with each factor.

A second limitation of previous literature concerns the elicitation of preferences in the multi-play paradigm. The majority of previous studies have presented participants with a binary choice: just one option is selected and then betting on that option is simulated multiple times. This design leaves decision-makers with no way to indicate indifference between the options, which is difficult to justify in two-alternative problems that possess very similar expected values (Regenwetter, Dana, & Davis-Stober, 2011). Additionally, others have found that decision-makers in some cases prefer to select a combination of risky and safe options when given the opportunity (Zeelenberg, Beattie, van der Pligt, & de Vries, 1996). Moreover, in everyday practice there are many decisions that permit the decision-maker to distribute his or her preference across the available options (e.g., a buffet table or investment strategy). Indeed, the repeated-play paradigm can be conceptualized as an "imagined" multi-play problem made real and in this situation the decision-maker can indeed distribute his or her preferences over time. For these reasons, in the current experiment we collected multi-play preferences in two different ways: First, following the typical procedure (e.g., Wedell & Böckenholt, 1990), we asked participants to indicate a single option that they preferred to allocate all of their 100 plays to. In addition, we asked participants to indicate a distribution of 100 plays across the two possible options in any combination, including indifference (e.g., Bristow, 2011; Thaler, Tversky, Kahneman, & Schwartz, 1997; Zeelenberg et al., 1996).

Third, a serious concern with the traditional all-or-none binary response is that changes in the *strength* of a decision-maker's preference may go undetected if they do not reach some threshold and thereby reverse preference. Given that we conceptualize preference as existing on some kind of graded scale, then it is appropriate to implement such a scale when trying to measure this preference. Thus, after each all-or-none binary choice, we asked participants to report the strength of their preference on a 7-point scale that included an "indifference" option. Note that marked differences between the binary choices and the graded strength of preference responses would suggest that the choice gap is in some sense merely an artifact of the scale being used to collected preferences in typical description and experience tasks.

Fourth, from a libertarian paternalism perspective it is important to learn whether experience-based choices can be "improved" in relation to normative standards such as value or utility maximization (Thaler & Sunstein, 2008). As noted earlier, the evidence from studying preferences in the description paradigm suggests that biases decrease in the multi-play paradigm – what some have called the "long-run rationality" hypothesis (e,g., Keren, 1991; Keren & Wagenaar, 1987; but see Chen & Corter, 2006). For example, one representative gamble from this literature is a choice between a 50% chance of \$250 vs. a 99% chance of \$100. The expected value of the first option is \$125 whereas the expected value of the second option is \$99. In this gamble the first option is the maximizing one since it has the higher expected value. Keren and Wagenaar (1987) found that 33% of participants selected the maximizing option when the gamble was to be played once, but this rate of maximization increased to 65% when the gamble was to be played 10 times.

Interestingly, a close analysis of the literature reveals somewhat conflicting behavior when rare events are present. Note first that almost every problem previously examined that contains a rare event has confounded the maximizing option and the more uncertain option. For example, in the gamble described above, the maximizing option is the 50% chance of \$250, which is also the more uncertain of the alternative options (50% vs. 99%). Curiously, in the few choice problems where this confound was absent, the rate of maximization actually *decreased* under multi-play conditions (Barron & Erev, 2003; Wedell & Böckenholt, 1990). These observations suggest an alternative interpretation of existing data in situations in which rare events are present: the multi-play format may reduce the tendency to overweight rare events. In order to rigorously evaluate whether multi-play framing actually produces more normative behavior we constructed a systematic set of choice problems that avoided the confound.

1.3. The current experiment

In summary, the purpose of the current experiment was twofold: first, to test the extent to which the description-experience gap is a function of the different parameters inherent in the traditional paradigms (i.e., the number of consequential decisions and/or consequential outcomes); second, to clarify whether preferences tend to be more rational in description and experience formats when framed in the long run. In order to answer these questions, we measured preferences in single-, multi-, and repeated-play choice paradigms (Table 1). In order to be able to make all the required comparisons, we collected two versions of multi-play: one in which participants allocated all of their 100 plays to a single option (a binary measure) and one in which participants could allocate their 100 plays across the options (a distributed measure).

Contrasting the single- vs. multi-play *binary* preferences allowed us to assess any influence of the number of consequential outcomes. Contrasting the multi- vs. repeated-play *distributed* preferences allowed us to assess any influence of the number of consequential decisions. To properly judge the direction in which preferences shifted under the different conditions, we measured both choice and preference strength on a set of 32 choice problems that varied systematically on a number of important attributes including whether the maximizing choice was the safer or riskier option.

Our hypotheses were speculative in light of the mixed evidence of multi-play framing on description-based preferences and absence of previous investigation of multi-play formatting on experience-based preferences. Given that previous literature has shown that (1) overweighting of rare events may decrease in description-based choice when single-play is reframed as multi-play (Keren, 1991; Keren & Wagenaar, 1987; Wedell & Böckenholt, 1990) or repeated play (Jessup et al., 2008), and that (2) underweighting of rare events may decrease in experience-based choice when repeated-play is reframed as single-play (Camilleri & Newell, 2011a; Ungemach et al., 2009), our tentative hypothesis was that the description-experience gap would be reduced in the multi-play paradigm. Analysis of whether this hypothesis obtained with distributed and/or binary multi-play measures would reveal whether potential preference shifts are due to the number of consequential decisions, outcomes, or both. We also predicted rates of maximization to be greatest in the multi-play format under the description format, and tentatively extended this prediction to the experience format.

2. Methods

2.1. Participants

The participants were 203 online American workers recruited from the Amazon's Mechanical Turk website. The median age of the participants was 30 years and 56% were female. Eighty-three percent indicated at least some university education. Just over half reported having full-time employment and the median household income was approximately US\$30,000.

2.2. Materials

2.2.1. Decision task

The decision task was a virtual money machine game with two alternative options. In the description-based paradigm, the options were labeled with the potential outcomes and their probabilities (e.g., "*x*% chance of *y*, otherwise 0"). In the experience-based paradigm, the options were presented with only single-letter labels (e.g., "A"); however, each option was associated with a distribution of outcomes in accordance with the outcome distribution shown to those playing the description version of the task. Participants played 40 trials during which time they were presented with a series of outcomes that were randomly arranged but together perfectly reflected the distribution. In order to ensure that participants' sample choices reflected their preference rather than information search (i.e., the "exploration–exploitation tension"), we presented participants with full-feedback: after each sample participants saw the outcome of the option that was selected, which was added to the running total, and also saw the forgone outcome of the option that remained unselected (Camilleri & Newell, 2011b). In this way, participants could learn about the outcome distribution of an option without expending resources taking samples from that option.

2.2.2. Choice problems

Each choice problem 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 (Appendix A). The 32 problems were formed by crossing the following characteristics: rare event rarity (5% vs. 10% vs. 15% vs. 20%); choice domain (gains vs. losses); maximizing option, that is, the option with the higher expected value (safe vs. risky option); rare event desirability (desired vs. not desired). For example, Choice Problem 1 was between a 100% chance of −20.1 or 20% chance −95.5 otherwise 0. This problem incorporates the following features: 20% rare outcome rarity; loss domain; risky option maximizing, and the rare event – in this case −95.5 – is undesirable. The maximizing option always had an expected value (EV) that was 1 unit higher than the EV of the non-maximizing option. The "base" safe option was chosen to be 20.1 in order to ensure that the safe outcome was always smaller than the non-zero risky outcome. The remaining 31 safe options were selected by incrementally adding 1.2 units to the base safe option. Note that given these constraints the only value in Appendix A that was free to vary was the risky outcome. Allocation of safe and risky options to the left and right of screen was counterbalanced and the problem order randomized.

2.3. Design and procedure

The experiment took on average 23 min to complete, which produced an effective average hourly wage rate of $1.25. No other payment was offered. The goal of the task was defined as earning the most number of "points" during the experiment. To minimize the effect of fatigue, each participant faced only sixteen choice problems (either the odd or even questions in Appendix A). Half of the sixteen problems were presented in the description format and the other half were presented in the experience format. The order of the problems and the formats was random.

An overview of the experimental procedure is shown in Fig. 1 separated into four stages. Note that we varied Choice Format (Description vs. Experience), Number of Choices (One vs. Many), and Number of Outcomes (One vs. Many) within-subjects. In the Description condition the alternative options were presented and each was labeled with a summary description of its outcome distribution. In the Experience conditions the same options were presented but were unlabeled. The repeated-play choice data was collected only in the Experience condition by asking participants to "*choose repeatedly between the two options with the goal of earning the most number of points*". The outcome of each sample was displayed and added to a cumulative points total, which was always displayed. The outcome of the foregone option was also displayed but did not add to the cumulative points total. Preference was taken as the distribution of plays during this sampling stage (i.e., average of all 40 trials).

After learning about the alternative outcome distributions – by inspection in the description condition and by sampling in the experience condition – all participants were presented with the three additional questions described below. Note that the presentation order of these questions, which correspond to stages 2, 3, and 4 in Fig. 1, was random.

The single-play data were collected by asking participants to "click on the option that you would choose if you were going to receive a single outcome from 1 single play from one of these options". After making the binary choice for which no feedback was given, participants were asked to report the strength of their preference on a 7-point scale ranging from "Strong Preference" for the left option to "Strong Preference" for the right option with "Indifference" as the middle point.

The binary multi-play data were collected by asking participants to "click on the option that you would choose if you were going to receive the combined outcome of 100 plays from one of these options". After making the binary choice for which no feedback was given, participants were also asked to report the strength of their preference on the same 7-point scale.

The distributed multi-play data were collected by presenting participants with a slider bar and instructions directing them to "*use the slider to distribute 100 plays between the two options*". The slider was anchored with the left option label and right option label (e.g., "A" and "B") and each also displayed a value reflecting the number of plays that were currently allocated to that alternative. The default value for each option was 0. The slider was activated when the participant initially clicked on it and subsequently the values updated as the participant moved the slider handle so that the two values always summed to 100. No feedback was given after the allocation had been finalized.

3. Results

The responses that we collected can be expressed in three different ways: (1) preference for the safe option vs. the risky option, (2) preference for the option if rare events are underweighted vs. overweighted, and (3) preference for the maximizing option vs. non-maximizing option. Each of these expressions can be of value depending on whether one is, respectively, interested in risk attitudes, the weighting of rare events, or rationality. As shown in Appendix A, the set of 32 problems we designed intentionally uncorrelated these qualities so that we could potentially investigate all three simultaneously. Given our research question, however, we were particularly interested in the weighting of rare events and maximizing tendency.

Following precedent (e.g., Rakow et al., 2008), the option consistent with underweighting rare events – that is,

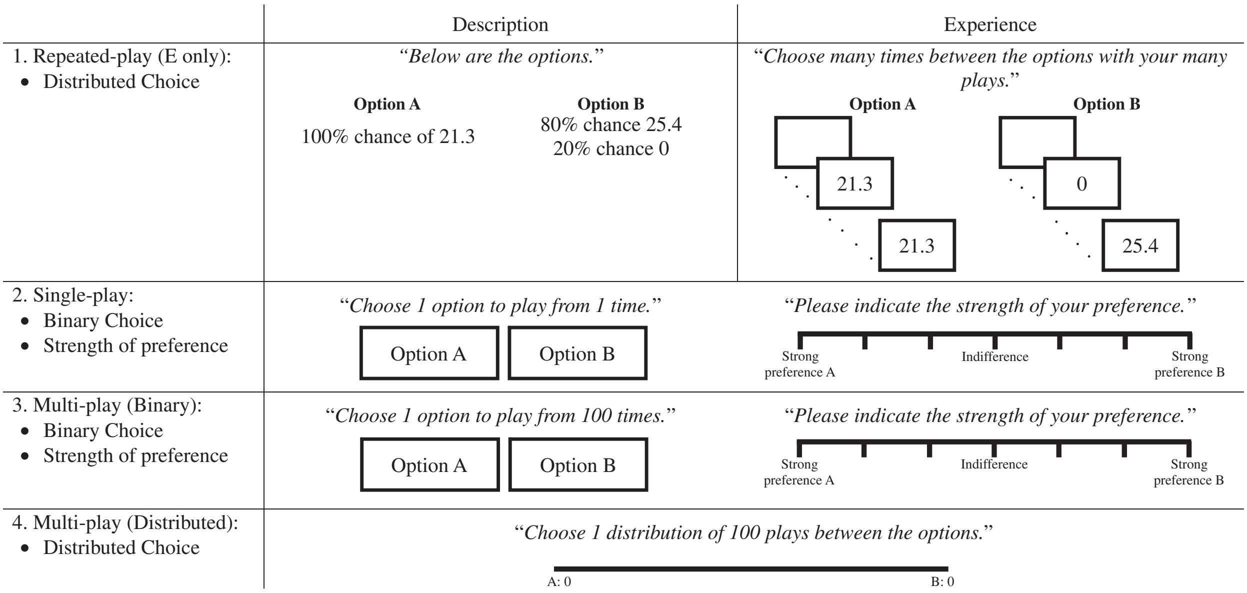

Fig. 1. Summary of the four key stages of the experimental procedure using Problem 3 as an example. Participants completed sixteen choice problems, half in the description format and half in the experience format. Stages 2–4 were randomly ordered. See text for further details of the procedure.

underweighting small probabilities – was the risky option when the rare event was undesirable and the safe option when the rare event was desirable (see Appendix A). Note that "underweighting" rare events and "overweighting" rare events are two anchors along a psychological dimension with the middle point being "appropriate weighting". Thus, "less underweighting" does not immediately imply "overweighting" unless the middle-point of this dimension is crossed. In the following we take as our DV preference for the option if rare events are underweighted. However, given that the current problems were binary choices, we could have just as easily selected the option consistent with *overweighting* rare events, which would simply have inverted the presented values.

The traditional description-experience gap contrasts preferences in the single-play description condition with preferences in the repeated-play experience condition. In our dataset, the proportion of choices consistent with underweighting rare events in the single-play description condition was 0.51 compared with 0.68 in the repeated-play experience condition (see Appendix B). This large difference represents the traditional description-experience gap but, as we have argued, confounds the number of consequential decisions and number of consequential outcomes in each task. In the current experiment comparison of preferences in the single-play task with preferences in the multi-play *binary* task permits insight into the impact of one vs. many outcomes on the choice gap while holding constant the number of decisions. In addition, comparison of preferences in the multi-play *distributed* task with preferences in the repeated-play task permits insight into the impact of one vs. many decisions on the choice gap while holding constant the number of outcomes.

3.1. Single-play vs. multi-play (binary)

In this section we examine the impact of multiple *outcomes* by comparing the two left-most choice paradigms shown in Table 1: single-play vs. multi-play. In the following Section 3.2, we examine the impact of multiple *decisions* by comparing the two bottom-most choice paradigms shown in Table 1: multi-play vs. repeated play. Recall that in order to make these comparisons the multi-play preferences were collected in two ways (Fig. 1): an allocation of all 100 plays to a *single* option (i.e., binary choice), and also by an allocation of 100 plays *across* the options (i.e., distributed choice). In this section we focus on the binary multi-play choice and in the following section we focus on the distributed multi-play choice.

3.1.1. The weighting of rare events

The first question of interest was whether rare events are weighted differentially across the different formats and choice tasks. Fig. 2 shows the average strength of preference for the option consistent with underweighting rare events in the Description and Experience conditions separated by problem Domain and Rare Outcome Desirability[3]. As can be seen, in every case the bar corresponding to the experience condition is taller than the bar corresponding to the description condition, although the extent of this gap appears to vary depending on the choice domain and rare outcome desirability. Moreover, the bars all tend to be taller in the single-play condition than the multi-play (binary) condition. To statistically examine the effect of number of outcomes, we analyzed the data via a mixed-effect model using the Markov Chain Monte Carlo (MCMC) method with Restricted Maximum Likelihood (REML). The mixed-effect model was preferred because it enables the modeling of correlated data – inherent to the within-subject nature of our design – without the violation of important regression

[3] The correlation between choices and strength of preference was strongly positive (Spearman's $\rho = .86$, $p < .001$). As a result, similar inferences follow if the binary choice DV is used, however, preference strength was preferred because it was measured on a graded scale and therefore better captured preferences.

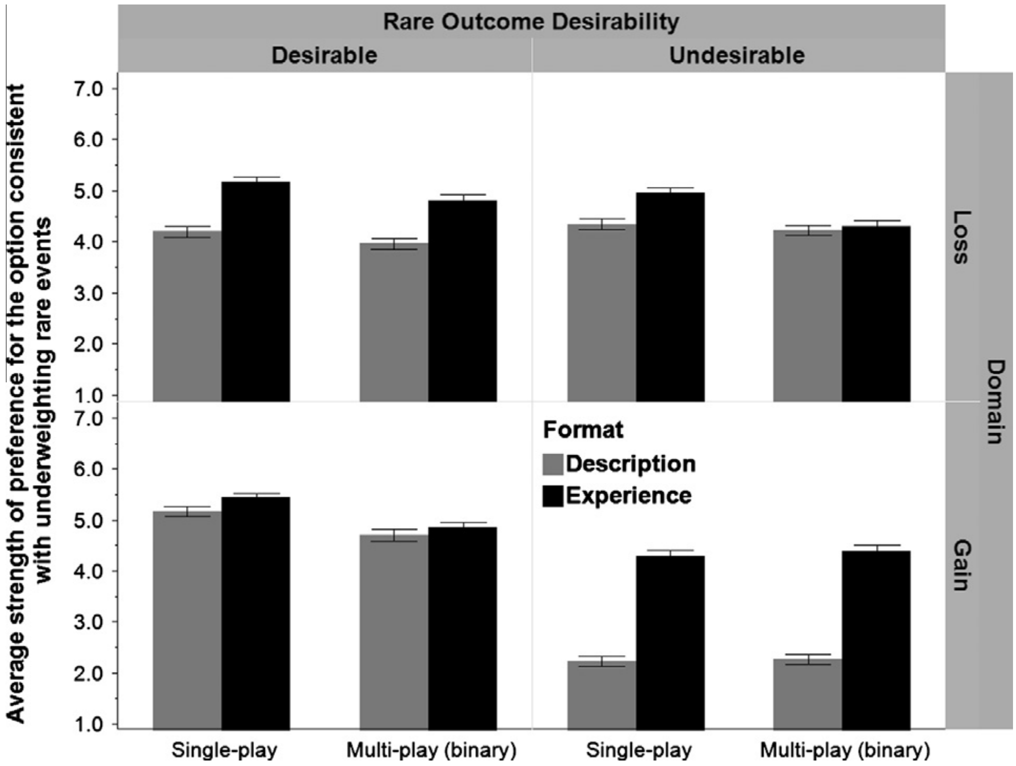

Fig. 2. The average strength of preference (range: 1–7) for the option consistent with underweighting rare events in the Description (gray bars) and Experience (black bars) conditions separated by task, problem Domain (Gain [lower panel] vs. Loss [upper panel]) and Rare Outcome Desirability (Desirable [left panel] vs. Undesirable [right panel]). The error bars indicate the standard error of the mean.

assumptions (Demidenko, 2004). The critical value was set at $\alpha = .05$ and polynomials were centered. The fixed factors were: Number of Outcomes, Format, Domain, Rare Outcome Probability, Rare Outcome Desirability, Maximizing Option, Question Order, and Component Order. The first six of these variables were crossed to the third degree to form interaction terms. Participant ID was entered as a random effect. The dependent variable was the average strength of preference for the option consistent with underweighting rare events.[4]

The specified model revealed a number of significant effects and produced an R^2 of .300 (adjusted $R^2 = .290$). Below we discuss some of the key significant effects; for the full list of effects and associated *p*-values please refer to Appendix C. Consistent with expectations from previous research, the participant's preferences were more consistent with underweighting rare events in the Experience condition than in the Description condition ($p < .0001$). More interestingly, however, the systematic set of problems that we used afforded us the opportunity to discover that this description-experience choice gap was more evident in the gain domain when the rare event was undesirable and also in the loss domain when the rare event was desirable ($p < .0001$; see also Fig. 2).

With respect to our more central research question, the choice gap tended to be smaller in the multi-play frame than in the single-play frame ($p = .06$). Put another way, the choice gap tended to be smaller when the preference was formed in the context of a single choice corresponding to many outcomes rather than just one outcome. Most importantly, a follow-up contrast revealed that this effect was primarily due to a reduced tendency to underweight rare events in the experience condition ($p < .0001$).

In order to obtain an improved understanding of when underweighting occurred, we fitted the data to the two parameters of Prospect Theory (PT; Kahneman & Tversky, 1979). PT is a highly successful model of description-based choice that incorporates nonlinear functions for probability and prospect weighting. 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 Appendix D).

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" between 0.3 and 0.7. The darker shading represents the regions with the better fit. As can be seen by the varied shading, some parameter combinations were more successful than others. The regions with the best fit for the Description condition are for outcome weighting values below 1

[4] Similar results are obtained if the average choice over the final 20 trials or the final trial itself is used. Average choice over the 40 trials was preferred because it encompasses all of the observations.

across the entire range of probability weighting values. These values are consistent with the behavioral data (Fig. 2), which revealed, on average, indifference between the options associated with overweighting and underweighting. In contrast, the regions with best fit for the Experience condition are for outcome weighting values between 0.8 and 1.6, and for probability weighting values above 1 implying underweighting of small probabilities. Again, these values are consistent with the behavioral data (Fig. 2), which revealed, on average, stronger preference for the option associated with underweighting rare events.

There are two particularly noteworthy features of the single-play and multi-play experience figures: First, the shading is darker in the single-play condition, which implies that PT was better able to account for choices made in this frame (this is also true of the description figures). Second, and more importantly, the darkest shaded areas in the two figures are not equal: the darkest band extends much higher in the single-play figure, which corresponds to higher probability weighting parameter values and thus greater underweighting of rare events. This difference therefore provides yet another line of evidence suggesting that underweighting of rare events was less pronounced in the multi-play condition.

3.1.2. Maximizing tendency

A second question of interest was whether choices made under multi-play conditions pushed participants to make more choices in line with maximizing expected value. In order to test this we replicated the mixed-model analysis described in Section 3.1.1 but this time with the dependent variable coded as the average strength of preference for the *maximizing* option (the data are shown in Appendix B).

The specified model revealed a number of significant effects and produced an R^2 of .201 (adjusted R^2 = .190). Below we again discuss some of the key significant effects; for the full list of effects and associated *p*-values please refer to Appendix E. Format (p = .05) and Number of Outcomes (p = .007) were both revealed to be significant predictors but there was no interaction between these variables (p = .48). Follow-up contrasts, however, indicated that the number of maximizing choices increased when the number of outcomes was larger for those in the

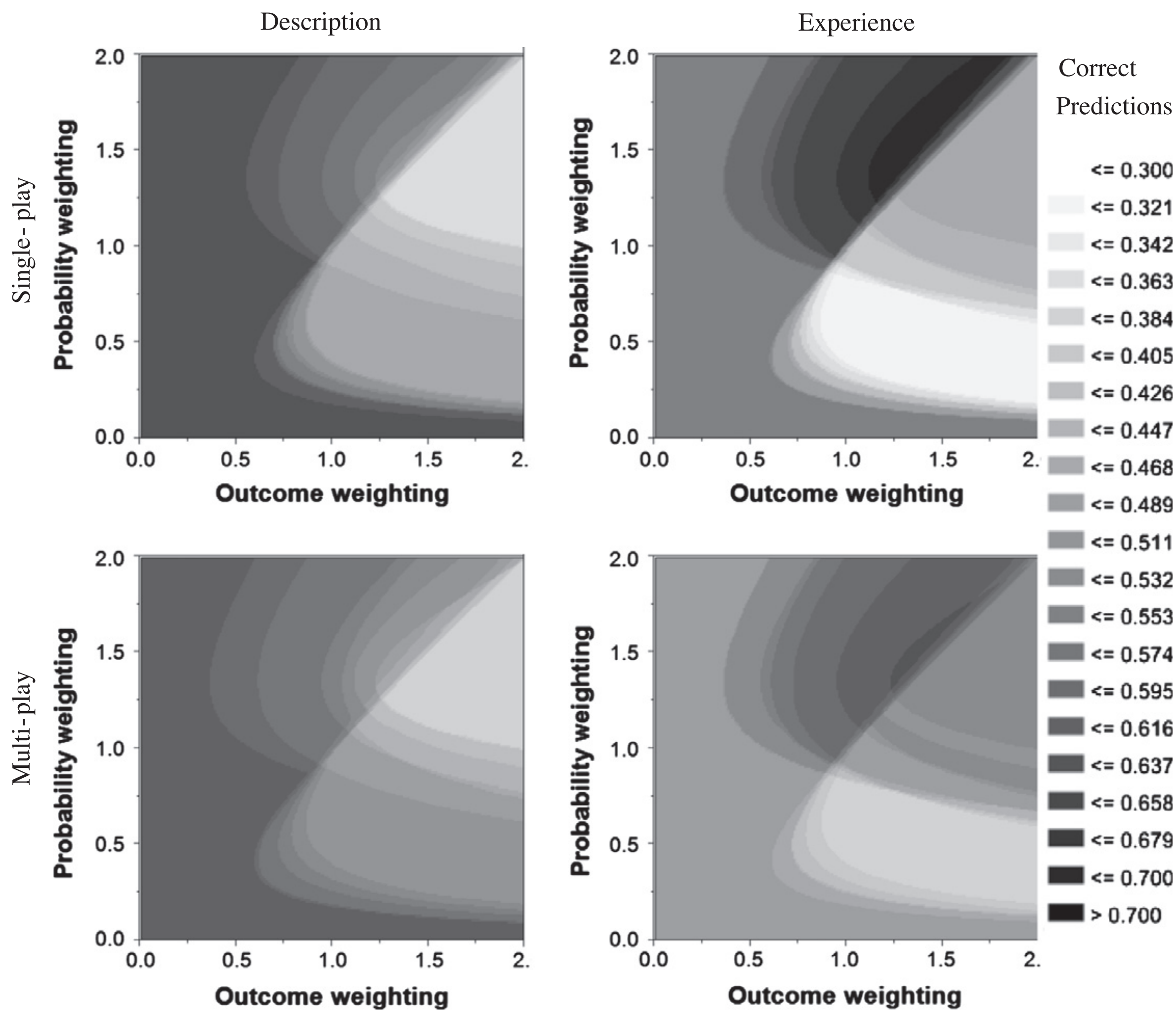

Fig. 3. Contour plots showing the proportion of correct predictions when the data from the Description and Experience condition were fitted to Prospect Theory separately for the single- and multi-play tasks. The proportion of correct predictions was calculated for each combination of value- and weighting-function parameters between 0 and 2, in steps of .01. The regions with the darker shading indicate the combinations providing the higher fit. The legend to the right of the figure indicates how each level of shading corresponds to the average proportion of correct response. The problem domain was ignored by assuming gain-loss symmetry (i.e., $\alpha = \beta$ and $\gamma = \delta$).

Description condition ($F_{(1,6200)} = 6.01$, $p = .01$) but not for those in the Experience condition ($F_{(1,6200)} = 2.14$, $p = .14$). Thus, the tendency to choice maximize increased under the multi-play frame, but only in the Description condition.

The conclusion from the results reported here in Section 3.1 is that the number of outcomes produced by a choice – one or many – does indeed influence preferences: specifically, compared to the single-play frame, the multi-play frame causes reduced underweighting of rare events in the experience condition (Section 3.1.1) but greater maximizing in the description condition (Section 3.1.2).

3.2. Repeated-play vs. multi-play (distributed)

In the previous section we examined the impact of the number of *outcomes* on preference, which relied on a comparison between the single-play and multi-play (binary) conditions. In this section we examine the impact of the number of *decisions*, which relies on a comparison between the repeated-play and multi-play (distributed) preferences (i.e., the bottom row of Table 1).

3.2.1. The weighting of rare events

The first question of interest was whether rare events are weighted differentially across the different choice tasks. Fig. 4 shows the average allocation of plays to the option consistent with underweighting rare events in the description and experience formats. Note that there was no repeated-play description condition. Additionally, the description data presented in Fig. 4 did not enter into the analyses in this section but are displayed for ease of comparison with Fig. 2. As can be seen, the bars corresponding to the experience repeated-play condition are always taller than the bars corresponding to the experience multi-play condition, implying a reduced tendency to underweight rare events in the latter case. To statistically examine the effect of number of choices, we again analyzed the data via a mixed-effect model as in Section 3.1.1, but in this model Number of Outcomes was replaced with Number of Choices and Format was removed (since there was no repeated-play description condition). The dependent variable was the average percentage of choices allocated to the option consistent with underweighting rare events.

The specified model revealed a number of significant effects and produced an R^2 of .257 (adjusted $R^2 = .247$). Below we discuss some of the key significant effects; for the full list of effects and associated *p*-values please refer to Appendix F. Of central importance for the current investigation, the percentage of choices allocated to the option consistent with underweighting rare events was smaller in the multi-play (distributed) condition than in the repeated-play condition ($p < .0001$). As in the previous analysis, the type of choice problem moderated this difference: the difference in distribution of plays was much greater when the rare outcome was undesirable ($p < .001$), particularly in the loss domain ($p < .001$).

3.2.2. Maximizing tendency

Although we had no specific hypothesis, our procedure allowed us to assess whether choices made under multi-play conditions pushed participants to make more choices in line with maximizing expected value. In order to test this we replicated the mixed-model analysis described in Section 3.2.1 but this time with the dependent variable coded as the average percentage of choices allocated to the maximizing option (the data are shown in Appendix B).

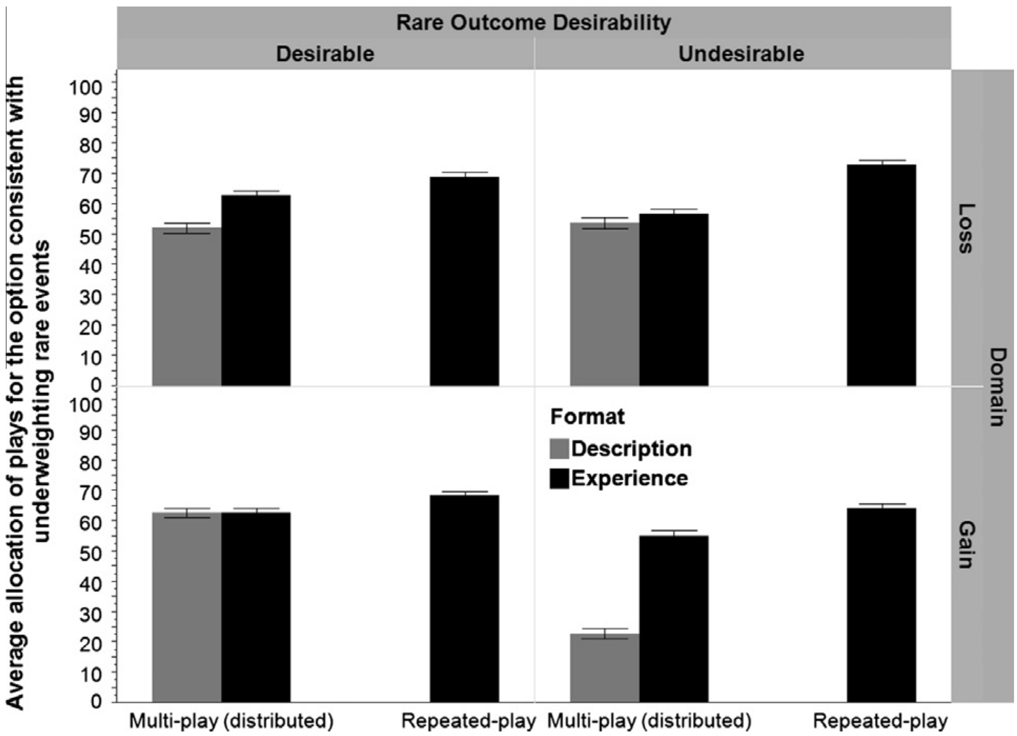

Fig. 4. The average percentage of choices allocated to the option consistent with underweighting rare events in the description and experience multi-play (distributed) and repeated-play conditions. Note that there was no description repeated-play condition. The error bars indicate the standard error of the mean.

The specified model revealed a number of significant effects and produced an R^2 of .201 (adjusted $R^2 = .190$). Below we again discuss only the key significant effect; for the full list of effects and associated *p*-values please refer to Appendix G. Most interestingly, the tendency to select the maximizing option was higher in the multi-play condition than the repeated-play condition ($p < .0001$).

The conclusion from the results reported here in Section 3.2 is that when only experience-based choices are examined the number of decisions that go on to produce multiple outcomes – one or many – does indeed influence preferences (Table 1): specifically, when making an experience-based choice, the multi-play frame causes reduced underweighting of rare events (Section 3.2.1) and greater maximizing (Section 3.2.2) relative to the repeated play frame.

4. Discussion

Why do decision-makers tend to form different preferences to objectively equivalent choices depending on whether they learn about the options from a summary description or sampled experience? Does it matter whether the choice is to be made for the short run (one outcome) or the long run (many aggregated outcomes)? We investigated the extent to which the description-experience choice gap is produced by differences in the paradigms these formats of choice are usually framed in (i.e., single-play vs. repeated-play). By testing the description and experience choice formats in a multi-play frame – where one consequential choice produces many consequential outcomes – we found a reduction in the size of the description-experience gap. The reduction was attributable to the reduced overweighting of rare events in the experience condition and greater proportion of choice maximizing in the description condition.

4.1. The description-experience choice gap

Consistent with past research, when averaged across all problems, we observed a description-experience choice gap; that is, the tendency to prefer options consistent with underweighting rare events was stronger in the experience condition than in the description condition (Hertwig & Erev, 2009; Rakow & Newell, 2010). Thus, the current data help to refute any suggestion that the description-experience gap is entirely due to the unrepresentative samples observed when learning occurs through repeated, consequential choices (this is in contrast to a choice task in which costless samples are taken; see Camilleri & Newell, 2011a).

One strength of the current experiment was that we collected both choice data (binary scale) as well as strength of preference data (graded scale, including indifference). Comparison of these two measures allowed us to judge whether the description-experience gap is, in some sense, merely an artifact of the scale used to infer choice. For example, differences in the choice data that were not reflected in the strength of preference data would suggest that the gap may be a red herring; a product of the binary scale being used. However, in all cases we found that the choice data and strength of preference data correlated strongly (see Footnote 3). Additionally, the graded strength of preference data allowed us to move away from less efficient non-parametric statistical analysis; as such, we recommend this measure to future investigators.

4.2. The "broad bracketing" effect

Another important feature of our experiment was that we collected preference data using three different elicitation formats: single-play (one decision, one outcome), multi-play (one decision, many outcomes), and repeated-play (many decisions, many outcomes; see Fig. 1). We found that underweighting of rare events in the experience condition was reduced (as was the choice gap) when repeated-play preferences were compared to multi-play distributed preferences, and also when single-play preferences were compared to multi-play binary preferences. These results are consistent with previous experiments that close the gap when contrasting description- and experience-based choices in the same choice paradigm (e.g., Camilleri & Newell, 2011a; Jessup et al., 2008). We are the first to show this using the multi-play choice task. We are also the first to be able to implicate both the number of consequential decisions and the number of consequential outcomes involved in the task as sources of the attenuation.

The structure of our problem set also allowed us to disentangle two explanations regarding the impact of moving to a multi-play format in description-based choice. According to the long-run rationality hypothesis, decision-makers are more likely to maximize – that is, prefer the higher expected value option – under the multi-play frame (Wedell, 2011). In contrast, we noted that in reviewing the literature we could not rule out the alternative explanation that the multi-play frame merely reduced the tendency to overweight rare events in cases where rare events were present. Keeping in mind that there was not strong evidence that participants in the description condition overweighted rare events to begin with, we found evidence in favor of the long-run rationality hypothesis: the multi-play format was indeed associated with a greater proportion of maximizing[5] (Wedell & Böckenholt, 1990). Importantly, however, this shift was *only* observed for those in the description-condition. We are the first to show that reframing single-play choices as multi-play does *not* tend to increase maximization when relying on one's experience of the outcome distribution.[6]

The theoretical explanation usually provided to explain the preference shift observed between single and multi-play in the description task is "broad bracketing" (Benartzi & Thaler, 1995; Thaler et al., 1997). The argument is that the single-play format induces a short run mindset that is susceptible to the effects of short run biases, namely, loss aversion. In contrast, the multi-play format induces a long

[5] Note, however, that when given the opportunity to distribute plays across both safe and risky options, participants adopted such a hedging, or diversification, strategy on 77% of occasions. In cases where alternative option's EVs are unequal, as in the current study, such distribution cannot maximize EV.

[6] Although purely explorative, in Section 3.2.2 we showed that increased maximization did occur in the experience condition when moving from *repeated*-play to multi-play frame.

run mindset that reduces the effects of loss aversion by the mental accounting act of aggregating outcomes. To make this explanation clear, we return to the problem posed by Samuelson to his colleague: a 50% chance of winning \$200 vs. 50% chance of losing \$100. A loss averse decision-maker such as Samuelson's colleague who weights losses 2.5 times as much as he weights gains would scoff at Samuelson's bet since it provides negative expected utility (Expected Utility = $(.5 \times \$200) + (.5 \times -\$100 \times 2.5) = -\$25$). However, when presented with two plays at the bet the same decision-maker might find it more appealing if it is first aggregated into the range of possible outcomes over two plays since his expected utility will no longer be negative (Expected Utility = $(.25 \times \$400) + (.5 \times \$100) + (.25 \times -\$400 \times 2.5 = 0)$.

The broad bracketing effect, induced by multi-play choice framing, may also operate in the experience format where it similarly attenuates biases prevalent with short-run choice. However, rather than operating by alleviating myopic loss aversion, the broad bracket effect may manifest itself in experience-based choice by attenuating decision-maker's reliance on small samples of recalled outcomes from memory. This explanation is couched in the idea that rare events are underweighted in experience-based choice because decision-makers tend to rely on a small sample of observed outcomes (Erev et al., 2010). Small samples are more likely to under-represent rare events in skewed distributions such as the ones used in the current experiment (Hertwig & Pleskac, 2008). For example, as Hertwig and Pleskac (2010, p. 226) explain, a decision-maker sampling 10 times from a distribution in which the rare event occurs 10% will observe the rare event more than once, less than once, or exactly once .26, .35, and .39 of the time, respectively. Thus, the person is more likely to underestimate than overestimate (.35 vs. .26, a difference of 9 percentage points) and this difference will increase with reliance on smaller samples.

The reliance on small samples can refer either to "external" samples taken from the environment or "internal" samples taken from memory (Camilleri & Newell, 2011a). Here we argue that decision-makers may rely on a larger *internal* sample of outcomes, which would necessarily reduce the under-representation of rare events. For example, going back to the example given by Hertwig and Pleskac (2010), a decision-maker who relied on a sample of 40 outcomes rather than 10 would be expected to reduce their expected tendency of rare event underestimation from 9 percentage points to 5. Thus, the observations made in our experience condition can be captured with the suggestion that repeated-play formats induce a narrow bracket and reliance on a very small set of samples whereas the multi-play choice formats induce a broad bracket and reliance on a relatively larger set of sample outcomes.[7,8] The implication of this finding is that the sequential nature of the repeated decisions in the experience paradigm is central to the underweighting choice behavior displayed in previous literature.

[7] Note that maximization would only be guaranteed if the decision-maker equally considered all of the sampled outcomes of a perfectly representative distribution.

[8] Note that this argument would not explain preference shifts in cases such as Samuelson's problem where outcomes are equally likely.

4.3. The importance of problem characteristics

A third strength of the current experiment is that the selection of choice problems used were not cherry picked from previous sets that are known to "work". Our concern with studies that use the same small set of problems is that we do not learn what characteristics of the choice problems promote or suppress the tendency to form different preferences as a function of presentation format. With that thought in mind, we constructed an original set of 32 problems that crossed several characteristics plausibly implicated in the strength of the choice gap: rare event desirability (desired vs. not desired), rare event rarity (5% vs. 10% vs. 15% vs. 20%), choice domain (gains vs. losses), and maximizing option (safe vs. risky option).

The lesson of this endeavor, which is obvious from Figs. 2 and 4, is that not all problems are created equally. We found that the size of the gap was systematically mediated by the problem type. Although there were several interactions between these variables (see Appendices C and E), the one that stood out was Domain (Gain vs. Loss) by Rare Outcome Desirability (Desirable vs. Undesirable). Specifically, the description-experience choice gap was only obvious in the gain domain when the rare event was undesirable and in the loss domain when the rare event was desirable. These results are not entirely consistent with the findings of Abdellaoui, L'Haridon, and Paraschiv (2011), who concluded that rare event underweighting in experience-based choice (when compared to description-based choice) is greater in the gain domain but no different in the loss domain. The reasons for this disparity remains unclear, although entirely different methods were used in the two studies: here we inferred weighting from choice behavior whereas Abdellaoui, L'Haridon, and Paraschiv (2011) relied on determining certainty equivalents to directly assess the parameters of Prospect Theory. An important take-home message is that studies should seek to systematically vary the properties of the choice problems they use and be mindful of such properties when making comparisons across different studies.

According to Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), when making description-based choices, decision-makers are risk *seeking* when faced with desirable but rare wins and risk *averse* when faced with undesirable but rare losses. This formulation has been used to explain why the same person would purchase both lottery tickets and insurance. In datasets like our own, such a "fourfold" pattern would be revealed through description-based choice behavior consistent with overweighting rare events. In our data we did not observe such behavior (Figs. 2–4); our participants were apparently not interested in lottery tickets or insurance. The failure of PT to account for the description format preferences as expected by the fourfold pattern is also revealed by our model fitting exercise, summarized in Fig. 3, which shows no clear-cut region of best fit, specifically in the region corresponding to probability weighting parameters less than 1.

An alternative account that may do somewhat better in explaining our description-based choice data is regret avoidance. According to this explanation, decision-makers approach choices with the intent of minimizing their possible future regret (Josephs, Larrick, Steele, & Nisbett, 1992). Note that there is an asymmetry between the gains and losses domain (Zeelenberg et al., 1996): In the gain domain with rare big wins, the risky option represents a *high* regret potential because the decision-maker will frequently obtain the undesirable outcome (i.e., no lottery prize) and also know for certain that they could have done better selecting the safe option (e.g., by buying a burger). In contrast, selecting the safe option represents a *low* regret potential because the decision-maker will never obtain the undesirable outcome and also never know what the outcome from the risky option would have been. Thus, decision-makers are expected to prefer the safe option and our participants had a strong preference to do this. In contrast, in the loss domain with rare big losses, the risky option represents a *medium* regret potential because the decision-maker will infrequently obtain the undesirable rare event (e.g., disaster) and know for certain that they could have done better selecting the safe option (i.e., buy insurance). Importantly, selecting the safe option also represents a *medium* regret potential because the decision-maker will always obtain an undesirable outcome but never know what the alternative outcome from the risky option would have been. Thus, decision-makers are expected to be largely indifferent between options and our participants were on average indifferent to such problems.

5. Conclusions

Our experiment adds to the literature demonstrating that decision biases are attenuated when decisions are framed in the long run. Previous work has shown that such long-run framing can reduce the impact of myopic loss aversion in description-based choice. Here we have argued that long-run framing may also reduce the reliance on small samples of outcomes in experience-based choice. One implication of our findings is that preference reversals produced by different formats of choice – specifically, description and experience – may be reduced when decision-makers make a single choice that will produce many consequential, aggregated outcomes.

Our observations may also have practical implications for choices that are made repeatedly, including investment allocation and food choices. The results of this study suggest that each individual choice in a sequence of similar repeated choices will tend to underweight the possibility of rare events, which can in some cases lead to sub-optimal outcomes in the long run (e.g., over-investment in bonds, over-indulgence in chocolate mousse cake). Removing the sequential nature of the choices and requiring a single allocation choice that will have multiple realizations over a long time horizon may overcome at least part of this bias.

Acknowledgments

We thank Rick Larrick, Timothy Schofield, Kevin Bird, Christoph Ungemach, Brett Hayes, and Fred Westbrook for their valuable advice. The support of the Australian Endeavour Award to the first author, and the Australian Research Council Discovery Grant (DP110100797) and Future Fellowship (FT110100151) awarded to the second author are gratefully acknowledged.

Appendix A

Choice problem	Domain	Rare outcome rarity (%)	Rare event desirability	Maximizing option	Option favored if rare events under weighted	Safe outcome (*S*)	Probability (*S*)	Expected value (*S*)	Risky outcome (*R*)	Probability (*RS*)	Expected Value (*R*)
1	Loss	20	Undesirable	Risky	Risky	−20.1	1.00	−20.1	−95.5	0.20	−19.1
2	Gain	20	Desirable	Safe	Safe	20.1	1.00	20.1	95.5	0.20	19.1
3	Gain	20	Undesirable	Safe	Risky	21.3	1.00	21.3	25.4	0.80	20.3
4	Loss	20	Desirable	Risky	Safe	−21.3	1.00	−21.3	−25.4	0.80	−20.3
5	Loss	15	Undesirable	Safe	Risky	−22.5	1.00	−22.5	−156.7	0.15	−23.5
6	Gain	15	Desirable	Risky	Safe	22.5	1.00	22.5	156.7	0.15	23.5
7	Gain	15	Undesirable	Risky	Risky	23.7	1.00	23.7	29.1	0.85	24.7
8	Loss	15	Desirable	Safe	Safe	−23.7	1.00	−23.7	−29.1	0.85	−24.7
9	Loss	10	Undesirable	Risky	Risky	−24.9	1.00	−24.9	−239.0	0.10	−23.9
10	Gain	10	Desirable	Safe	Safe	24.9	1.00	24.9	239.0	0.10	23.9
11	Gain	10	Undesirable	Safe	Risky	26.1	1.00	26.1	27.9	0.90	25.1
12	Loss	10	Desirable	Risky	Safe	−26.1	1.00	−26.1	−27.8	0.90	−25.0
13	Loss	5	Undesirable	Safe	Risky	−27.3	1.00	−27.3	−566.0	0.05	−28.3

(*continued on next page*)

Appendix A (*continued*)

Choice problem	Domain	Rare outcome rarity (%)	Rare event desirability	Maximizing option	Option favored if rare events under weighted	Safe outcome (*S*)	Probability (*S*)	Expected value (*S*)	Risky outcome (*R*)	Probability (*RS*)	Expected Value (*R*)
14	Gain	5	Desirable	Risky	Safe	27.3	1.00	27.3	566.0	0.05	28.3
15	Gain	5	Undesirable	Risky	Risky	28.5	1.00	28.5	31.1	0.95	29.5
16	Loss	5	Desirable	Safe	Safe	−28.5	1.00	−28.5	−31.1	0.95	−29.5
17	Loss	20	Undesirable	Safe	Risky	−29.7	1.00	−29.7	−153.5	0.20	−30.7
18	Gain	20	Desirable	Risky	Safe	29.7	1.00	29.7	153.5	0.20	30.7
19	Gain	20	Undesirable	Risky	Risky	30.9	1.00	30.9	39.9	0.80	31.9
20	Loss	20	Desirable	Safe	Safe	−30.9	1.00	−30.9	−39.9	0.80	−31.9
21	Loss	15	Undesirable	Risky	Risky	−32.1	1.00	−32.1	−207.3	0.15	−31.1
22	Gain	15	Desirable	Safe	Safe	32.1	1.00	32.1	207.3	0.15	31.1
23	Gain	15	Undesirable	Safe	Risky	33.3	1.00	33.3	38.0	0.85	32.3
24	Loss	15	Desirable	Risky	Safe	−33.3	1.00	−33.3	−38	0.85	−32.3
25	Loss	10	Undesirable	Safe	Risky	−34.5	1.00	−34.5	−355.0	0.10	−35.5
26	Gain	10	Desirable	Risky	Safe	34.5	1.00	34.5	355.0	0.10	35.5
27	Gain	10	Undesirable	Risky	Risky	35.7	1.00	35.7	40.8	0.90	36.7
28	Loss	10	Desirable	Safe	Safe	−35.7	1.00	−35.7	−40.7	0.90	−36.6
29	Loss	5	Undesirable	Risky	Risky	−36.9	1.00	−36.9	−718.0	0.05	−35.9
30	Gain	5	Desirable	Safe	Safe	36.9	1.00	36.9	718.0	0.05	35.9
31	Gain	5	Undesirable	Safe	Risky	38.1	1.00	38.1	39.1	0.95	37.1
32	Loss	5	Desirable	Risky	Safe	−38.1	1.00	−38.1	−39.1	0.95	−37.1

Appendix B

Choice problem	Description					Experience						
	Single-play		Multi-play (binary)		Multi-play (distributed)	Single-play		Multi-play (binary)		Multi-play (distributed)	Repeated-play
	Choice[a]	SoP[b]	Choice[a]	SoP[b]	Choice[c]	Choice[a]	SoP[b]	Choice[a]	SoP[b]	Choice[c]	Choice[c]
1	0.56	4.47	0.56	4.53	0.56	0.68	4.89	0.59	4.35	0.56	0.72
2	0.15	1.96	0.08	1.74	0.17	0.50	3.90	0.54	4.10	0.51	0.62
3	0.60	4.51	0.51	4.07	0.49	0.75	5.30	0.63	4.63	0.60	0.75
4	0.22	2.44	0.20	2.29	0.27	0.52	3.95	0.59	4.41	0.54	0.61
5	0.62	4.47	0.60	4.44	0.62	0.79	5.21	0.68	4.82	0.62	0.68
6	0.19	1.96	0.19	2.26	0.22	0.46	3.79	0.38	3.38	0.41	0.52
7	0.47	3.94	0.53	4.00	0.53	0.74	5.13	0.56	4.43	0.59	0.77
8	0.33	2.96	0.35	3.14	0.39	0.70	5.12	0.82	5.56	0.69	0.77
9	0.55	4.18	0.45	3.73	0.48	0.68	4.80	0.44	3.64	0.46	0.69
10	0.17	2.25	0.39	3.11	0.29	0.63	4.68	0.65	4.68	0.61	0.70
11	0.58	4.37	0.75	4.86	0.63	0.67	4.98	0.40	3.93	0.53	0.68
12	0.19	2.17	0.12	1.85	0.15	0.50	4.12	0.55	4.38	0.50	0.64
13	0.69	4.96	0.52	3.96	0.50	0.59	4.33	0.51	4.12	0.56	0.76
14	0.26	2.47	0.30	2.66	0.25	0.63	4.78	0.59	4.43	0.59	0.62
15	0.51	3.98	0.53	4.17	0.50	0.75	5.06	0.63	4.54	0.62	0.78
16	0.10	1.77	0.08	1.60	0.12	0.53	3.98	0.57	4.15	0.53	0.64
17	0.20	2.46	0.30	2.94	0.33	0.33	3.08	0.33	3.23	0.39	0.37
18	0.55	4.14	0.66	4.64	0.55	0.30	2.80	0.27	3.07	0.35	0.32
19	0.30	2.81	0.35	3.26	0.42	0.19	2.38	0.33	3.00	0.34	0.31
20	0.35	3.39	0.46	3.70	0.43	0.31	3.00	0.23	2.73	0.36	0.29

Appendix B (*continued*)

Choice problem	Description					Experience					
	Single-play		Multi-play (binary)		Multi-play (distributed)	Single-play		Multi-play (binary)		Multi-play (distributed)	Repeated-play
	Choice[a]	SoP[b]	Choice[a]	SoP[b]	Choice[c]	Choice[a]	SoP[b]	Choice[a]	SoP[b]	Choice[c]	Choice[c]
21	0.25	2.58	0.33	2.98	0.32	0.18	2.52	0.28	2.74	0.32	0.29
22	0.55	4.30	0.66	4.66	0.63	0.28	2.93	0.48	3.74	0.43	0.35
23	0.30	3.11	0.41	3.57	0.41	0.17	2.19	0.38	3.04	0.34	0.27
24	0.34	3.18	0.39	3.41	0.39	0.28	2.87	0.26	3.07	0.35	0.30
25	0.29	3.05	0.39	3.39	0.45	0.28	2.80	0.41	3.43	0.44	0.36
26	0.43	3.43	0.53	3.90	0.43	0.22	2.67	0.33	3.08	0.36	0.34
27	0.26	2.84	0.50	3.74	0.38	0.22	2.73	0.34	3.17	0.39	0.35
28	0.41	3.55	0.43	3.61	0.44	0.24	2.62	0.28	2.88	0.32	0.25
29	0.23	2.82	0.52	4.11	0.44	0.24	2.53	0.43	3.45	0.41	0.28
30	0.51	4.02	0.43	3.65	0.43	0.19	2.45	0.30	2.94	0.29	0.24
31	0.30	2.87	0.20	2.59	0.24	0.13	2.13	0.27	2.96	0.34	0.26
32	0.57	4.28	0.59	4.50	0.54	0.38	3.19	0.42	3.81	0.51	0.38

[a] Proportion selecting the risky option (0 = Safe option; 1 = Risky option).
[b] Average strength of preference for the risky option (1 = "Strong preference for the safe option", 4 = "Indifference", 7 = "Strong preference for the risky option").
[c] Percentage of plays distributed to the risky option.

Appendix C

Source	*df*	*F*	*p*
Format	1	318.9017	<.0001
Number of Outcomes	1	34.6985	<.0001
Domain	1	44.4157	<.0001
Maximizing Option	1	2.0993	0.1474
Rare Outcome Probability	3	1.7416	0.1562
Rare Outcome Desirability	1	60.5572	<.0001
Format * Number of Outcomes	1	3.4512	0.0633
Format * Domain	1	16.2654	<.0001
Format * Maximizing Option	1	1.5422	0.2143
Format * Rare Outcome Probability	3	0.4571	0.7123
Format * Rare Outcome Desirability	1	36.2879	<.0001
Number of Outcomes * Domain	1	1.4846	0.2231
Number of Outcomes * Maximizing Option	1	0.0649	0.7990
Number of Outcomes * Rare Outcome Probability	3	0.4955	0.6854
Number of Outcomes * Rare Outcome Desirability	1	7.3733	0.0066
Domain * Maximizing Option	1	15.5375	<.0001
Domain * Rare Outcome Probability	3	4.4121	0.0042
Domain * Rare Outcome Desirability	1	287.4258	<.0001
Maximizing Option * Rare Outcome Probability	3	0.7336	0.5318
Maximizing Option * Rare Outcome Desirability	1	76.0661	<.0001
Rare Outcome Probability * Rare Outcome Desirability	3	0.8873	0.4468
Format * Number of Outcomes * Domain	1	1.8874	0.1695
Format * Number of Outcomes * Maximizing Option	1	1.8727	0.1712
Format * Number of Outcomes * Rare Outcome Probability	3	0.5239	0.6658
Format * Number of Outcomes * Rare Outcome Desirability	1	0.4785	0.4891
Format * Domain * Maximizing Option	1	0.1533	0.6954
Format * Domain * Rare Outcome Probability	3	1.5090	0.2100
Format * Domain * Rare Outcome Desirability	1	167.9844	<.0001

(*continued on next page*)

Appendix C (*continued*)

Source	*df*	*F*	*p*
Format * Maximizing Option * Rare Outcome Probability	3	1.9944	0.1125
Format * Maximizing Option * Rare Outcome Desirability	1	3.8139	0.0509
Format * Rare Outcome Probability * Rare Outcome Desirability	3	2.8069	0.0382
Number of Outcomes * Domain * Maximizing Option	1	0.4729	0.4917
Number of Outcomes * Domain * Rare Outcome Probability	3	0.8311	0.4765
Number of Outcomes * Domain * * Rare Outcome Desirability	1	13.0476	0.0003
Number of Outcomes * Maximizing Option * Rare Outcome Probability	3	1.0303	0.3779
Number of Outcomes * Maximizing Option * Rare Outcome Desirability	1	8.2427	0.0041
Number of Outcomes * Rare Outcome Probability * Rare Outcome Desirability	3	0.1135	0.9522
Domain * Maximizing Option * Rare Outcome Probability	3	2.6518	0.0470
Domain * Maximizing Option * Rare Outcome Desirability	1	1.5103	0.2191
Domain * Rare Outcome Probability * Rare Outcome Desirability	3	1.3929	0.2429
Maximizing Option * Rare Outcome Probability * Rare Outcome Desirability	3	9.6063	<.0001
Component Order	5	2.1292	0.0590
Problem Order	15	2.4336	0.0015

Appendix D

Prospect Theory calculates the weighted value of each option and then chooses the most attractive alternative (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). The expected value of each outcome, j, is given by:

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

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

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

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

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

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

Appendix E

Source	*df*	*F*	*p*
Format	1	3.9223	0.0477
Number of Outcomes	1	7.6886	0.0056
Domain	1	1.5172	0.2181
Maximizing Option	1	308.5748	<.0001
Rare Outcome Probability	3	8.5589	<.0001
Rare Outcome Desirability	1	1.3625	0.2445
Format * Number of Outcomes	1	0.4871	0.4853
Format * Domain	1	0.0030	0.9567
Format * Maximizing Option	1	34.2518	<.0001

Appendix E (*continued*)

Source	*df*	*F*	*p*
Format * Rare Outcome Probability	3	2.3353	0.0718
Format * Rare Outcome Desirability	1	2.5816	0.1082
Number of Outcomes * Domain	1	0.0536	0.8169
Number of Outcomes * Maximizing Option	1	6.4740	0.0110
Number of Outcomes * Rare Outcome Probability	3	0.2283	0.8768
Number of Outcomes * Rare Outcome Desirability	1	0.0641	0.8001
Domain * Maximizing Option	1	253.9779	<.0001
Domain * Rare Outcome Probability	3	0.7271	0.5357
Domain * Rare Outcome Desirability	1	13.9931	0.0002
Maximizing Option * Rare Outcome Probability	3	0.4980	0.6837
Maximizing Option * Rare Outcome Desirability	1	184.1840	<.0001
Rare Outcome Probability * Rare Outcome Desirability	3	0.6348	0.5925
Format * Number of Outcomes * Domain	1	0.0970	0.7555
Format * Number of Outcomes * Maximizing Option	1	0.3511	0.5535
Format * Number of Outcomes * Rare Outcome Probability	3	0.2016	0.8953
Format * Number of Outcomes * Rare Outcome Desirability	1	1.5453	0.2139
Format * Domain * Maximizing Option	1	130.8028	<.0001
Format * Domain * Rare Outcome Probability	3	0.3994	0.7534
Format * Domain * Rare Outcome Desirability	1	0.0113	0.9155
Format * Maximizing Option * Rare Outcome Probability	3	2.3440	0.0710
Format * Maximizing Option * Rare Outcome Desirability	1	283.0177	<.0001
Format * Rare Outcome Probability * Rare Outcome Desirability	3	2.7438	0.0416
Number of Outcomes * Domain * Maximizing Option	1	11.2875	0.0008
Number of Outcomes * Domain * Rare Outcome Probability	3	0.1343	0.9396
Number of Outcomes * Domain * Rare Outcome Desirability	1	0.4489	0.5029
Number of Outcomes * Maximizing Option * Rare Outcome Probability	3	0.0902	0.9655
Number of Outcomes * Maximizing Option * Rare Outcome Desirability	1	30.1360	<.0001
Number of Outcomes * Rare Outcome Probability * Rare Outcome Desirability	3	1.0086	0.3877
Domain * Maximizing Option * Rare Outcome Probability	3	1.3011	0.2722
Domain * Maximizing Option * Rare Outcome Desirability	1	42.2109	<.0001
Domain * Rare Outcome Probability * Rare Outcome Desirability	3	2.0727	0.1016
Maximizing Option * Rare Outcome Probability * Rare Outcome Desirability	3	1.4587	0.2237
Component Order	5	0.7814	0.5629
Problem Order	15	1.3707	0.1519

Appendix F

Source	*df*	*F*	*p*
Number of Choices	1	259.9699	<.0001
Domain	1	31.8478	<.0001
Maximizing Option	1	1.1703	0.2794
Rare Outcome Probability	3	2.8384	0.0366
Rare Outcome Desirability	1	13.2862	0.0003
Number of Choices * Domain	1	0.0104	0.9189
Number of Choices * Maximizing Option	1	0.0618	0.8037
Number of Choices * Rare Outcome Probability	3	0.1405	0.9358
Number of Choices * Rare Outcome Desirability	1	43.9409	<.0001
Domain * Maximizing Option	1	11.2327	0.0008
Domain * Rare Outcome Probability	3	4.7423	0.0026
Domain * Rare Outcome Desirability	1	61.7990	<.0001
Maximizing Option * Rare Outcome Probability	3	1.6583	0.1738

(*continued on next page*)

Appendix F (*continued*)

Source	*df*	*F*	*p*
Maximizing Option * Rare Outcome Desirability	1	33.6927	<.0001
Rare Outcome Probability * Rare Outcome Desirability	3	2.4596	0.0609
Number of Choices * * Domain*Maximizing Option	1	0.0142	0.9053
Number of Choices * Domain * Rare Outcome Probability	3	0.1319	0.9411
Number of Choices * Domain * Rare Outcome Desirability	1	11.7290	0.0006
Number of Choices * Maximizing Option * Rare Outcome Probability	3	0.5079	0.6768
Number of Choices * Maximizing Option * Rare Outcome Desirability	1	16.7358	<.0001
Number of Choices * Rare Outcome Probability * Rare Outcome Desirability	3	1.3517	0.2557
Domain * Maximizing Option * Rare Outcome Probability	3	2.0780	0.1009
Domain * Maximizing Option * Rare Outcome Desirability	1	4.3139	0.0379
Domain * Rare Outcome Probability * Rare Outcome Desirability	3	3.4791	0.0153
Maximizing Option * Rare Outcome Probability * Rare Outcome Desirability	3	4.4125	0.0042
Component Order	5	1.7725	0.1149
Problem Order	15	3.4471	<.0001

Appendix G

Source	*df*	*F*	*p*
Number of Choices	1	15.1772	<.0001
Domain	1	2.8511	0.0914
Maximizing Option	1	46.6410	<.0001
Rare Outcome Probability	3	4.5235	0.0036
Rare Outcome Desirability	1	0.5765	0.4485
Number of Choices * Domain	1	0.2165	0.6418
Number of Choices * Maximizing Option	1	39.5348	<.0001
Number of Choices * Rare Outcome Probability	3	0.1211	0.9477
Number of Choices * Rare Outcome Desirability	1	0.0307	0.8609
Domain * Maximizing Option	1	59.5036	<.0001
Domain * Rare Outcome Probability	3	0.4620	0.7088
Domain * Rare Outcome Desirability	1	11.4925	0.0007
Maximizing Option * Rare Outcome Probability	3	2.4650	0.0605
Maximizing Option * Rare Outcome Desirability	1	546.6707	<.0001
Rare Outcome Probability * Rare Outcome Desirability	3	1.3640	0.2518
Number of Choices * Domain * Maximizing Option	1	9.4236	0.0022
Number of Choices * Domain * Rare Outcome Probability	3	0.1899	0.9033
Number of Choices * Domain * Rare Outcome Desirability	1	0.0495	0.8239
Number of Choices * Maximizing Option * Rare Outcome Probability	3	1.1759	0.3173
Number of Choices * Maximizing Option * Rare Outcome Desirability	1	234.5334	<.0001
Number of Choices * Rare Outcome Probability * Rare Outcome Desirability	3	0.7143	0.5434
Domain * Maximizing Option * Rare Outcome Probability	3	3.4483	0.0159
Domain * Maximizing Option * Rare Outcome Desirability	1	31.7140	<.0001
Domain * Rare Outcome Probability * Rare Outcome Desirability	3	1.6613	0.1731
Maximizing Option * Rare Outcome Probability * Rare Outcome Desirability	3	2.6603	0.0465
Component Order	5	1.2208	0.2964
Problem Order	15	1.5492	0.0797

References

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two Prospect Theory specifications? *Management Science, 57*, 1879–1895. http://dx.doi.org/10.1287/mnsc.1110.1368.

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. http://dx.doi.org/10.1002/bdm.443.

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics, 110*, 73–92.

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

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

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

Chen, Y.-J., & Corter, J. E. (2006). When mixed options are preferred in multiple-trial decision making. *Journal of Behavioral Decision Making, 19*, 17–42.

Demidenko, E. (2004). *Mixed models: Theory and applications*. New York: John. Wiley & Sons.

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.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451–482.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

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.

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Charter & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 209–236). Oxford, England: Oxford University Press.

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

Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the self from the negative consequences of risky decisions. *Journal of Personality and Social Psychology, 62*, 26–37.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

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

Keren, G. (1991). Additional tests of utility theory under unique and repeated conditions. *Journal of Behavioral Decision Making, 4*, 297–304.

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

Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes, 117*, 286–295.

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

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

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

Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (1998). Judgment and decision making. *Annual Review of Psychology, 49*, 447–477.

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.

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

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

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*, 491–517.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, & happiness*. New Haven: Yale University Press.

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

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

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

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53–85.

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.

Wedell, D. H. (2011). Evaluations of single- and repeated-play gambles. In J. J. Cochran (Ed.), *Wiley encyclopedia of operations research and management science*. John Wiley & Sons, Inc..

Zeelenberg, M., Beattie, J., van der Pligt, J., & de Vries, N. (1996). Consequences of regret aversion: Effects of expected feedback on risky decision making. *Organizational Behavior and Human Decision Processes, 65*, 148–158.