Journal of Behavioral Decision Making, *J. Behav. Dec. Making*, **29**: 489–498 (2016)
Published online 16 July 2015 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1893

Experience-Based Decisions Favor Riskier Alternatives in Large Sets

TAKAO NOGUCHI[1]* and THOMAS T. HILLS[2]
[1]*Department of Experimental Psychology, University College London, UK*
[2]*Department of Psychology, University of Warwick, UK*

ABSTRACT

Research on risky choice has been predominantly based on studies of choices between two alternatives, but their findings are often generalized to environments with more than two alternatives. One prominent claim of this research is that choices differ with respect to risk when alternatives are described (the description paradigm) as opposed to sampled (the sampling paradigm). Here, we demonstrate that the difference in choices is sensitive to the number of alternatives in a choice set: with a growth in set size, alternatives with rare payoffs become more likely to be chosen in the sampling paradigm. This increased risk-taking is due to the statistical property of payoffs: with a growth in set size, at least one riskier alternative becomes much more likely to deliver an attractive sample payoff at a higher frequency than its underlying probability. The increased risk-taking eventually diminishes the difference between the description and the sampling paradigms in the gain domain. Further, we show that the increase in risk-taking is difficult to avoid, even if each alternative is thoroughly sampled. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS description–experience gap; decision from experience; search-amplified risk; sampling error; too much choice

Over the past decade, research with two-alternative environments has led to the observation that in decisions from experience—in particular, the sampling paradigm—where a choice is made after sampling a series of payoffs (such as \$0, \$0, \$0, \$9, and \$0), individuals make a choice as if they under-weight small probabilities (Hertwig, Barron, Weber, & Erev, 2004). This under-weighting has been juxtaposed against over-weighting of small probabilities in decisions from description (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), where payoffs and their probabilities are described (e.g., \$9 with a 10% probability, otherwise nothing). This difference in the weighting of small probabilities—termed the description–experience gap (Hertwig & Erev, 2009)—has been rigorously examined and consistently confirmed in two-alternative environments (e.g., Abdellaoui, L'Haridon, & Paraschiv, 2011; Erev et al., 2010; Gonzalez & Dutt, 2011; Gottlieb, Weiss, & Chapman, 2007; Hilbig & Gloeckner, 2011; Lajarraga, Hertwig, & Gonzalez, 2012; Ungemach, Chater, & Stewart, 2009).

The description–experience gap has been used to explain a variety of phenomena related to decision making outside the laboratory, including those involving the financial crisis (Hertwig & Erev, 2009) and perceived terrorist threats (Yechiam, Barron, & Erev, 2005). The basic logic is that in decisions from experience, rare events are often under-weighted, which leads individuals to discount the likelihood of financial risk or potential threats of violence. Although the aforementioned examples represent loss (negative) domains, this observation can be framed for gain (positive) domains as well. As an illustration, suppose an individual is assessing how likely he is to lose weight from a specific diet. One method of assessment is to recall other individuals who have tried the diet (Galesic, Olsson, & Rieskamp, 2012; Tversky & Kahneman, 1973). If the diet rarely leads to weight loss, then the probability of weight loss may be under-weighted and the individual may infer that the diet is a waste of time.

The previous three examples generalize the description–experience gap outside the laboratory. This generalization requires that the under-weighting and a subsequent choice be independent of other features in naturalistic environments. One of the most prominent features associated with choices outside the laboratory is set size, where choices often involve tens, if not hundreds, of alternatives (e.g., pension plans and pasta) (Schwartz, 2004). Continuing with our diet example, an individual should under-weight the probability of weight loss and choose not to try the diet regardless of the number of diets he considers. As we demonstrate in the succeeding text, however, this intuition is misleading.

Previous research indicates that the under-weighting of small probabilities can be sensitive to set size. Ert and Erev (2007), for example, report that when individuals can observe foregone payoffs from alternatives they did not choose, a larger set tends to lead the individuals to choose the alternative that, most recently, delivered the largest payoff. This tendency to choose an alternative with a large payoff is, in this case, consistent with over-weighting of small probabilities (Ert & Erev, 2007). Similarly, Hills, Noguchi and Gibbert (2013) found that larger and more diverse sets lead individuals to choose alternatives that delivered larger payoffs, which were associated with smaller probabilities of payoff. The present study extends these studies by testing the potential role of set size on the description–experience gap.

A general explanation can be provided formally as follows: suppose an individual is faced with two alternatives, a safer alternative that delivers L for sure, and a riskier alternative that delivers H with probability p and 0 with probability $1-p$. Let the expected payoffs of the two alternatives be equal, such that $H > L$, and $H \times p = L$. If each alternative is sampled once and the alternative with the highest sample is chosen, then the probability of choosing H, $P(H)$,

*Correspondence to: Takao Noguchi, Department of Experimental Psychology, University College London, WC1E 6BT, UK.E-mail: t.noguchi@ucl.ac.uk

is p. If each alternative is replicated, such that there are two safer and two riskier alternatives, then $P(H) = 1 - (1 - p)^2$. More generally, when each alternative is replicated n times, $P(H) = 1 - (1 - p)^n$, which approaches 1 as n increases (Figure 1a). Thus, with a growth in set size, the probability of choosing a riskier alternative increases. Although this example makes potentially unrealistic assumptions (e.g., sampling each alternative only once), in the remainder of this article, we demonstrate that the results of this example are robust.

We will also investigate the contrasting loss domain, where a payoff is either zero or negative. Here, theoretical accounts of previous findings imply that set size is less likely to influence risk-taking (Ert & Erev, 2007). Indeed, in the previous mathematical example, if H and L were losses, with $H < L$, then individuals would choose riskier alternatives with $P(H) = 1 - p^n$, and as in the gain domain, the probability of choosing a riskier alternative approaches 1 as n increases (Figure 1b). Unlike in the gain domain, however, $P(H)$ is already above .5 when p is less than .5. Thus, even in a small set, riskier alternatives are more likely to be chosen than safe alternatives. Therefore, we hypothesize that set size in the loss domain will not have the same impact on the description–experience gap as in the gain domain.

In addition to investigating the influence of set size on choices, the present study will further investigate a prominent underlying cause: sampling error. Formally, sampling error is defined as the difference between a mean sample and an expected payoff (Hertwig & Erev, 2009). For example, if an alternative with a .16 probability of a \$5.19 payoff delivers a series of samples, 0.00, 0.00, and 5.19, the mean sample is $(0.00 + 0.00 + 5.19)/3 = 1.73$. Because this alternative has the expected payoff of $.16 \times 5.19 = 0.8304$, the sampling error is $1.73 - 0.8304 = 0.8996$. Although the description–experience gap has been attributed to a variety of factors (e.g., Hertwig & Erev, 2009; Hills & Hertwig, 2010), previous studies with two-alternative environments have demonstrated that sampling error plays a large role in influencing choices. Ungemach et al. (2009), for example, demonstrated that when sampling error is minimized, the size of the description–experience gap is substantially reduced in the two-alternative environment. Here, we will investigate how sampling error changes with a growth in set size.

In what follows, we first describe two experiments that examine the effects of set size on risk-taking and the role of sampling error. Then, using a simulation, we examine how robust the effects of set size are on sampling error.

METHOD

In Experiments 1 and 2, we manipulated set size in both description and sampling paradigms, for both gain and loss domains. Both experiments employed a 2 (between-participant, set size: small or large) × 2 (between-participant, paradigm: description or sampling) × 2 (within-participant, domain: gain or loss) design.

Following Experiment 1, Experiment 2 was carried out to replicate the findings with a different population in a different location. American participants took part in Experiment 1 online, and British university students took part in Experiment 2 in the laboratory. Additionally, to ensure that the findings do not depend on a particular structure of the payoffs, we altered the structures of alternatives between Experiments 1 and 2, as explained in the succeeding text.

Participants

In Experiment 1, 131 participants (73 men, 56 women, and 2 unspecified) were recruited through Mechanical Turk (http://www.mturk.com). Their age ranged from 18 to 69 years with a mean of 30.63. In Experiment 2, 101 students (47 men, 53 women, and 1 unspecified) were recruited through the participant panel at the University of Warwick. Their age ranged from 18 to 52 years with a mean of 22.7.

Previous studies (e.g., Hills et al., 2013) suggest that the predicted effect can be sufficiently detected with 100 participants. Thus, in advance of collecting the data, we decided to test 100 participants but over-recruited due to technical reasons. Here, we examine the data collected from all participants.

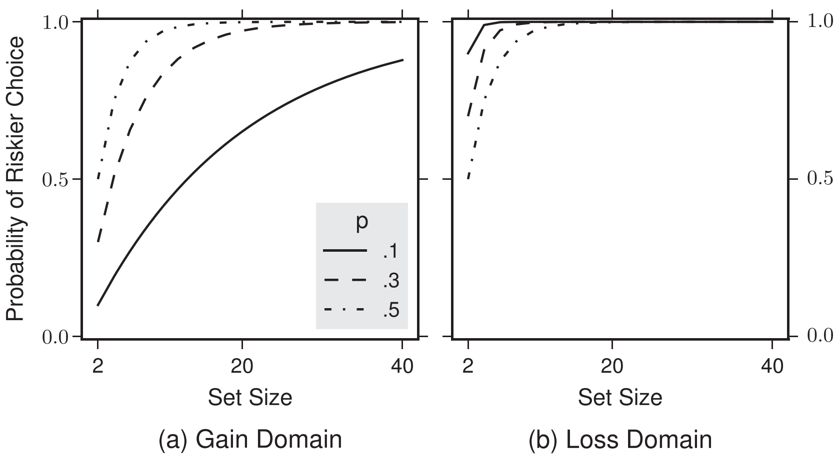

Figure 1. The influence of set size on choices for risky alternatives: (a) gain domain and (b) loss domain. Here, half of the alternatives in a choice set is always risky, delivering payoff with probability p; in the set size of 40, for example, 20 alternatives are risky and the other 20 are safe

DOI: 10.1002/bdm

Apparatus

Alternatives were independently and randomly generated for each trial for each participant. Half of the alternatives within a trial (1 alternative in a small set and 16 alternatives in a large set) were safer alternatives, and the other half were riskier. Each alternative delivered a non-zero payoff with a certain probability and otherwise nothing. Safer and riskier alternatives were characterized by different probabilities of non-zero payoff.

A safer alternative had a probability of non-zero payoff, uniformly distributed between .800 and .995. Similarly, a riskier alternative had a probability of non-zero payoff, uniformly distributed between .005 and .200. This range of probability follows the definition of rare events by Hertwig et al. (2004). We drew a random number from the uniform distributions to determine the probability of non-zero payoff, independently for each alternative and for each participant. Thus, in a particular trial, for example, a safer alternative could have .943 probability of non-zero payoff.

Also, all the safer alternatives within a trial had the same expected payoff—a random draw from a uniform distribution between 0.50 and 1.00 for the gain domain and between −0.50 and −1.00 for the loss domain. To derive the expected payoff for the riskier alternatives, the expected payoff for the safer alternatives was multiplied by 0.9 in Experiment 1 and by 1.1 in Experiment 2. We used the different multipliers in each experiment to alter the structures of alternatives. Continuing the previous example, the safer alternative with .943 probability of non-zero payoff could have the expected payoff of 0.83. Then, all the riskier alternatives in the same trial would have the expected payoff of $0.83 \times 0.9 = 0.747$ in Experiment 1. Thus, in comparison with the safe alternatives, in Experiment 1, riskier alternatives were associated with lower expected values, and in Experiment 2, riskier alternatives were associated with higher expected values.

The expected payoff for an alternative was then divided by the probability of non-zero payoff to derive the non-zero payoff amount. Thus, if the safer alternative had an expected payoff of 0.83 and a .943 probability of non-zero payoff, the amount of the non-zero payoff was $0.83/.943 = 0.880$. The probability and the amount of payoff were rounded to the nearest two decimal points. Thus in the above example, the alternative has a .94 probability of 0.88 payoff and a $1-.94=.06$ probability of 0.00 payoff. Example alternatives generated with this procedure are listed in Table A1.

Each alternative was presented as a box on the screen, and a set of alternatives was presented to participants as an array of boxes. An example screenshot is given in Figure 2. The left panel in this figure illustrates a small set with two alternatives, and the right panel illustrates a large set with 32 alternatives. Locations of alternatives were randomized independently for each trial and for each participant.

Procedure

Participants were instructed that their payment would depend on their choices during the experiment (see Appendix B for the instruction). In each experiment, participants made six choices, three involving gains and three involving losses. The gain and loss trials were interleaved and presented in a random order. At each trial, participants saw either 2 or 32 alternatives and were allowed to draw samples from the alternatives as many times as they wanted before choosing one of the alternatives. Every time a sample was drawn, information about the alternative was presented for 500 ms. In the description paradigm, the information displayed the probability and the amount of non-zero payoff (e.g., 94%, \$0.88; Figure 2a). In the sampling paradigm, the information presented was a sample, randomly drawn with replacement from the payoff distribution associated with that alternative (Figure 2b). For example, about 94 out of 100 samples from an alternative with a 94% probability of \$0.88 were \$0.88 and the rest were \$0.00. Participants did not learn about the payoffs from their final choices until the end of the experiment. In Experiment 1, the payoffs from the six choices were summed and added to the base fee of \$1.00. After adding the base fee, the fee ranged from \$0.00 to \$14.57, with a mean of \$1.80. In Experiment 2, the payoff from one choice was randomly selected and added to the base fee of £4.00. After adding the base fee, the fee ranged from £0.00 to £8.00, with a mean of £4.08.

RESULTS

Analyses were confined to trials where participants sampled at least two alternatives and chose an alternative they had

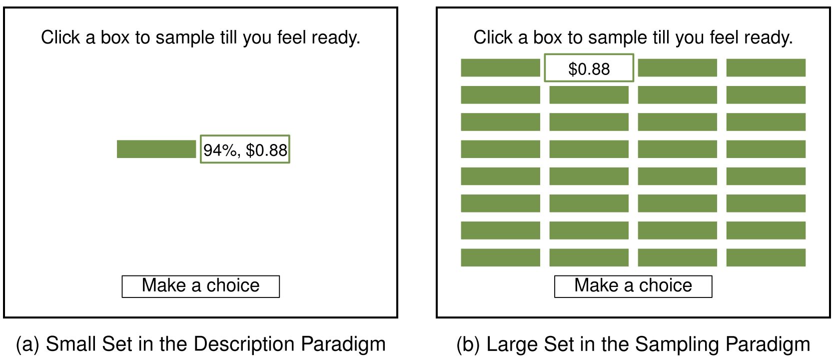

Figure 2. Example screenshots showing how samples are presented: (a) small set in the description paradigm and (b) large set in the sample paradigm. Font size is enlarged for illustration purposes

DOI: 10.1002/bdm

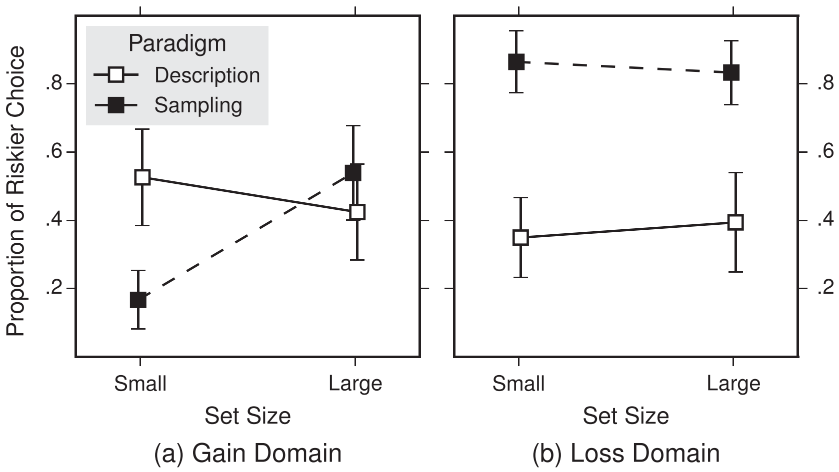

Figure 3. Proportion of choices for riskier alternatives in Experiment 1: (a) gain domain and (b) loss domain. Error bars are 95% confidence intervals

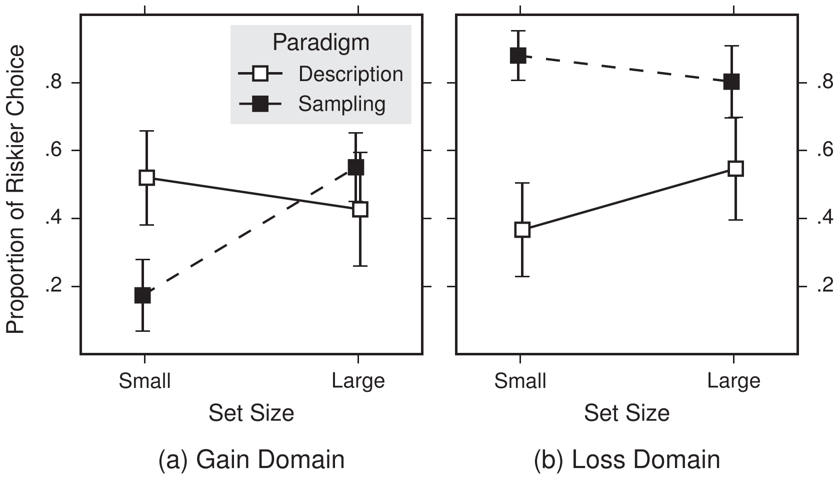

Figure 4. Proportion of choices for riskier alternatives in Experiment 2: (a) gain domain and (b) loss domain. Error bars are 95% confidence intervals

sampled at least once.[1] This represented 686 out of 786 choices (=131 participants × 6 choices) in Experiment 1, and 555 out of 606 choices (=101 participants × 6 trials) in Experiment 2.

Risk-taking

Figures 3 and 4 present the proportion of choices for riskier alternatives for each experiment. Both experiments replicated the description–experience gap in both gain and loss domains for the small set. In the gain domain, riskier alternatives were more frequently chosen in the description paradigm than in the sampling paradigm. Alternatively, in the loss domain, riskier alternatives were more frequently chosen in the sampling paradigm than in the description paradigm.

For the large set, the description–experience gap disappeared in the gain domain, but not in the loss domain. Mixed-effect logistic regressions—with by-participant intercepts and slopes as random factors—indicate a significant three-way interaction between the domain, the paradigm, and the set size (Experiment 1: $\chi^2(1) = 6.71$, $p < .01$; Experiment 2: $\chi^2(1) = 11.70$, $p < .001$). For the gain domain, the effect of paradigm differs between the small and large sets (Experiment 1: $\chi^2(1) = 13.22$, $p < .001$; Experiment 2: $\chi^2(1) = 11.84$, $p < .001$). Specifically, with a growth in set size, riskier alternatives become more frequently chosen in the sampling paradigm (Experiment 1: $\chi^2(1) = 19.24$, $p < .001$; Experiment 2: $\chi^2(1) = 18.90$, $p < .001$) but not in the description paradigm (*ps* > .39). This increased risk-taking eliminates the description–experience gap for the large set.

In the loss domain, riskier alternatives in the small set were chosen more than 80 % of the time in the sampling paradigm. We expected riskier alternatives to be chosen more often in the large set, but the proportion of choices for riskier alternatives was already near ceiling for the small set. As a result, a growth in set size does not significantly influence choices for riskier alternatives (*ps* > .07). Here, the description–experience gap persists, providing reassuring support for prior generalizations from the two-alternative

[1]The main results and our conclusion are not affected when we conducted the identical analysis with further confinement to trials where at least one riskier and one safer alternative were sampled.

DOI: 10.1002/bdm

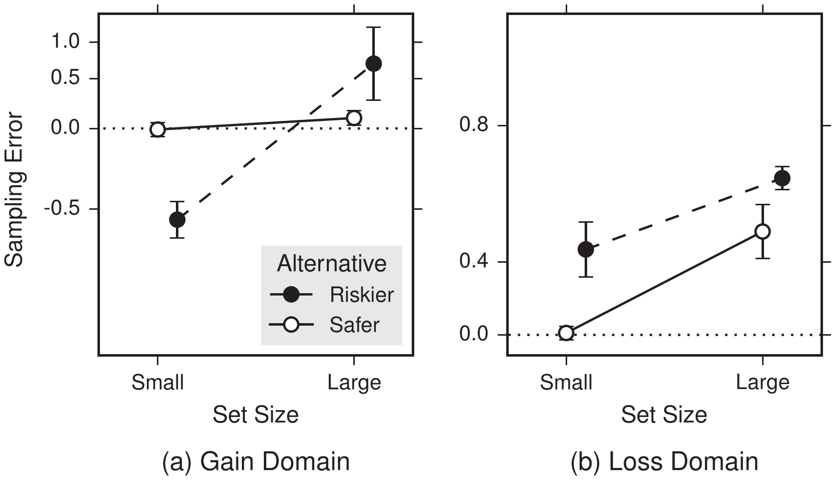

Figure 5. Sampling error in Experiment 1: (a) gain domain and (b) loss domain. For illustration purposes, the sampling error is mean-averaged across trials for each participant before being averaged across participants. Error bars are 95% confidence intervals

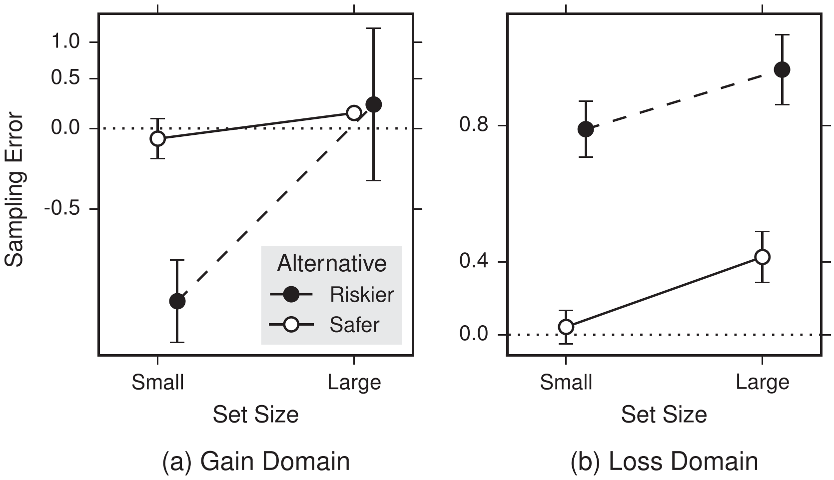

Figure 6. Sampling error in Experiment 2: (a) gain domain and (b) loss domain. Otherwise, the figure is presented as in Figure 5

environments in the loss domain (e.g., Hertwig & Erev, 2009; Yechiam et al., 2005).

The results for the sampling paradigm are also in line with previous studies. Ert and Erev (2007), in particular, suggested that the effect of set size on risk-taking depends on whether or not rare payoffs are attractive. In the gain domain where we observe the increased risk-taking, rare payoffs are attractive, but in the loss domain where we do not observe the increased risk-taking, rare payoffs are not attractive. Also, we note that the non-significant results in the description paradigm do not necessarily indicate zero effect: a growth in set size might have another effect on choices than tested in the present study (see, for example, Benartzi & Thaler, 2001; Iyengar & Lepper, 2000; Noguchi & Hills, 2015).

Sampling error

The increased risk-taking in the gain domain may be due to sampling error associated with a growth in set size. To investigate this possibility, the sampling error was calculated for each alternative, and we identified the largest sampling error separately for riskier and safer alternatives in each trial and for each participant. Here, positive error indicates that an alternative has produced samples with a higher mean than its expected payoff; negative error indicates the opposite.

Sampling error is displayed in Figures 5 and 6.[2] In all cases, the sampling error increases with a growth in set size, and the increment is largest for riskier alternatives in the gain domain. Mixed-effect linear regressions reveal a significant interaction effect, indicating that the difference between riskier and safer alternatives depends on the interaction between the set size and the domain (Experiment 1: $\chi^2(1) = 50.12$, $p < .001$; Experiment 2: $\chi^2(1) = 15.22$, $p < .001$). For the gain domain, the set size has different effects on riskier and safer alternatives (Experiment 1: $\chi^2(1) = 37.12$, $p < .001$;

[2]The distribution of sampling errors was positively skewed in the gain domain and negatively skewed in the loss domain. Therefore, in fitting statistical models, we rescaled the sampling error by log-transforming their values after correcting for the different lower and upper bounds for the gain and loss domains, respectively. For the gain domain, we subtracted the smallest possible sampling error, − 1.1, and log-transformed the resulting value. For the loss domain where the largest possible sampling error is 1.1, we subtracted 1.1 from the sampling error and negated to make the distribution positively skewed, log-transformed and negated again. Spacing between the ticks along the vertical axis in Figures 5 and 6 reflects this transformation, but we labeled the ticks with the sampling error before the transformation.

DOI: 10.1002/bdm

Experiment 2: $\chi^2(1) = 12.59$, $p < .001$). While the sampling error increases for both riskier and safer alternatives, the increment is larger for a riskier alternative (Experiment 1: $\chi^2(1) = 53.71$, $p < .001$; Experiment 2: $\chi^2(1) = 21.24$, $p < .001$) than for a safer alternative (Experiment 1: $\chi^2(1) = 4.47$, $p = .03$; Experiment 2: $\chi^2(1) = 5.57$, $p = .02$).

For the loss domain, the set size does not have significantly different effects on riskier and safer alternatives (Experiment 1: $\chi^2(1) = 2.59$, $p = .11$; Experiment 2: $\chi^2(1) = 0.27$, $p = .61$). Here, the sampling error for both riskier and safer alternatives is significantly larger in the large set than in the small set (Experiment 1: $\chi^2(1) = 42.55$, $p < .001$; Experiment 2: $\chi^2(1) = 11.26$, $p < .001$).

Therefore, in the gain domain, riskier alternatives appear better than safer alternatives with a growth in set size. This relative attractiveness of riskier alternatives explains the increase in risk-taking: with a growth in set size in the gain domain, the sampling error increases most for riskier alternatives. To confirm, we take a difference between the sampling errors for a riskier and a safer alternative for each trial. This difference in the sampling error is a measure of how attractive riskier alternatives appeared in relation to safer alternatives. Then we entered the difference into a mixed-effect logistic regression to predict choices of riskier alternatives.

The effects of the sampling error on choices do not significantly differ between the gain and loss domains (Experiment 1: $\chi^2(1) = 3.32$, $p = .07$; Experiment 2: $\chi^2(1) = 2.46$, $p = .12$) or between the small and large sets (Experiment 1: $\chi^2(1) = 2.28$, $p = .13$; Experiment 2: $\chi^2(1) = 0.20$, $p = .65$). Overall, the difference in the sampling error is a significant predictor of choices (Experiment 1: $\chi^2(1) = 44.79$, $p < .001$; Experiment 2: $\chi^2(1) = 58.35$, $p < .001$). Therefore, when the sampling error is larger for riskier alternatives than for safer alternatives, riskier alternatives are more likely to be chosen.

Thus, the sampling error significantly increases with a growth in set size in both the gain and loss domains. In the gain domain, however, the increment is more severe for riskier alternatives and is associated with the increased risk-taking.

Sampling behavior

Additionally, we have examined the number of alternatives participants sampled and the number of samples participants drew from each alternative. In a large set, participants did not always sample all the 32 alternatives. On average, participants sampled 18.04 alternatives (95% confidence interval [CI] [15.53, 20.83]) in Experiment 1 and 13.96 alternatives (95% CI [11.47, 16.83]) in Experiment 2. Also, the number of samples per alternative tends to decrease with a growth in set size: for Experiment 1, the mean number of samples is 2.80 (95% CI [2.10, 3.73]) in the small set and 1.59 (95% CI [1.38, 1.83]) in the large set, and for Experiment 2, it is 2.01 (95% CI [1.51, 2.68]) in the small set and 1.20 (95% CI [1.10, 1.31]) in the large set. These results replicate previous findings by Hills et al. (2013).

DISCUSSION OF EXPERIMENTAL RESULTS

As hypothesized, a growth in set size increases risk-taking in the gain domain in the sampling paradigm but not in the description paradigm. As a consequence, the description–experience gap is eliminated in the gain domain. In the loss domain, on the other hand, choices appear uninfluenced by set size.

Investigation of the sampling paradigm revealed that these effects are predicted by changes in the sampling error, which show the largest relative changes for riskier alternatives in the gain domain. We here note that as sampling error predicts choices in all the conditions, choices in the large set are not likely to be randomly made. This observation is also consistent with the evidence reported in Hills et al. (2013), which found that the relative attractiveness of chosen alternatives increases with a growth in set size: choices become more predictable.

Because the increased risk-taking is driven by the sampling error, our results leave open the possibility that the increased risk-taking can be alleviated if individuals sample more per alternative. Previous studies have demonstrated a close relationship between the sampling error and the number of samples (e.g., Hertwig & Pleskac, 2010; Ungemach et al., 2009), and therefore, the effects of set size on the sampling error may be attenuated as more samples are drawn from an alternative. In the next section, we use a simulation to ask whether or not a particularly motivated sampler might escape the increased risk-taking in the gain domain by simply sampling more.

SIMULATION

The purpose of this simulation is to explore the relationship between the number of samples per alternative and the sampling error in the gain domain. To this end, we simulated alternatives generated with the same procedures as in Experiments 1 and 2 but allow each alternative to be sampled a fixed number of times. We also investigated these results over a range of set size.

Method

Pooled across Experiments 1 and 2, the mean number of samples is 3.94 in the small set and 1.74 in the large set. Thus, we simulated trials where all the alternatives are sampled twice, four times and also as many time as 10 and 20. Then as in the results section in the preceding text, the sampling error is computed for each alternative and the largest sampling error within a trial is identified for riskier and safer alternatives separately.

We simulated 10^5 trials for each number of samples for each set size. Set size in the simulation ranged between 2 and 32 alternatives with a step size of 2. As in the experiments, half of alternatives in a set are riskier and the other half are safer in each simulated trial.

DOI: 10.1002/bdm

Results and discussion

The simulation results are summarized in Figures 7 and 8. The four lines in each panel monotonically increase with a growth in set size, indicating that the sampling error increases regardless of the number of samples.

Consistent with the results from the experiments, riskier alternatives show a more drastic increase in the sampling error than safer alternatives. In particular, the mean sampling error for riskier alternatives is negative when set size is small, indicating that riskier alternatives appear worse on average than their expected payoffs. The sampling error, however, becomes positive with a growth in set size, suggesting that riskier alternatives start appearing increasingly better than their expected payoffs. Specifically, once six or more alternatives are sampled, riskier alternatives appear better on average than its expected payoff, regardless of the number of samples taken per alternative.

The simulation results also show that the sampling error for riskier alternatives has a non-monotonic relationship with the number of samples. In a set with two alternatives, the sampling error comes closer to zero with more samples. Also, in a set with 16 or more alternatives, the sampling error comes closer to zero with more samples. In a set with more than 2 but less than 16 alternatives, however, the sampling error is not necessarily reduced with more samples. In fact, the sampling error can be increased with more samples. In a set of six alternatives in Experiment 1, for example, the mean sampling error increases from 0.26 to 0.40 as the mean number of samples increases from 2 to 20.

GENERAL DISCUSSION

Psychological experiments often simplify the complexity of the environments in which we live. This simplification is necessary for researchers to isolate and manipulate variables of interest while holding other variables constant. However, this simplification can lead to the neglect of variables most likely to influence behavior outside the laboratory. In the present study, we demonstrate that one such variable, set size, which is widely recognized to be increasing in our modern world (Schwartz, 2004), has a substantial and unavoidable impact on decision making.

Specifically, using three different methods (a simple mathematical model, two experiments, and a simulation), we show that a growth in set size increases risk-taking in

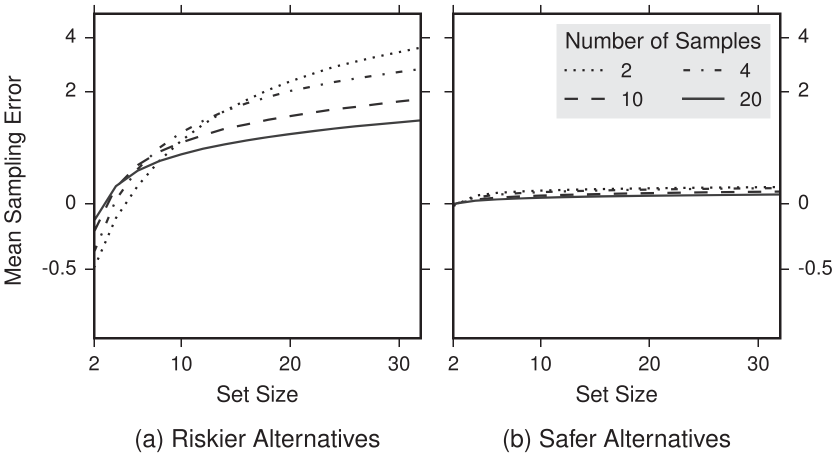

Figure 7. Simulation result for Experiment 1: (a) riskier alternatives and (b) safer alternatives. The horizontal axis represents the number of alternatives, half of which are riskier and the other half are safer. Therefore, if the number of alternatives is 10, five of them are riskier and the other five are safer. Sampling error is mean-averaged across the 10^5 simulated trials

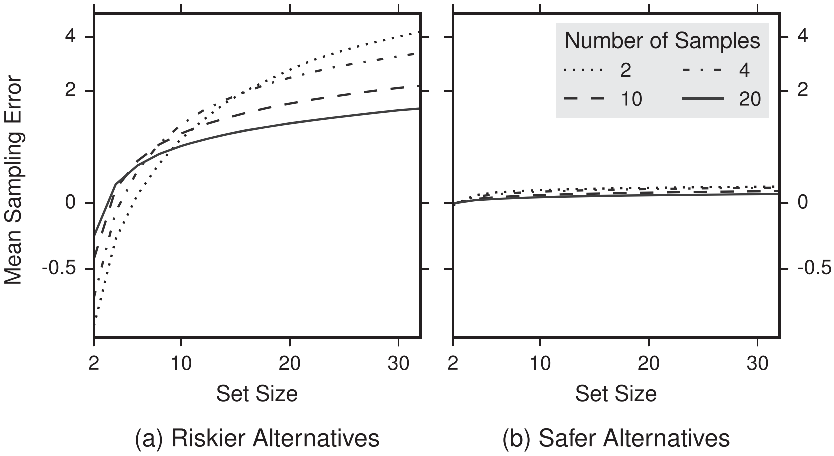

Figure 8. Simulation result for Experiment 2: (a) riskier alternatives and (b) safer alternatives. The horizontal axis represents the number of alternatives, half of which are riskier and the other half are safer

DOI: 10.1002/bdm

the gain domain in the sampling paradigm. This increased risk-taking eliminates the description–experience gap. Notably, however, we also show that risk-taking is not increased for the loss domain with a growth in set size, as risk-taking is already high in small sets. In both domains, however, the results are explained by sampling error.

In addition, our simulation results demonstrate that the sampling error is not necessarily reduced with more samples and that even fairly large numbers of samples (20 samples per alternative for 30 alternatives) are still subject to sampling error sufficient to increase risk-taking. Moreover, with certain set sizes, sampling error can increase as each alternative is sampled more. These findings illustrate the difficulty in preventing risk-taking from being increased with a growth in set size. For such prevention, detailed knowledge of choice environments, such as set size and probabilities of payoffs, may be required. Without such knowledge, the increase in risk-taking may be inevitable.

Our findings may appear at odds with previous studies, which report that individuals learn to avoid riskier alternatives in a large set (Ert & Erev, 2007; Grosskopf, Erev, & Yechiam, 2006). In the previous studies, however, participants were presented with information about the payoffs participants earned and could have earned by choosing other alternatives, immediately after making each choice. This information can help participants learn that payoffs observed before a choice are not necessarily representative payoffs (Ert & Erev, 2007) and indeed may help participants to categorize risky and safe alternatives.

To illustrate the distinction between the present and previous findings, consider searching among multiple financial portfolios before choosing one in which to invest your pension. With a growth in set size, it becomes more likely that any particular portfolio has had a lucky streak in the recent past. The present study indicates that with a growth in set size, individuals become more likely to choose the portfolios with the lucky streaks. The previous studies indicate, in contrast, if individuals make repeated choices between portfolios and are provided with information immediately after each choice, individuals may be able to learn that portfolios with recent lucky streaks do not necessarily perform well in the future. Consequently, individuals may learn to avoid portfolios with lucky streaks. This learning may require 100 or more repeated choices (Grosskopf et al., 2006). Thus, the present study highlights that individual investors, who do not often make choices between portfolios repeatedly, may take higher risks when the set size is large.

The increased risk-taking with a growth in set size is potentially a growing concern. The world wide Web now provides consumer ratings and personal experiences for countless alternatives, ranging from treatments for depression to suggestions about how to get your partner back. Immediate feedback and information on forgone outcomes are often not available, and as a result, it may be becoming increasingly likely for alternatives with rare successes to be chosen. The breadth of potential choice domains suggests that the impact of observing rare success may be large and potentially dangerous. Thus, the effects of set size are extremely important in understanding how individuals make choices in complex many-alternative environments, and we believe it is a critical direction for future research.

APPENDIX A: LIST OF EXAMPLE ALTERNATIVES

Table A1. Example alternatives generated for one trial in the gain domain in Experiment 1

Alternative	Probability of non-zero payoff	Amount of non-zero payoff	Probability of nothing	Expected payoff	Payoff variance
Safer	.92	0.62	.08	0.57	0.33
Safer	.95	0.60	.05	0.57	0.33
Safer	.94	0.60	.06	0.56	0.32
Safer	.94	0.60	.05	0.56	0.32
Safer	.88	0.64	.12	0.56	0.32
Safer	.84	0.68	.16	0.57	0.34
Safer	.96	0.59	.04	0.57	0.32
Safer	.86	0.65	.14	0.56	0.32
Safer	.85	0.67	.15	0.57	0.33
Safer	.91	0.62	.09	0.56	0.32
Safer	.83	0.68	.17	0.56	0.33
Safer	.82	0.61	.18	0.50	0.26
Safer	.89	0.64	.11	0.57	0.33
Safer	.89	0.64	.11	0.57	0.33
Safer	.96	0.59	.04	0.57	0.32
Safer	.89	0.64	.11	0.57	0.33
Riskier	.06	8.16	.94	0.49	59.07
Riskier	.01	44.19	.99	0.44	1914.09
Riskier	.11	4.76	.89	0.52	18.22
Riskier	.08	6.68	.92	0.53	38.05
Riskier	.05	10.40	.95	0.52	97.88
Riskier	.04	14.31	.96	0.57	189.05
Riskier	.16	3.10	.84	0.50	7.03

(*Continues*)

DOI: 10.1002/bdm

Table A1. (Continued)

Alternative	Probability of non-zero payoff	Amount of non-zero payoff	Probability of nothing	Expected payoff	Payoff variance
Riskier	.08	6.11	.92	0.49	31.84
Riskier	.18	2.85	.82	0.51	5.72
Riskier	.16	3.16	.84	0.51	7.30
Riskier	.08	6.21	.92	0.50	32.89
Riskier	.08	6.57	.92	0.53	36.81
Riskier	.02	21.30	.98	0.43	435.91
Riskier	.04	12.66	.96	0.51	147.97
Riskier	.09	5.84	.91	0.53	28.52
Riskier	.19	2.65	.81	0.50	4.86

In this particular trial, all the safer alternatives are assigned with expected payoff of 0.56, and all the riskier alternatives are assigned with expected payoff of 0.51. Please note that expected payoff listed varies due to rounding.

APPENDIX B: INSTRUCTION FOR EXPERIMENT 2

In the following task, you will be presented with sets of boxes. Each box is a money machine that pays off a fixed non-zero amount some percentage (%) of time.

You can sample these boxes as many times as you want to determine which box you want to choose to receive your payoff.

During the sampling phase, you will not receive any payoffs. When you decide to make a final choice, the boxes you have not sampled disappear from the screen and you will be asked to choose one of the boxes you have sampled. You will get one draw from this box "for real" as your payoff for that set of boxes.

You will be asked to make five choices in this task, and the earnings from your final choices are recorded throughout the task.

After you make the fifth choice, one of your choices is randomly selected, and you will be paid the outcome of the selected choice. You will learn how much you will be paid at the end of the task.

When you click on "Start" below, you will be first taken to a practice choice. Your earning from this practice choice is not counted toward your payment or included among the five choices in this task.

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.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*, 79–98.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., . . . Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*, 15–47.

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

Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social sampling explains apparent biases in judgments of social environments. *Psychological Science*, *23*, 1515–1523.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review*, *118*, 523–551.

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

Grosskopf, B., Erev, I., & Yechiam, E. (2006). Foregone with the wind: Indirect payoff information and its implications for choice. *International Journal of Game Theory*, *34*, 285–302.

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*, 517–23.

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

Hilbig, B. E., & Gloeckner, A. (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, *138*, 390–396.

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

Hills, T. T., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review*, *20*, 1023–1031.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 143–150.

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

Lajarraga, Y., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*, 334–342.

Noguchi, T., & Hills, T. T. (2015). Description–experience gap in choice deferral. Manuscript submitted for publication.

Schwartz, B. (2004). *The paradox of choice: Why more is less*. New York: Harper Perennial.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

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.

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

DOI: 10.1002/bdm

Authors' biographies:

Takao Noguchi is a research associate at the University College London. His research interests include the psychological process of judgment and decision making.

Thomas T. Hills is a professor of psychology at the University of Warwick. His research focuses on using statistical and model-based approaches to quantify the interface between mind and environment.

Authors' addresses:

Takao Noguchi, Department of Experimental Psychology, University College London, UK.

Thomas T. Hills, Department of Psychology, University of Warwick, UK.