Context effects and risk amplification: Why more is risky

Takao Noguchi (t.noguchi@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry, CV4 7AL, UK

Thomas T. Hills (thomhills@gmail.com)
Department of Psychology, University of Warwick
Coventry, CV4 7AL, UK

Abstract

Research on risky choice has been dominantly based on studies of choice between two alternatives, with the findings 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 experienced (the experience paradigm): Individuals appear to make decisions as if they over-weight small probabilities in the description paradigm, but under-weight the same probabilities in the experience paradigm. Here, we show that the under-weighting in the experience paradigm is sensitive to the choice set size in the gain domain. Two experiments show that as set sizes increase, choices systematically favour risky alternatives in the experience paradigm. Using simulations of three choice models, we further demonstrate that this risk-amplification is independent of choice and search strategies and is predicted by the statistical structure of pay-offs. The results suggest caution in generalising findings from two-choice environments to many-choice environments and further indicate a robust and systematic problem with increasing choice set sizes.

Keywords: context effect; description-experience gap; decision from experience; search-amplified risk; too much choice

Over the past decade, research with two-choice environments has led to the claim that in the experience paradigm, where decisions are made after experiencing a series of sample pay-offs (e.g., \$0, \$0, \$0, \$9, and \$0), individuals make a decision 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 the description paradigm (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), where pay-offs 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-choice environments (e.g., Ungemach, Chater, & Stewart, 2009; Lajarraga, Hertwig, & Gonzalez, 2012; Gonzalez & Dutt, 2011; Newell & Rakow, 2007; Erev et al., 2010; Gottlieb, Weiss, & Chapman, 2007; Abdellaoui, L'Haridon, & Paraschiv, 2011; Hilbig & Gloeckner, 2011).

The under-weighting of small probabilities in the experience paradigm has been used to explain a variety of phenomenon related to decision making outside the laboratory, including those involving perceived terrorist threats (Yechiam, Barron, & Erev, 2005) and the recent financial crisis (Hertwig & Erev, 2009). However, because the number of decision alternative is often more than two outside the laboratory, those generalisations implicitly assume the principle of independence from irrelevant alternatives (IIA; Arrow, 1963): the order of preferences for alternatives should not be changed by the inclusion of non-preferred alternatives in a choice set.

Violations of the principle of IIA, however, have been reported in various areas of decision making (e.g., Huber, Payne, & Puto, 1982; Simonson, 1989; Tversky, 1972), and we examined whether the principle of IIA holds in differing set sizes. If a risky alternative is preferred between one risky and one safe alternative, the principle of IIA implies that a risky alternative is still preferred between many risky and safe alternatives. Here, a risky alternative has a small probability of a large pay-off, and a safe alternative has a large probability of a small pay-off.

Prior empirical evidence suggests that decisions can change systematically with set sizes. Ert and Erev (2007), for instance, report that when individuals can observe foregone pay-offs from the alternatives they did not choose, a larger set tends to lead the individuals to choose the alternative that, most recently, delivered the largest pay-off. Further, Hills, Noguchi, and Gibbert (2013) found that larger and more diverse choice sets lead individuals to sample more alternatives but fewer times per alternative and subsequently to choose alternatives which delivered a larger sample pay-off. As large pay-offs are often associated with risky alternatives in decision research, these studies imply that increasing set sizes facilitates choice for a risky alternative.

Based on the above work, we hypothesized that in the experience paradigm, safe alternatives would be more often chosen when set sizes are small, but risky alternatives would be more likely to be chosen with large set sizes. In addition to this increase in risky choice violating the principle of IIA, we further predict that the effect of set size will diminish the description-experience gap. As discussed above, the description-experience gap is partly explained by under-weighting of small probabilities in the experience paradigm, which predicts that safe alternatives are more frequently chosen than risky alternatives. If risky alternatives become more frequently chosen as set-sizes grow, the description-experience gap will become smaller and eventually diminish.

In Experiments 1 and 2, we manipulated the set sizes in both description and experience paradigms, for both gain and loss domains, and for conditions in which the expected pay-off of an alternative was positively or negatively associated with the probability of pay-off. We then examined the robustness of these results and their potential cause by simulating decision models for various set sizes and information search strategies. Our principle goal was to lay a foundation for identifying environments to which findings from two-choice environments may be best generalised.

Experiments 1 and 2

We report how we determined the number of participants, all data exclusions (if any), all manipulations, and all measures in the experiments. Both experiments employed a 2 (between-participant, set size: small or large) $\times$ 2 (between-participant, paradigm: description or experience) $\times$ 2 (within-participant, domain: gain or loss) design. Experiment 1 was conducted online and the pay-off was in American dollars. Experiment 2 was carried out in a laboratory and the pay-off was in British pounds. The two experiments differed in the relationship between risk and expected pay-off to ensure that the results do not depend on particular structure of pay-offs. Additionally, the experiments differed in how the participation fee was calculated (explained below).

Participants

In Experiment 1, 131 participants (73 males, 56 females and 2 unspecified) were recruited through Mechanical Turk (`http://www.mturk.com`). Their age ranged from 18 to 69 with a mean of 30.63. In Experiment 2, 101 students (57 males and 44 females) were recruited through the participant panel at the University of Warwick. Their age ranged from 18 to 52 with a mean of 22.7. We decided in advance of collecting the data to test exactly 100 participants for both experiments, but over-recruited due to technical reasons.

Apparatus

The alternatives were independently and randomly generated for each trial for each participant. Half of the alternatives within a trial (1 alternative in the small set size, and 16 alternatives in the large set size) were *safe* alternatives. The probability of pay-off for each safe alternative was a random draw from a uniform distribution between 0.8 and 1.0. The other half of the alternatives were *risky*, with probability of pay-off drawn from a uniform distribution between 0.0 and 0.2. Each alternative was presented as a box on a screen, and a choice set was presented to the participant as an array of boxes.

For each trial, we draw a random number 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. This number is used as an expected pay-off for the safe alternatives. Thus within a trial, all the safe alternatives had the same expected pay-off. In Experiment 1 and 2 the expected pay-off for the safe alternatives was multiplied by 0.9 and 1.1, respectively, to derive the expected pay-off for the risky alternatives. For each alternative, the expected pay-off was divided by the probability of pay-off and rounded to the nearest two decimals to derive the non-zero pay-off amount that would appear on the screen.

Procedure

Participants were instructed that their payments would depend on their choices during the experiment. The two experiments asked participants to make six choices in total, three involving gains and three involving losses. The gain and loss trials were interleaved and presented in a random order. Each trial displayed 2 alternatives (small set size) or 32 alternatives (large set size).

At each trial, participants were asked to sample from the alternatives as many times as they wanted and then to choose one of the alternatives. Every time an alternative was sampled, information about the alternative was presented for 500 ms. In the description paradigm, the information displayed the probability and pay-off amount (e.g., 10%, \$9.00). In the experience paradigm, the information presented was a random sample from the pay-off distribution associated with that alternative. For example, when an alternative with a 10% chance of \$9.00 was sampled in the experience paradigm, 1 in 10 samples displayed \$9.00, otherwise \$0.00.

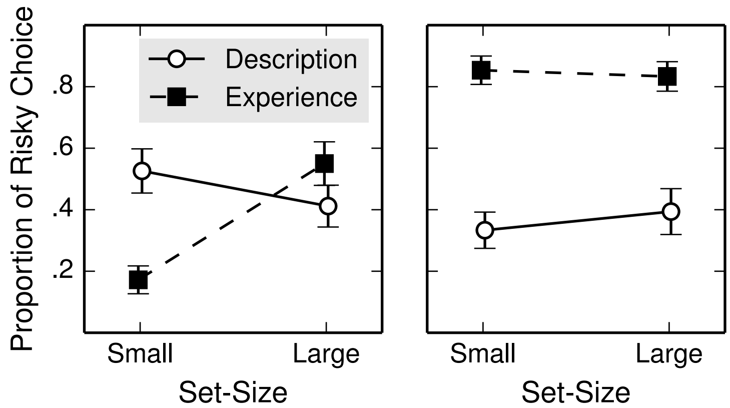
Participants did not learn about the pay-offs from their final choices until the end of the experiment, when the participation fee was calculated. In Experiment 1, the pay-offs from the six choices were summed and added to the base fee of \$1.00. The fee ranged from \$0.00 to \$14.57, with a mean of \$1.80. In Experiment 2, the pay-off from one choice was randomly selected and added to the base fee of £4.00. Participant could receive up to £8.00, and a mean fee was £4.08.

Results and Discussion

Analyses were confined to trials where participants chose an alternative they had sampled at least once. This represented 703 choices out of 786 (= 131 participants $\times$ 6 choices) in Experiment 1, and 562 choices out of 606 (= 101 participants $\times$ 6 trials) in Experiment 2.

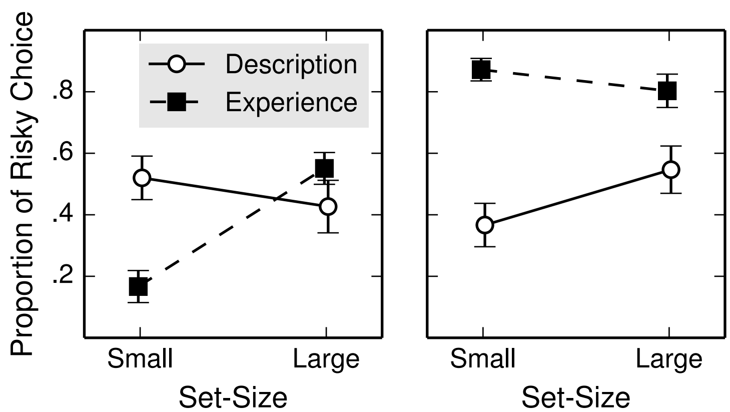
Increasing the set size eliminates the description-experience gap. Figures 1 and 2 show the results of Experiments 1 and 2, respectively. Both replicate the description-experience gap for the small set size in the gain and loss domains. In the gain domain, a risky alternative was more frequently selected in the description paradigm than in the experience paradigm. Alternatively in the loss domain, a risky alternative was more frequently chosen in the experience paradigm than in the description paradigm. These results are consistent with the over-weighting of small probabilities in the description paradigm and the under-weighting of small probabilities in the experience paradigm, results routinely noted in the literature (e.g., Hertwig & Erev, 2009; Ungemach et al., 2009; Hilbig & Gloeckner, 2011).

Statistical significance was examined with mixed-effect logistic regressions with maximal random effects (Barr, Levy, Scheepers, & Tily, 2013) to predict the choices of the risky alternative. Model fits indicate a significant three-way in-



(a) Gain Domain (b) Loss Domain

Figure 1: Proportion of risky choice in Experiment 1. Error bars are standard error.



(a) Gain Domain (b) Loss Domain

Figure 2: Proportion of risky choice in Experiment 2. Error bars are standard error.

teraction (Experiment 1: $\chi^2(1) = 7.01$, $p = .008$; Experiment 2: $\chi^2(1) = 11.83$, $p < .001$). For the gain domain, the effect of paradigm depends on set sizes (Experiment 1: $\chi^2(1) = 13.85$, $p < .001$; Experiment 2: $\chi^2(1) = 12.58$, $p < .001$). When the set size is increased, a risky alternative is significantly more frequently chosen in the experience paradigm (Experiment 1: $\chi^2(1) = 18.46$, $p < .001$; Experiment 2: $\chi^2(1) = 20.64$, $p < .001$) but not in the description paradigm ($ps > .344$). This amplified risk taking eliminates the description-experience gap in the large set size.

For the loss domain, the set size did not significantly influence the proportion of risky choice ($ps > .060$). Here, the description-experience gap persisted, providing reassuring support for prior generalisations from the two-choice environments in the loss domain (e.g., Yechiam et al., 2005; Hertwig & Erev, 2009).

Note that because the associations between the pay-offs and probability of pay-off differ between the two experiments, these results are not due to participants in one paradigm simply choosing the alternatives with the higher expected pay-off.

Given the sampled pay-offs, a better alternative is systematically chosen in the large set size. To show that the elimination of the description-experience gap in the gain domain is not due to a greater tendency for random choice in the large set size, we computed the mean sample pay-offs participants observed for each alternative. Using this mean pay-off, the sampled alternatives were ranked in a descending order within each trial. Ranks were then normalised as follows:

$$\text{relative rank} = \frac{\text{rank} - 1}{n - 1},$$

where n is the number of unique alternatives sampled. Random choice is indicated by a mean relative rank of the chosen alternative close to .50. The mean relative rank of the chosen alternative in the experience paradigm for the large set size was .85 (SE = 0.02) in Experiment 1 and .90 (SE = 0.02) in Experiment 2. These were both significantly higher than .50 (using a mixed-effect linear regression with maximal random effects on the logit-transformed ranks, Experiment 1: $t(325) = 10.40$, $p < .001$; Experiment 2: $t(260) = 14.79$, $p < .001$), indicating that participants consistently chose higher ranked alternatives among those they sampled and thus were *not* choosing randomly.

Simulating Decision Models over Increasing Set Sizes

Experiments 1 and 2 demonstrated a set size dependent violation of the principle of independence from irrelevant alternatives and suggest that as the set size increases, decisions tend to favour risky alternatives in the gain domain. This risk-amplification, however, could possibly be alleviated by employing different decision or information search strategies. Thus to understand whether and how decision and information search strategies impact on the risk-amplification, we examined the influence of set size using three of the best performing models for explaining choices in the experience paradigm reported in Hau, Pleskac, Kiefer, and Hertwig (2008).

Method

We simulated choices in the environments which reflect those in Experiments 1 and 2, with the following exception: the expected pay-offs for the risky and safe alternatives were made equal within each trial. This eliminates any confound associated with the relationship between the expected pay-off and the alternative categories (i.e., risky and safe). Simulated participants followed one of the three models in making a decision: the maximax model, the two-stage model of cumulative prospect theory (the two-stage model henceforth), or the natural mean model. Participants with the maximax model chose the alternative with the largest experienced pay-off.

Participants who followed the two-stage model first weighted the frequency of pay-offs and transformed pay-off amounts into a subjective value to derive the perceived expectation of utility, E:

$$\begin{aligned} E &= w(\text{experienced frequency of pay-off}) \\ &\quad \times v(\text{amount of pay-off}), \end{aligned}$$

where

$$w(f) = \frac{f^\gamma}{(f^\gamma + (1-f)^\gamma)^{1/\gamma}}$$

and

$$v(x) = \operatorname{sign}(x)\,|x|^\alpha.$$

The participants then chose the alternative with the largest perceived expectation of utility. We used parameter values reported in Hau et al. (2008): $\alpha = 0.94$ and $\gamma = 0.99$ for the gain domain; $\alpha = 0.86$ and $\gamma = 0.93$ for the loss domain.

Lastly, participants using the natural mean chose the alternative with the highest mean of sample pay-offs. This natural mean model is a special case of the two-stage model, where $\alpha = 1$ and $\gamma = 1$.

For each of the choice models and for a variety of set sizes (from 2 to 32) we simulated 10^4 participant choices with a fixed number of samples per alternative.

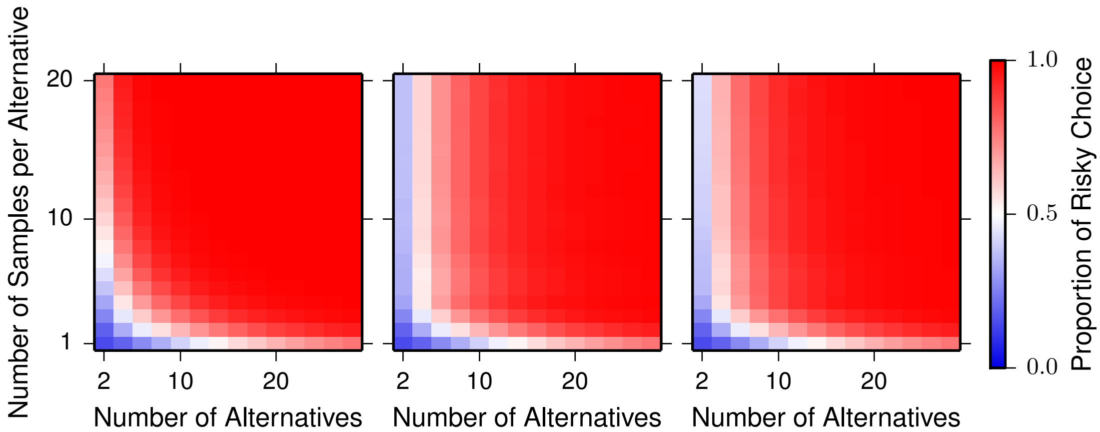
Results and Discussion

The simulation results are summarised in Figures 3 and 4 for the gain and loss domains, respectively. In both domains the three models make similar predictions over a large area of the simulation space. The most striking result is that, in the gain domain, the simulated participants chose a safe alternative only when the set size and the number of samples are relatively small. When the set size is larger than two, risky alternatives were chosen almost exclusively. The only exception is when the number of samples is near one per alternative and only when the number of alternatives is equal to or smaller than 12. When more than 12 alternatives are available, all three models predict that a risky alternative is more frequently chosen, even when the number of samples is as large as 20 samples per alternative. This is a clear and dramatic reversal of the inference based on two-choice experiments, where the experience paradigm leads individuals to under-weight small probabilities and to choose a safe alternative (Hertwig et al., 2004; Hertwig & Erev, 2009).

The simulation indicates that the risk-amplification for the gain domain is independent of the decision and information search strategies we examined. Further, the simulation also provides insight into the reason for the risk-amplification. To illustrate, suppose an individual samples each alternative once but the set sizes increase from 2 to 20 — with half risky and half safe. Further assume that a risky alternative pays \$9.00 with probability .10 and a safe alternative pays \$0.90 with probability 1. The frequency of pay-off from the risky alternative is determined by the binomial distribution. With two alternatives in a choice set, there is a .10 probability that a sample from the risky alternative is \$9.00, leading to the infrequent choice of a risky alternative. With 20 alternatives in a choice set, however, there is a $1-(1-.10)^{20} = .88$ probability that a sample from at least one risky alternative is \$9.00, leading to the more frequent choice of the risky alternative. Similar results follow for larger sample sizes and varying choice strategies.

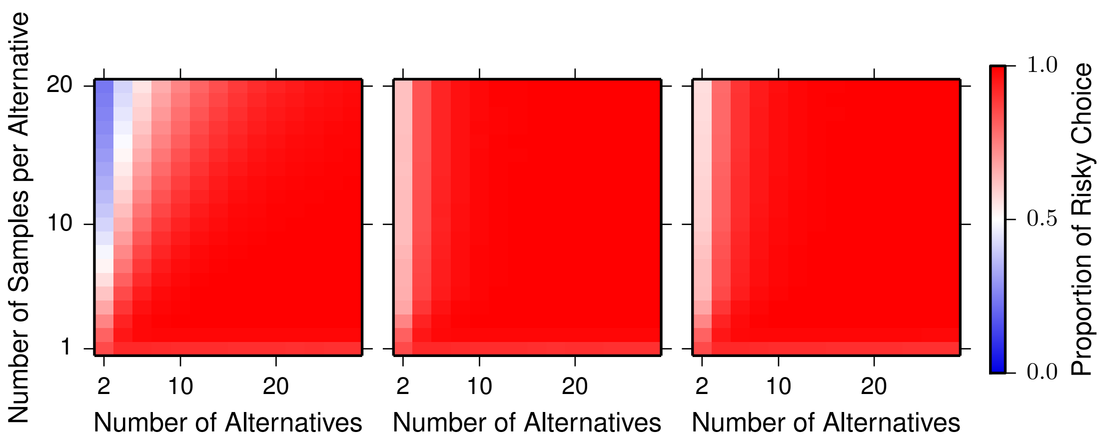
For the loss domain, the same statistical structure of pay-offs explains the smaller influence of the set size. Again, we assume that a risky alternative pays $-$\$9.00 with probability .10 and that a safe alternative pays $-$\$0.90 with probability 1. When an individual samples once per alternative, a risky alternative delivers \$0.00 with probability $1 - .10 = .90$ with two alternatives in a choice set, leading to the frequent choice of the risky alternative. With 20 alternatives in a choice set, at least one risky alternative delivers \$0.00 with probability greater than $1 - .10^{20} \approx 1$, leading to the choice of the risky alternative.

In both domains, the results follow straightforwardly from the binomial distribution of pay-offs associated with the risky



(a) Maximax (b) Two-stage Model (c) Natural Mean

Figure 3: Simulation results for the gain domain.



(a) Maximax (b) Two-stage Model (c) Natural Mean

Figure 4: Simulation results for the loss domain.

alternative. In particular, as the number of risky alternatives increases, so does the probability of encountering risky alternatives that pay off at their extremes.

General Discussion

Psychological experiments often simplify the complexity of the environments in which we live. This simplification is, of course, necessary for researchers to isolate and manipulate variables of interest while holding other variables constant. However, this simplification could lead to the neglect of the variables most likely to influence behaviour outside the laboratory. In the present study, we demonstrate that one such variable — set size — has a substantial and potentially unavoidable impact on decision making.

Specifically, our experimental results show that increasing the number of risky and safe alternatives amplifies the risks individuals take in the gain domain. This risk-amplification eliminates the description-experience gap — a finding replicated multiple times over the last decade using two-choice environments. Thus, the risk an individual takes crucially depends on the number of alternatives the individual considers.

As an illustration, suppose an individual is assessing how likely he or she is to benefit from a new diet. One method of assessment is to recall other individuals who have benefited from this diet (Tversky & Kahneman, 1973; Galesic, Olsson, & Rieskamp, 2012). If the benefits of the diet are rare, then a recall based on a limited set of other individuals may indicate that the diet is a waste of time. However, as the set size of known individuals who have tried the diet increases, the probability that someone will have a positive testimonial increases.

Our simulations demonstrate that risk-amplification is unavoidable despite changes in information search: Even if an individual dramatically increased their number of samples, the risk-amplification would persist. These results indicate the importance of research on the set size in decision making and further indicate that generalizations from two-choice environments may be at times both quantitatively and qualitatively misleading.

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.

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed. ed.). New York: John Wiley & Sons Inc.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

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 Behaviora 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.

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*, 493–518.

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.

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., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review*, 1-9.

Huber, J., Payne, J., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98.

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.

Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review*, *14*, 1133–1139.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

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.