Encoding context determines risky choice

Christopher R. Madan[1,*], Marcia L. Spetch[2],
Fernanda M. D. S. Machado[3], Alice Mason[3], & Elliot A. Ludvig[3]

[1] School of Psychology, University of Nottingham, Nottingham, UK
[2] Department of Psychology, University of Alberta, Edmonton, AB, Canada
[3] Department of Psychology, University of Warwick, Coventry, UK

* Corresponding author.
Christopher R. Madan
School of Psychology, University Park
University of Nottingham
Nottingham, NG7 2RD, UK.
Email: christopher.madan@nottingham.ac.uk.

Author Contributions

CRM, MLS, and EAL conceived and planned the experiments. MLS and FMDSM collected the data. All authors conducted analyses. CRM took the lead in writing the manuscript, with extensive input from MLS and EAL. All authors discussed the results and commented on the manuscript.

Acknowledgements

We would like to acknowledge Veronica Bergstrom for assistance with data collection and Yang Liu for helping with task programming. This research was funded by the Alberta Gambling Research Institute (held by MLS, EAL, CRM), the Natural Sciences and Engineering Research Council of Canada (held by MLS), and a Leverhulme Early Career Fellowship held by AM. Door images were extracted from “Irish Doors” on fineartamerica.com with permission from Joe Bonita.

Encoding context determines risky choice

Abstract

Both memory and choice are influenced by context: Memory is enhanced when encoding and retrieval contexts match, and choice is swayed by available options. Here, we assessed how context influences risky choice in an experience-based task in two main experiments (119 and 98 participants retained) and two additional experiments reported in the supplemental material (152 and 106 participants retained). Within a single session, we created two separate contexts by presenting blocks of trials in distinct backgrounds. Risky choices were context dependent; given the same choice, people chose differently depending on other outcomes experienced in that context. Choices reflected an overweighting of the most extreme outcomes within each local context, rather than the global context of all outcomes. When tested in the non-trained context, people chose according to the context at encoding and not retrieval. In subsequent memory tests, people displayed biases specific to distinct contexts: extreme outcomes from each context were more accessible and judged as more frequent. These results pose a challenge for theories of choice that rely on retrieval as guiding choice.

Keywords: risky decision making; memory; decisions from experience; memory biases; behavioral economics; context; encoding

Statement of Relevance

People make risky choices in a variety of contexts, whether gambling at a casino, selecting a stock portfolio, or deciding which traffic-prone route to drive on the way home. The context determines the range of available options and outcomes, influencing what people choose. Context, such as location or time of day, also influences what people remember. Here, in a series of experiments, we assess how people make risky choices when they learn about the odds and outcomes from their own experience. We show that people select differently even between the exact same options, when those options appear in different contexts. Moreover, we show that people's memories and risky choice depend on the context in which options are initially encountered, rather than the context at decision time. These results provide a novel demonstration of how memory for past outcomes influences choice with wide-reaching impacts for theories of memory and choice.

Introduction

People's decisions are often informed by prior experiences, reflecting the influence of memory on decision making (e.g., Ludvig et al., 2015; Murty et al., 2016; Shohamy & Daw, 2015). Context has a large impact on memory (see Stark et al., 2018, for a review), leading, for example, to reduced recall when the location changes between study and test (Hupbach et al., 2007; Smith et al., 1978) and playing a prominent role in computational models of memory recall (Howard & Kahana, 2002). Context also significantly influences choice: other available options in a context can lead to range adaptation (Bavard et al., 2018) or even preference reversal in multi-attribute choice (Huber et al., 1982). Some theories have posited that choice is determined by context-dependent samples drawn from memory (e.g., Stewart et al., 2006). Here we show that people choose differently between and remember differently about functionally identical pairs of risky options depending on the context. Moreover, we show that choice is determined by the set of available options present during encoding rather than at retrieval.

Contextual information from the local environment can influence choices. For example, when French music is playing in a supermarket people buy more French than German wine and vice versa when German music is playing (North et al., 1999). Similarly, locating polling stations in a school nudges people toward support of school funding (Berger et al., 2008; Pryor et al., 2014). The local context provided by other available options can also influence choice (Huber et al., 1982; Simonson, 1989; Simonson & Tversky, 1992; Spektor et al., 2019). Consumer preference between two multidimensional products can reverse when a third "decoy" option is introduced that is inferior along one dimension (e.g., cost or quality). Non-human animals also show similar local context effects in their choices (e.g., Shafir et al., 2012).

Experience-based risky choices are also influenced by the set of available values in a decision context. When making decisions based on experience, people tend to be more risk seeking for relative gains than losses—the opposite of decisions made from explicit descriptions (e.g., Ludvig & Spetch, 2011; Kahneman & Tversky, 1979; Konstantinidis, et al., 2018; Wulff et al., 2018). This pattern of experienced-based risky choice appears to be driven by overweighting of the most extreme (best and worst) outcomes in the decision context (Ludvig et al., 2014, 2018). This effect of extremes was confirmed by including other options in the decision context that potentially led to higher (or lower) outcomes, thereby eliminating the bias in risky choice. Moreover, these biases in choice correlate with biases in memory for the extreme outcomes (Madan et al., 2014).

People will sometimes even choose differently for identical decisions across experiments that have different ranges of possible outcomes, suggesting session-level context dependence (Ludvig et al., 2014; Stewart et al., 2015). For example, one decision in Ludvig et al. (2014) was between a fixed gain of 20 points and a risky option that had a 50/50 chance of winning 40 points or nothing. People were more risk averse for this decision in an experiment that included other, larger wins (such that winning nothing was the worst possible outcome) than in an experiment that also included losses (such that winning 40 was the best possible outcome). These differences in risky choice for the exact same decision across experiments involving different decision sets implicate the context as an important determinant of risky choice.

Here we tested whether peoples' choices shift with context changes even within a single experimental session and whether context-dependent effects on choice are based on the decision set present at encoding or retrieval. The main text reports two experiments and the supplementary material contains two additional experiments that replicate the main findings and refine what determines the decision context.

EXPERIMENT 1: LOCAL DECISION CONTEXTS

This experiment tested the stability of choice behavior by eliciting distinct decision contexts that alternated within a session. In memory research, discrete contexts are often elicited through distinct background images (e.g., Anderson & Bower, 1974; Ezzyat & Davachi, 2014). Inspired by this approach, the current experiment provided different contexts by alternating between blocks of decisions with distinct background images and choice options (Fig. 1). One choice (between a fixed gain of 20 points and a risky gain of 10 or 30 points) was common to both contexts and served as the target choice. In the Gain/Loss context, other values were a fixed loss of 20 points and a risky loss of 10 or 30 points. In the High/Low context, other values were a fixed gain of 60 points and a risky gain of 50 or 70 points. Thus, the target risky option provided the best possible outcome (+30) in the Gain/Loss context but the worst possible outcome (+10) in the High/Low context.

If decision contexts create discrete sets of memories, then the extreme-outcome rule predicts that the best and worst outcomes in each local context will be overweighted in memory and choice (Ludvig et al., 2014). This overweighting would produce more risk seeking for the target choice in the Gain/Loss context than in the High/Low context (see comparison outlined in orange/red in Figure 1). If people do not distinguish the contexts, risky choice should be identical in both cases, as the options yield the same values. In either case, we expected that people would show more risk seeking for the highest value decisions (+60 vs. +50/+70) and more risk aversion for the lowest value decisions (−20 vs. −10/−30).

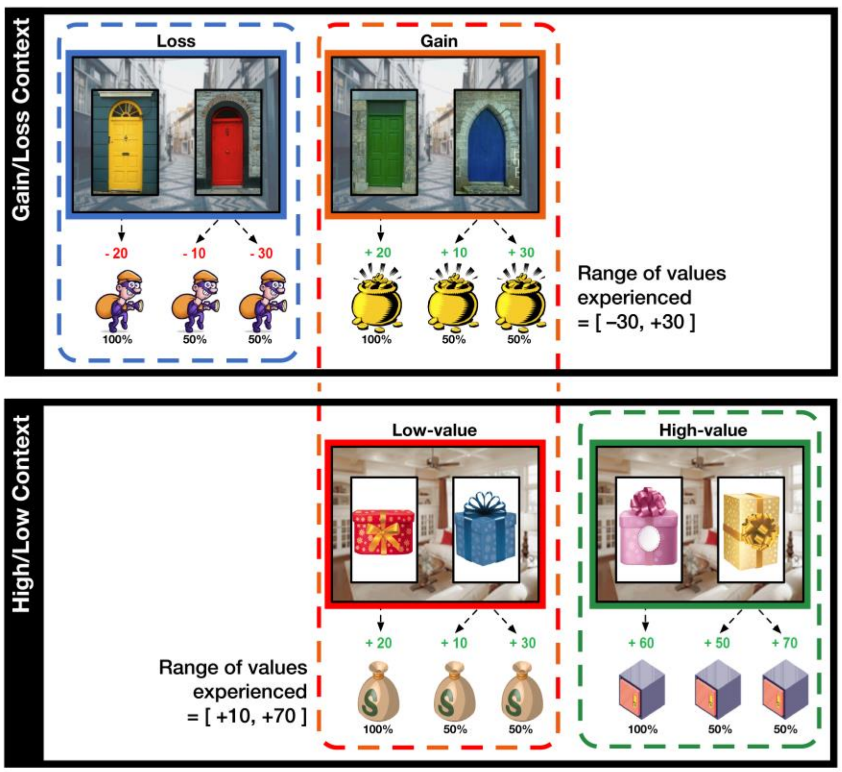

Figure 1. Illustration of the options, outcomes, and context manipulations used in Experiment 1. The computer screen first presented the choice options (e.g., two doors) along with a background image. After the participant made their choice, the chosen door was replaced with an outcome image (e.g., robber or pot of gold), indicating the number of points won or lost following the outcome contingencies shown; the unchosen door was no longer shown. To differentiate between the four option pairs (losses, gains, low value, high value), different option images (distinct doors or distinct gift boxes) and different outcome images (i.e., robber, pot of gold, bag of money, and safe, respectively) were used. The target choices, outlined by the orange/red dashed line, had identical values in the two contexts.

Methods

Participants

A total of 128 participants (99 females; age [$M \pm SD$] = 19.4±1.9 years old) were recruited from the University of Alberta psychology participant pool. An additional 52 participants were recruited but were instructed and paid according to an incorrect payment scheme; as such their data was excluded and not analysed. Informed consent was obtained, and participants received course credit and a cash bonus for participating. They were instructed in groups of up to 15 but performed the task in individual rooms. The number recruited exceeded the number needed (97) to detect a medium effect size (Cohen's d = 0.4) with an alpha of .01 according to a power analysis for this within-subjects design. Procedures were approved by the University of Alberta Research Ethics Board.

Procedure

The experiment consisted of six blocks of trials. Blocks providing a 'Gain/Loss' context, indicated by an outdoor background image, alternated with blocks providing a 'High/Low' context, indicated by an indoor background image (Fig. 1). Fixed options always led to the same outcome, whereas risky options provided two outcomes each with a 50% chance. In the Gain/Loss context, options were selected from four possible doors which led to either a fixed gain (+20), a risky gain (+10 or +30), a fixed loss (−20), or a risky loss (−10 or −30). In the High/Low context, options were selected from four possible gifts which led to either a fixed high-value gain (+60), a risky high-value gain (+50 or +70), a fixed low-value gain (+20), or a risky low-value gain (+10 or +30). As such, there were four different option pairs in the experiment: gain, loss, high value, and low value. Critically, as highlighted by the orange dashed box in Figure 1, the target choices—gain options in the Gain/Loss context and low-value options in the High/Low context—led to identical outcome values, but their relative values within their respective contexts differed. Participants could only learn about the odds and outcomes by selecting the options.

After a choice, the options disappeared, and feedback for the chosen option appeared for 1.2 s. Feedback consisted of the points earned or lost along with an outcome image. The order of the two contexts was counterbalanced across participants, as was the assignment of options to particular outcomes.

For each context, prior to the first block of choice trials, participants were pre-trained with 24 single-option trials to provide experience with the experimental procedure. For these trials, the outcomes associated with each risky option occurred equally often, preventing differences in initial experiences from influencing later choice (e.g., hot-stove or primacy effects; Denrell & March, 2001). Within this block, the gain or high-value options each appeared 8 times, whereas the loss or low-value options each appeared 4 times, such that participants ended the pre-training phase with a positive number of points in both contexts.

Each block of choices consisted of 56 trials and included a mixture of trial types: There were 32 decision trials, which required a choice between fixed and risky options from the same option pairs (16 of each) and 16 catch trials, which required a choice between options from different option pairs with substantially different expected values (e.g., fixed gain vs. fixed loss). On 8 single-option trials, there was only one option that had to be selected to continue; these trials guaranteed that all reward contingencies continued to be experienced, even if the options were initially unlucky, further limiting any hot-stove effects.

In all blocks, trial order was randomized, and each option appeared equally often on either side of the screen. Performance of lower than 60% on catch trials in either context, across the whole experiment, was used as an exclusion criterion, following established protocol from previous experiments (Ludvig et al., 2014; Ludvig & Spetch, 2011; Madan et al., 2014). Participants won or lost points on all trials and were paid $1 for every 2000 points to a maximum of $5 (Canadian).

After the choice task, memory for the outcomes associated with each option was tested in two ways. First, participants were shown the eight options in random order, and, for each option, were asked to report the first outcome that came to mind. Second, participants were again shown the eight options in random order and asked to judge the frequency in percent of each possible outcome (−30, −20, −10, +10, +20, +30, +50, +60, +70). For each option, these nine possible outcomes were displayed simultaneously, and participants typed a number from 0 to 100 below each respective outcome. For both memory tests, each option was presented against a uniform grey background on all trials. Stimuli and data from all experiments are available on the Open Science Framework at: https://osf.io/3mbwu/. All statistical results have been checked with statcheck (Epskamp & Nuijten, 2016).

Analysis

Data from 9 of the 128 participants were excluded from the analyses for scoring less than

60% on the catch trials, leaving 119 participants for the main analyses. The primary dependent measure was the proportion of risky choices in the final training blocks and in the test blocks. Two specific hypotheses were tested:

1. The *Decision-Context Hypothesis* supposes that the extreme outcomes in each context will be overweighted. As a result, risky choice should be higher for the Gain/High-value options (with a high extreme) than for the Loss/Low-value options (with a low extreme) in the corresponding context. In addition, the target choice that has identical outcomes (i.e., Low or Gain, pending the context) should differ across the two contexts with more risk-seeking for that choice in the Gain/Loss context than in the High/Low context. These directional predictions were assessed through three one-tailed, paired *t*-tests.

2. The *Contextual-Memory Hypothesis* supposes that, by the last block in each context, the extreme outcomes in each context will be more salient in memory. For the first-outcome-reported test, this hypothesis was assessed using four χ^2 tests—one for each risky option. For the frequency-judgment tests, this hypothesis was assessed using four one-tailed paired *t*-tests, again one for each risky option. Based on prior work, we expected a robust effect for the Loss/Low-value risky option, but a milder effect with Gain/High-value options, because we have previously found memory biases to be weaker for Gains/High-value outcomes than for the Loss/Low-value outcomes (e.g., Madan et al., 2014, 2017).

Results

Risky choice

Figure 2 shows the mean proportion of risky choice for each context and option pair. In the Gain/Loss context, participants were 10.6±6.6% [*M±95%C.I.*] more risk seeking for gains than losses [$t(118)=3.15$, $p=.001$, Cohen's $d=0.39$]. In the High/Low context, participants were 15.9±6.6% more risk seeking for high-value than low-value options [$t(118)=4.73$, $p<.001$, $d=0.54$]. These results qualitatively replicate our previous findings on an extreme-outcome effect, including evidence for greater differences in risky choice for high- vs. low-value gains than for gains vs. losses (Ludvig et al., 2014; Madan et al., 2014).

Critically, when comparing choice in the two contexts, participants were 11.3±6.3% more risk seeking for the target choices in the Gain/Loss context than in the High/Low context (i.e., comparison highlighted in Figure 1; orange/red bars in Figure 2), despite these options leading to the exact same outcome values [$t(118)=3.52$, $p<.001$, $d=0.40$]. Interestingly, the magnitude of the extreme-outcome effect in the final block of each context was uncorrelated between the two contexts [$r(117)= -.04$, $p=.69$], indicating that the two contexts had been learned relatively independently. Overall risk seeking collapsed across gains and losses, however, was correlated between the two contexts [$r(117)=.45$, $p<.001$].

Thus, participants' biases in risky choice shifted as the visually distinct contexts alternated between blocks. The effect was sufficiently pronounced that even for the exact same target choice (between +20 and a 50/50 chance of +10 or +30), risky choice shifted by more than 10% even within the same participants within the same session, determined by the decision context.

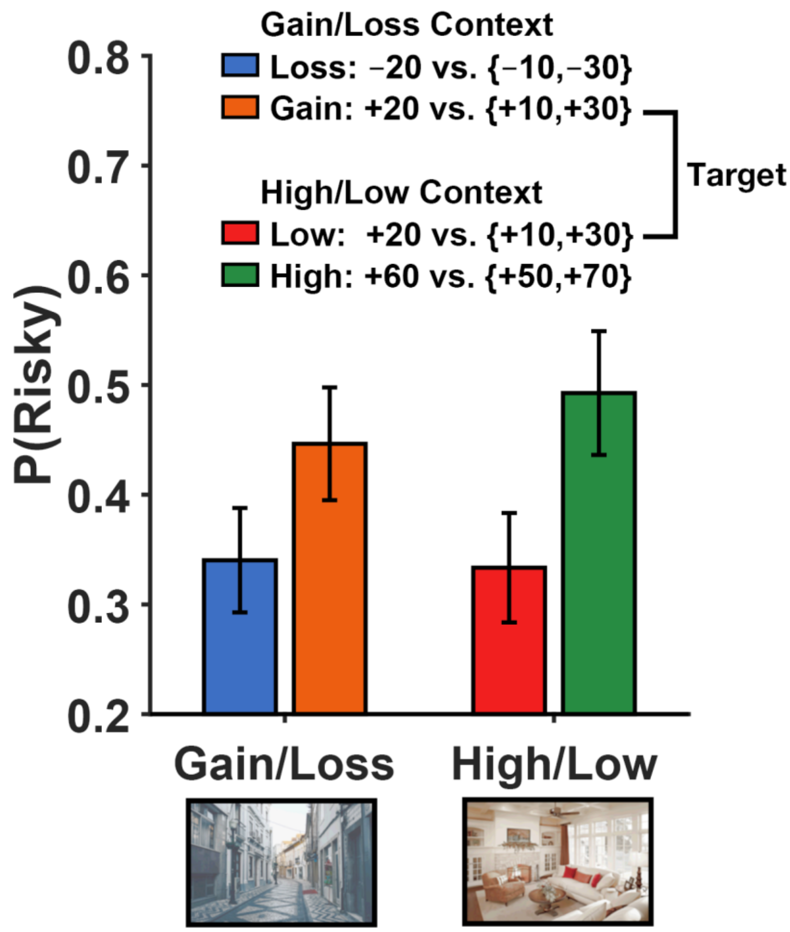

Figure 2. Proportion of risky choices for each decision set and their respective decision context, averaged across the last block in each context for Experiment 1. Error bars represent 95% confidence intervals.

Memory tests

Figure 3 shows how both memory tests suggested some overweighting of the extreme outcomes, convergent with prior findings (Madan et al., 2014), as well as some context dependence in overweighting. The memory biases were more robust for the loss/low-value decisions, also consistent with prior work.

In the first-outcome-reported test, for both the loss and low-value options, participants were significantly more likely to report the worse value (−30 and +10, respectively) [Loss: $\chi^2(1,N=88)=35.64$, $p<.001$; Low: $\chi^2(1,N=92)=31.70$, $p<.001$]. Participants did not exhibit a bias in their reported outcomes for gains [$\chi^2(1,N=85)=0.11$, $p=.74$], and there was only a weak trend toward responding with the better outcome for the high-value option [$\chi^2(1,N=99)=2.92$, $p=.088$]. Results were similar in the frequency-judgment test, where people reported a significantly larger percent for the worse outcome for the loss and low-value options [Loss: $t(102)=6.16$, $p<.001$, $d=1.06$; Low: $t(102)=7.02$, $p<.001$, $d=1.19$], but did not report a reliable difference in judged percent for the outcomes of the gain and high-value options [Gain: $t(102)=0.29$, $p=.39$, $d=0.05$; High: $t(102)=0.82$, $p=.21$, $d=0.14$]. Thus, by both measures, the worst outcome in each context seemed to be particularly salient in memory. The context-dependence of this salience is highlighted by the +10 outcome which was reported more often and judged as having a higher frequency in the High/Low context than in the Gain/Loss context.

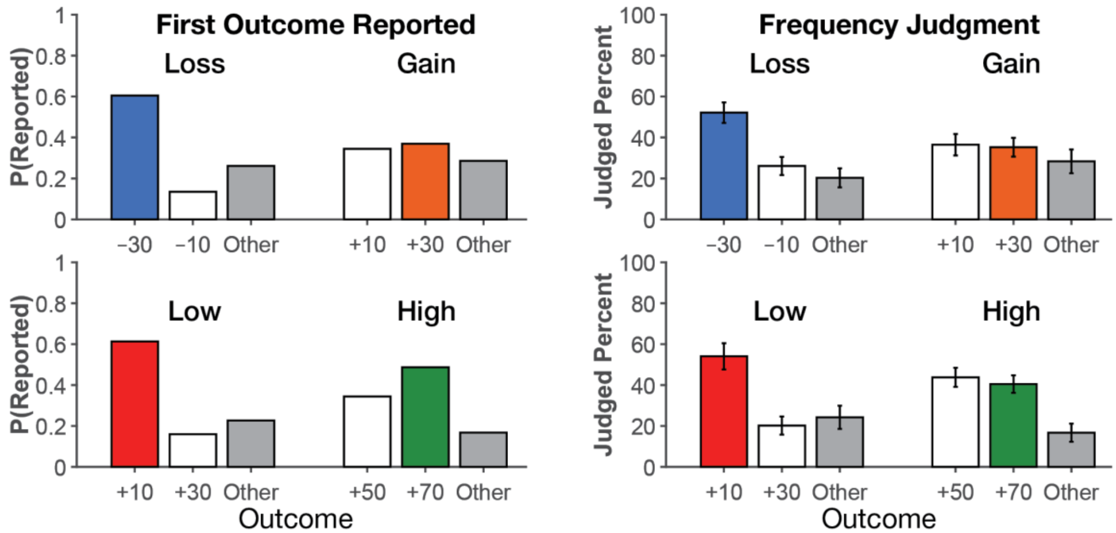

Figure 3. Results of the two memory tests for the risky options in the two decision contexts in Experiment 1. Participants were more likely to report the extreme outcomes first and judged the lowest outcome in each context as having occurred most frequently. Coloured bars are local extreme outcomes and white bars are non-extremes. The colour code matches the conditions in Figure 1 (Blue = Loss; Orange = Gain; Red = Low; Green = High). Error bars represent 95% confidence intervals.

Discussion

The context manipulation in this experiment successfully established distinct decision contexts. Participants made different risky choices even for option pairs that led to the exact same values; choices depended on the other values present in the same context, i.e., choices in the Gain and Low-value decisions (as highlighted in Figure 2). The memory tests also showed context dependence: people were more likely to report the extreme outcomes in each context as the first to come to mind and judged the worst outcome in each context as more frequent (see Fig. 3). Though we have previously demonstrated different risky choice for options leading to the same outcomes across experiments (e.g., Ludvig et al., 2014; Madan et al., 2014), this experiment is the first demonstration that risk preference for a given decision and related memory biases can differ across blocks of trials within a single session, based on the local context.

EXPERIMENT 2: ENCODING OR RETRIEVAL OF CONTEXTUAL CUES

Here we sought to extend the findings of Experiment 1 by testing whether the context of encoding or retrieval is crucial for determining which outcomes are overweighted in memory and choice. The results of Experiment 1 could be due to processes operating at either encoding or retrieval. From an encoding perspective, outcome values might be encoded relative to the other values present in the context during learning (Rangel & Clithero, 2012). Values at the extremes of that set may be given more weight during encoding, causing them to be retrieved/sampled more readily when the option is later re-experienced. An encoding account is also congruent with a selective-attention mechanism whereby goal-congruent items influence value integration (e.g., Kunar et al., 2017; Usher et al., 2019).

Alternatively, context-dependent biases could be due to retrieval processes during choice. For example, if outcome values are encoded together with an association to their learning context, then the context present during choice may retrieve a memory of other values associated with that context. This retrieved set of values may determine the comparison set for evaluating values during choice (as in Decision by Sampling; Stewart et al., 2006), with extreme values being given most weight. A retrieval-based interpretation is consistent with findings that risky choice can be altered by presenting reminders of previous outcomes (Bornstein et al., 2017; Ludvig et al., 2015).

To distinguish between encoding and retrieval hypotheses, we used the same design as Experiment 1, but with two modifications: (1) Choice stimuli and background images were changed to make the target options more interchangeable. Specifically, we used eight distinct doors (rather than four doors and four gifts) as choice stimuli and two distinct street scenes as background images for the two decision contexts. (2) After the six choice blocks, we presented two blocks of probe tests without feedback, in which the doors providing the target choice were presented in either their training context (Same) or untrained context (Reversed).

If the context of encoding is crucial, choices should be independent of the testing context. Participants should be more risk seeking for the target choices initially encountered in the Gain/Loss context than for those initially encountered in the High/Low value context, regardless of the test context. If the context of retrieval determines choice, however, then people should choose differently between the same pairs of doors in the two testing contexts. Specifically, participants should be more risk seeking for both target choices when tested in the Gain/Loss context than in the High/Low context. The design, hypotheses, analysis and expected choice results were pre-registered on the Open Science Framework: https://osf.io/kv458/.

Methods

Participants

A total of 103 participants (72 females; age [$M \pm SD$] = 20.8±3.4 years old) were drawn from the same participant pool, and recruitment and consent procedures were the same as in Experiment 1. Participants were paid $1 for every 200 points after the first 8000 earned up to a maximum of $5 (Canadian).

Procedure

General procedures were the same as in Experiment 1 with the following exceptions. The task consisted of 8 blocks. The first 6 blocks alternated between two contexts in which 4 possible doors appeared alone or in pairs against a background outdoor scene that was unique to each context; these will be referred to as the training blocks. The last two blocks were test blocks, one for each context. In these blocks, choices were not followed by feedback. Prior to these two blocks, participants were informed by an instruction screen that they would *not* receive feedback for their choices, but that points would still be won or lost in the same way as before.

Trials during the training blocks were identical to Experiment 1, except that all choice stimuli were doors, and the two backgrounds were distinct street scenes rather than an outdoor and indoor scene. In the test blocks, only the doors that led to the target choice of +20 versus a 50/50 chance of +10 or +30 appeared. These were tested in both contexts (order randomized across participants) without feedback. There were two test blocks of 16 trials each, providing a total of 8 trials with each target choice in each context.

Following the test blocks, participants were given the same two types of memory tests (first outcome reported and frequency judgement) described in Experiment 1.

Analysis

Five participants were excluded from the analysis for scoring less than 60% on the catch trials, leaving 98 participants. As per our pre-registration, comparisons were evaluated with an alpha of .01. The primary dependent measure was the proportion of risky choices in the final training blocks and in the test blocks. Four specific pre-registered hypotheses were tested:

1. The *Context-Replication Hypothesis*, which supposes that by the end of training the extreme outcomes in each context will be overweighted, was assessed through three one-tailed paired *t*-tests. First, we tested the prediction that risky choice would be higher for the higher value option (high or gain) than for the lower value option (low or loss) in both contexts in the final block of the training phase. Second, we compared risky choice for the target choice in the two contexts. We predicted more risk-seeking for that choice in the Gain/Loss context than in the High/Low context.

2. The *Encoding Hypothesis* supposes that the context effects are due to the way the doors were initially encoded in the training contexts. As a result, we predicted that, regardless of the test context, there would be more risk-seeking for the target choice learned in the Gain/Loss context than for the target choice learned in the High/Low context. This was assessed with two one-tailed paired *t*-tests, examining risky choice for the target choice in the two contexts during testing.

3. The *Retrieval Hypothesis* supposes that the context effects are due to the context in which outcomes are retrieved at the time of choice. As a result, the test context should matter, and, for the target choice, people should be more risk-seeking when tested in the Gain/Loss context (where the other options were worse) than in the High/Low context (where the other options were better). This was assessed through a two-way (Training Context by Test Context) repeated-measures ANOVA. This hypothesis predicted a main effect of Test Context.

4. The *Noise Hypothesis* supposes that the context shift in the test blocks makes people behave more randomly as the discrepant context makes them rely less on their prior feedback. As a result, choice should shift toward indifference whenever doors are tested outside their training context. This hypothesis was tested by calculating the difference between each individual's average absolute deviation from 50% in their risky choices in the two test contexts; a shift toward indifference with a context change should result in lower absolute deviation scores in the Reversed context. A one-tailed one-sample *t*-test was used to test for reliable differences from 0 across the two contexts.

Memory tests were analyzed in the same way as in Experiment 1. We did not preregister specific predictions for these tests.

Results

Risky choice

Figure 4 shows the mean proportion of risky choices for each context and option pair during the last training block with each context. In the Gain/Loss context, participants were 13.8±7.6% [M±95%$C.I.$] more risk seeking for gains than losses [$t(97)=3.62$, $p<.001$, $d=0.37$]. In the High/Low context, participants were 24.1±8.3% more risk seeking for high-value than low-value choices [$t(97)=5.72$, $p<.001$, $d=0.58$]. These results qualitatively replicate results from Experiment 1.

Critically, when comparing the two contexts, participants were 22.7±7.9% more risk seeking for the target choice in the Gain/Loss context than in the High/Low context [$t(97)=5.68$, $p<.001$, $d=0.57$]. The magnitude of the extreme-outcome effect was again uncorrelated between the two contexts [$r(97)= .037$, $p=.72$], indicating that the two contexts were learned relatively independently. Overall risk seeking (collapsing across all risky decisions) was slightly, but not significantly, correlated between the two contexts [$r(97)=.191$, $p=.058$].

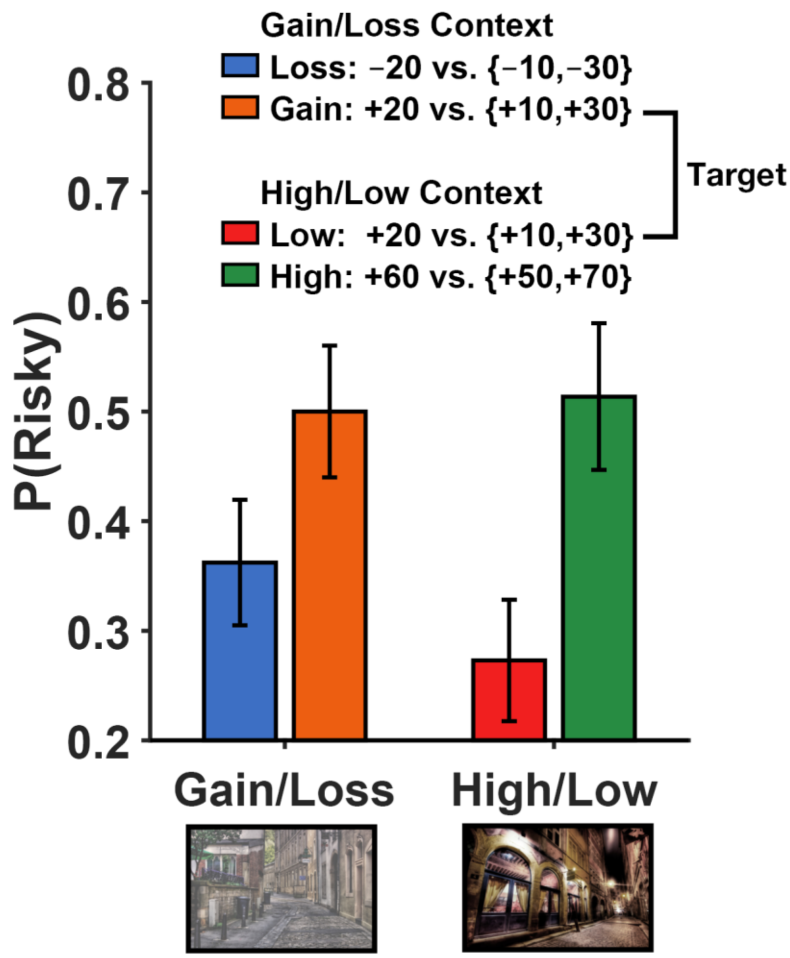

Figure 4. Proportion of risky choices for each of the four option pairs separated by their respective decision contexts and averaged across the last block in each context for Experiment 2. Error bars represent 95% confidence intervals.

Test blocks

Figure 5 shows the mean risky choice for the target choices when they were presented without feedback during testing. The test context had no discernable effect. When tested in the Same context, participants were 22.2±9.8% more risk seeking for the target choice trained in the Gain/Loss context than in the High/Low context [$t(97)=4.50$, $p<.001$, *d*=0.46]. Similarly, when tested in the Reversed context, participants were 22.5±10.0% more risk seeking for the target choice trained in the Gain/Loss context than in the High/Low context [$t(97)=4.48$, $p<.001$, $d=.45$]. A two-way ANOVA confirmed a main effect of Choice [$F(1,97)=21.1$, $p<.001$, $\eta_p^2=.18$], but no effect of Test Context [$F(1,97)=0.51$, $p=.48$, $\eta_p^2=.005$] and no interaction [$F(1,97)=0.015$, $p=.90$, $\eta_p^2=.00$]. There was no evidence in support of the noise hypothesis: The average deviation from indifference (0.5) did not differ for risky choices conducted in the Same context [35.7±3.0%] from the risky choices conducted in the Reversed context [36.0±2.8%; $t(97)=0.28$, $p=.78$, $d=0.03$]. These data support the notion that the encoding context is more important than the retrieval context in determining later choice.

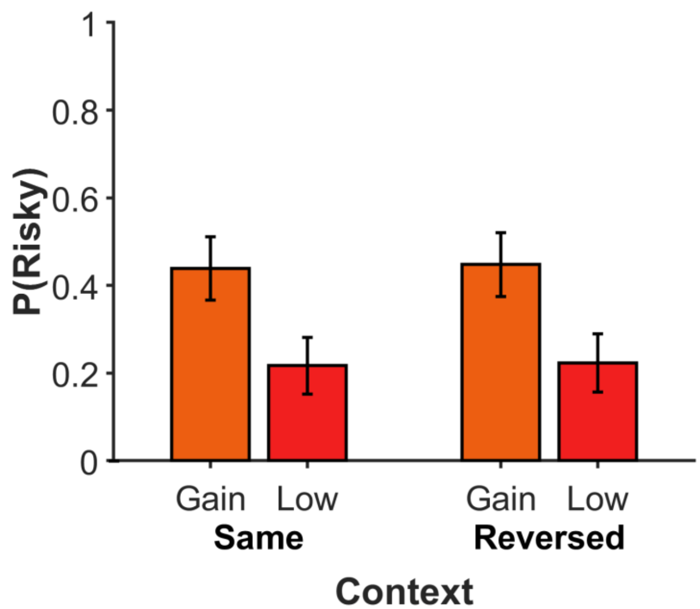

Figure 5. Results of the probe choice tests in Experiment 2. Proportion of risky choices for the target choice (+20 vs +10/+30) trained in the Gain/Loss context or the High/Low context when tested without feedback in the Same or Reversed context.

In an additional exploratory analysis, we sought to solidify the argument for/against the encoding/retrieval hypotheses, respectively. Here we tested whether risky choices in different conditions of the test blocks were independent. The Encoding Hypothesis predicts that the proportion of risky choices for gain and low-value decisions should be highly correlated between the Same and Reversed contexts because the choices should be invariant to test context. In addition, the Encoding hypothesis predicts low correlations between risky choices for gain and low-value option pairs within each test context, as these would have been encountered independently in training. In contrast, the Retrieval Hypothesis predicts the opposite: low correlations for each option pair across test contexts, but high correlations between the gain and low-value decisions within a context.

Figure 6 shows how these results strongly support the Encoding Hypothesis: Correlations were very strong when comparing the proportion of risky choices made for the gain decisions in the Same or Reversed test context [$r(97)=.901$, $p<.001$] and similarly high for the low-value decisions [$r(97)=.920$, $p<.001$]. In contrast, correlations between risky choices for gain and low-value decisions within each context were very low, suggesting that these decisions were independent of each other despite having identical outcome values [Same context: $r(97)=.014$, $p=.89$; Reversed context: $r(97)=.002$, $p=.99$].

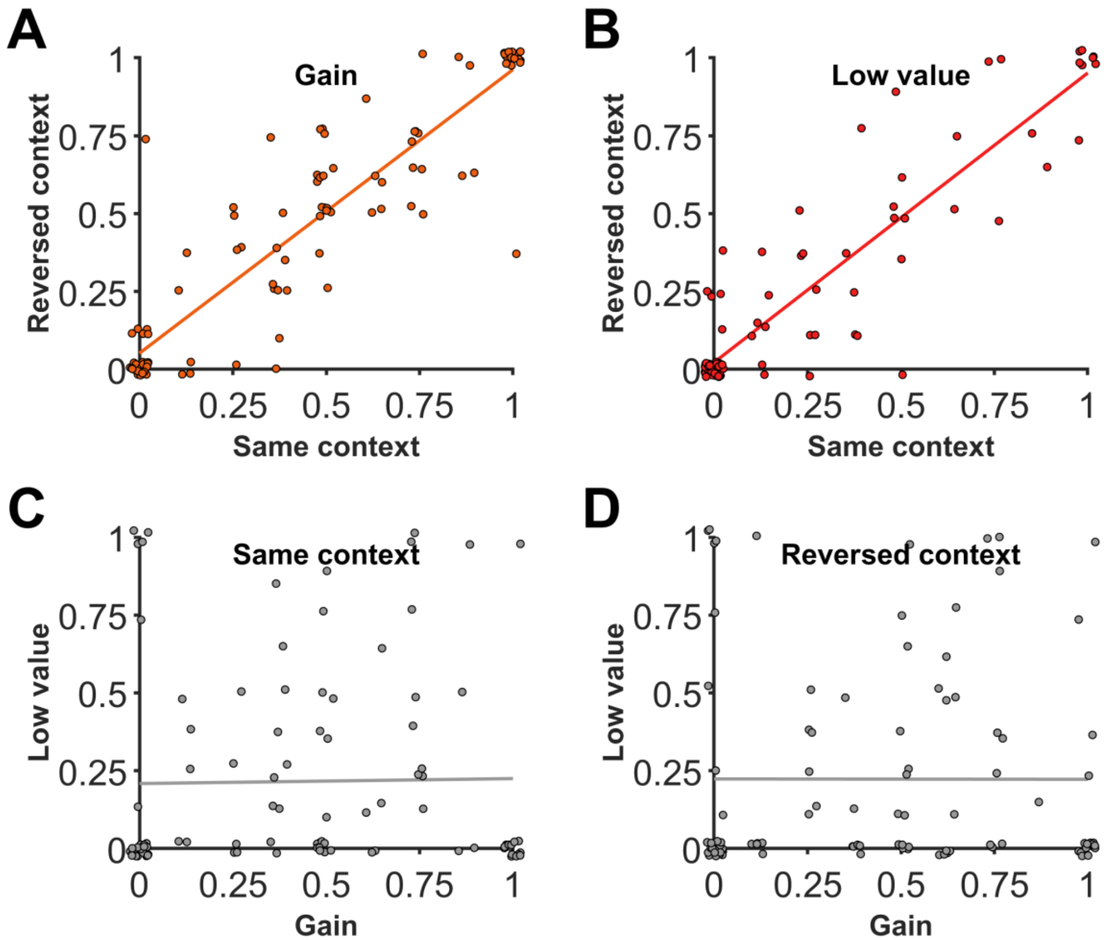

Figure 6. Proportion of risky choices made in the test blocks for (A) gain decisions and (B) low-value decisions, between Same and Reversed contexts. The opposite comparison is shown in the next panels, with the proportion of risky choices in the (C) Same context and (D) Reversed context, between gain and low-value decisions. Each dot represents an individual participant; dot locations are jittered to reduce overlap.

Memory tests

Figure 7 shows the results of the memory tests were similar to those seen in Experiment 1, with context-dependent overweighting of the extreme loss and low-value outcomes. In the first-outcome-reported test, for both the loss and low-value options, participants were significantly more likely to report the worse value (−30 and +10, respectively) [Loss: $\chi^2(1,N=71)=8.80$, $p=.003$; Low: $\chi^2(1,N=76)=23.21$, $p<.001$]. Differences in reporting of outcomes were not significant for the risky gains [$\chi^2(1,N=76)=1.90$, $p=.17$], nor for the risky high-value option [$\chi^2(1,N=78)=2.51$, $p=.11$]. The frequency-judgment test also showed a context-dependent bias in which people reported higher percentages for the worse outcome for the loss and low-value options [Loss: $t(93)=5.10$, $p<.001$, $d=0.526$; Low: $t(90)=6.19$, $p<.001$, $d=.65$], but no reliable difference in judged percent for the outcomes of the gain and high-value options [Gain: $t(92)=0.07$, $p=.948$, $d=0.01$; High: $t(91)=0.58$, $p=.56$, $d=0.06$]. Thus, by both measures, the worst outcome in each context was particularly salient in memory. The context dependence of this salience is highlighted by the +10 outcome which was reported more often [$\chi^2(1,N=91)=8.01$, $p=.005$] and judged as having a higher frequency ($t(88)=4.07$, $p<.001$, $d=0.43$) in the High/Low context (where it was the worst outcome) than in the Gain/Loss context (where it was an intermediate outcome).

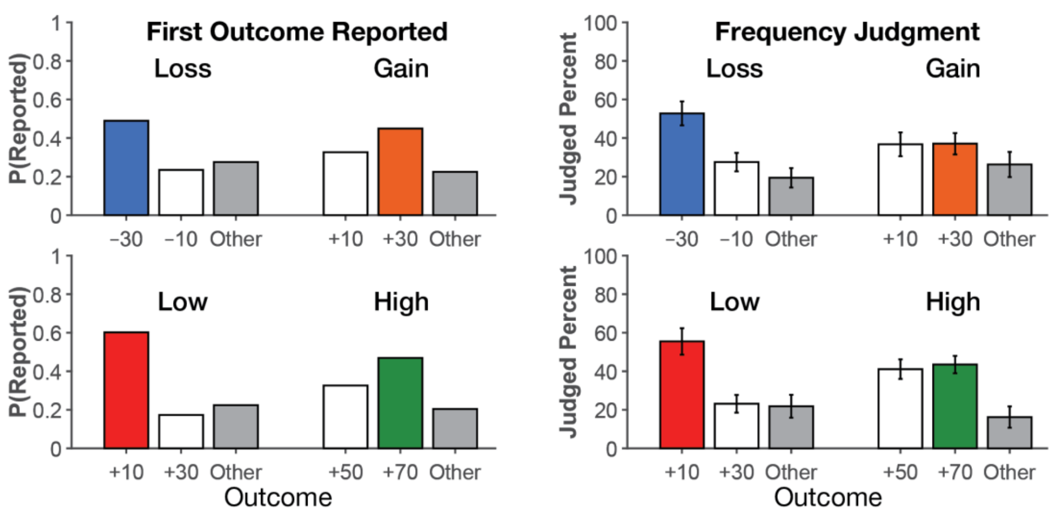

Figure 7. Results of the two memory tests for each decision context in Experiment 2. Participants were more likely to report the extreme outcomes first and judged the lowest outcome in each context as having occurred most frequently. Coloured bars are local extreme outcomes and white bars are non-extremes. The colour code matches the conditions in previous figures (Blue = Loss; Orange = Gain; Red = Low; Green = High). Error bars represent 95% confidence intervals.

Supplemental experiments

Two additional experiments are reported in the Supplemental Material that address alternative explanations related to the necessary and sufficient conditions for creating distinct decision contexts (see Table S1). The results show that distinct background images are not necessary for establishing a local decision context, but temporal grouping of the choices is not sufficient to discretize the contexts. The distinct visual cues from the choice stimuli, however, are sufficient, and may even be necessary, to distinguish the contexts (see Exp. S2). These distinct visual cues may also serve as retrieval cues for the decision context in which they were encoded. Together with Exp. 2, these results clearly show that choice is determined by the decision context during encoding, and not the decision context at retrieval.

General Discussion

Here, in two experiments, we demonstrated that people's risky choices are not stable, even within a single experimental session, but rather depend on the other outcomes experienced during the context of encoding. Risky choice was biased by the most extreme outcomes in a particular decision context, rather than the global context of the whole experiment, and people also remembered those outcomes more strongly. Even for the exact same decisions (between +20 and a 50/50 chance of +10 or +30), changes in context substantially shifted both risky choice (>10% in Exp. 1 and >20% in Exp. 2) and memory for extremes, even for the same participants within a single session. Moreover, when tested in the opposite context, people chose in line with the initial training context, suggesting that the context of encoding is critical for this memory-based choice.

These findings have theoretical implications for memory-based theories of experience-based decision making (e.g., Shohamy & Daw, 2015; Weber & Johnson, 2006). For example, according to Decision by Sampling Theory (Stewart et al., 2006), the values of options presented at choice are compared to a small sample in working memory; the sample comes both from other values in the immediate context and from values stored in long-term memory. Our results suggest that such samples would have to come from values presented in the encoding context rather than in the context at the time of choice. Thus, our results pose significant challenges for retrieval-based models of how memory affects choice, but are more consistent with a recent reinforcement-learning (RL) model that assumes that the influence of context on value operates during the learning process (Spektor et al., 2019).

The current results show how unstable choices can be and add to the growing evidence that choices depend on properties of the decision context (e.g., Huber et al., 1982; Simonson & Tversky, 1992). An important open question is how to pull the various context effects into a single process model of risky choice. One possibility is inspired by recent RL models that have attempted to integrate aspects of episodic memory (e.g., Gershman & Daw, 2017). Exactly how to incorporate other context effects from the decision-making literature is not clear, but may require real-time integration mechanisms as in decision-field theory or the drift-diffusion model (Ratcliff & McKoon, 2008; Roe et al., 2001). Our work, however, suggests how important modeling context effects will be for creating a reliable model of human decision-making when learning from experience.

References

Anderson, J. R., & Bower, G. H. (1974). Interference in memory for multiple contexts. *Memory & Cognition, 2*, 509-514.

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.

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature Communications*, *9*, 4503.

Berger, J., Meredith, M., & Wheeler, S. C. (2008). Contextual priming: Where people vote affects how they vote. *Proceedings of the National Academy of Sciences*, *105*, 8846-8849.

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications, 8*, 15958.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science, 12*, 523–538.

Epskamp, S., & Nuijten, M. B. (2016). statcheck: Extract statistics from articles and recompute p values. R package. http://CRAN.R-project.org/package=statcheck

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*, 1179-1189.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. Annual Review of Psychology, 68, 101-128

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.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269-299.

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

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, *14*, 47-53.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47,* 263-292.

Konstantinidis, E., Taylor, R. T., & Newell, B. R. (2018). Magnitude and incentives: revisiting the overweighting of extreme events in risky decisions from experience. *Psychonomic Bulletin & Review*, *25*(5), 1925-1933.

Kunar, M. A., Watson, D. G., Tsetsos, K., & Chater, N. (2017). The influence of attention on value integration. *Attention, Perception & Psychophysics*, *79*, 1615–1627.

Ludvig, E.A., Madan, C.R., & Spetch, M.L. (2014). Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making*, *27,* 146–156.

Ludvig, E.A., Madan, C.R., & Spetch, M.L. (2015). Priming memories of past wins induces risk seeking. *Journal of Experimental Psychology: General*, *144*, 24-29.

Ludvig, E.A., & Spetch, M.L. (2011). Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events? *PLOS ONE, 6,* e20262.

Madan, C.R., Ludvig, E.A., & Spetch, M.L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review, 21,* 629–636.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2019). Comparative inspiration: From puzzles with pigeons to novel discoveries with humans in risky choice. *Behavioural Processes*, *160*, 10-19.

Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of Experimental Psychology: General*, *145*, 548-558.

North, A. C., Hargreaves, D. J., & McKendrick, J. (1997). In-store music affects product choice. *Nature*, *390*, 132.

Pryor, B., Mendez, J. M., & Herrick, R. (2014). Let's be fair: Do polling locations prime votes? *Journal of Political Sciences & Public Affairs*, *2*, 3.

Rangel, A., & Clithero, J. A. (2012). Value normalization in decision making: theory and evidence. *Current Opinion in Neurobiology*, *22*, 970-981.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873-922.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*(2), 370-392.

Shafir, S., Waite, T. A., & Smith, B. H. (2002). Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis*). *Behavioral Ecology and Sociobiology*, *51*, 180-187.

Shohamy, D., & Daw, N.D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences*, *5*, 85-90.

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

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*, 281-295.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342-353.

Spektor, M. S., Gluth, S., Fontanesi, L., & Rieskamp, J. (2019). How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychological Review*, *126*(1), 52-88.

Stark, S. M., Reagh, Z. M., Yassa, M. A., & Stark, C. E. (2018). What's in a context? Cautions, limitations, and potential paths forward. *Neuroscience Letters*, *680*, 77-87.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology, 53(1),* 1-26.

Stewart, N., Reimers, S., & Harris, A. J. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61*(3), 687-705.

Usher, M., Tsetsos, K., Glickman, M., & Chater, N. (2019). Selective Integration: An Attentional Theory of Choice Biases and Adaptive Choice. *Current Directions in Psychological Science*, *28*, 552-559.

Weber, E. U. & Johnson, E. J. (2006). Constructing preferences from memory. in *The Construction of Preference* (Eds. S. Lichtenstein & P. Slovic), pp. 397–410. Cambridge University Press.

Wulff, D. U., Mergenthaler, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description–experience gap. *Psychological Bulletin, 144*, 140-176.

Supplemental Materials for: Encoding Context Determines Risky Choice

OVERVIEW

These supplemental materials present two additional experiments that replicate and extend the two experiments presented in the main text. Both supplemental experiments dispense with the use of background images to distinguish the contexts. Table S1 summarizes the methodological details and key results. Exp. S1 recreates Exp. 1 from the main text, except omits the background images and does not have distinct doors to represent the options with the same outcomes. The results show that some visual distinctiveness is necessary to create separate contexts, and temporal grouping alone is not sufficient. Exp. S2 replicates Exp. 2 from the main text, except omitting the background images; the doors, however, are visually different in the two contexts. In addition, the post-training test trials were different and placed the two sets of target choices (i.e., gains and low-value option pairs) in the same temporal context. Results exactly match the key results for Exp. 2, with local context driving the overweighting of extremes in memory and choice, and the effect being driven by the context at encoding.

As shown in Table S1, all experiments included the temporal grouping of alternating blocks of two option pairs during training. As such, based solely on the underlying temporal structure of all experiments (i.e., ignoring the visual features), all four experiments are identical. Exp. 1 and 2 had distinct backgrounds that served as visual signals for the current decision context. Exp. 1, 2, and S2 used distinct choice stimuli that could also visually signal the current context. These three experiments all demonstrated that the range of values experienced within a block dictated choice, indicating that decision contexts were functionally distinct—which we refer to in Table S1 as evidence of a local decision context. Removing both the background images and the distinct choice stimuli, while still retaining temporal groupings of trial types, eliminated the effects of local decision context on choice, and led to behavior consistent with a global decision context.

Additionally, both Exp. 2 and S2 included test blocks at the end, where the choice stimuli (doors) were presented in new contexts, either by using a mismatched background image (Exp. 2) or by presenting choices involving a mix of doors used for the Gain and Low-value option pairs (Exp. S2). In both cases, choices were congruent with the risky choices during training, suggesting that people made choices according to the encoding context, rather than retrieval context.

Table S1. Details of methodology and primary results from the two main and two supplemental experiments.

	Exp. 1	Exp. 2	Exp. S1	Exp. S2
Method:				
Background Images	√	√	×	×
Temporal Grouping	√	√	√	√
Distinct Choice Stimuli	√	√	×	√
Results:				
Global/Local	L	L	G	L
Encoding/Retrieval	-	E	-	E

EXPERIMENT S1: NECCESSITY OF DISTINCT VISUAL CUES

The context effects seen in Experiment 1 clearly indicated that participants were able to segregate contexts that were visually distinct (different background cues and visually distinct choice options) and temporally segregated by alternating blocks of decision sets. This experiment tested whether participants also discretize contexts based on the temporal structure of the blocks alone.

Methods

Participants

A total of 155 participants (109 females; age [$M \pm SD$] = 19.3±2.4 years old) were drawn from the same participant pool at the University of Alberta, and all recruitment, consent, and payment procedures were the same as in Experiment 1.

Procedure

The structure of the experiment was identical to that in Experiment 1, but there were no visual cues to differentiate between the two choice contexts. Instead all choices were presented against a uniform gray background. Moreover, all choice stimuli were doors, and the same two target doors served as the gain doors in the gain/loss context and the low-value doors in the high/low context. The number and composition of trials in each block, and the procedural details of each trial were the same as in Experiment 1. Three participants were excluded because they scored less than 60% on the catch trials, leaving 152 participants for the main analyses. After the last block of choice trials, recall and frequency memory tests were conducted with the six doors.

Results

Risky choice

Figure S1 shows the mean proportion of risky choices for each option pair. In the blocks with gain and loss choices, participants were only 0.2±4.6% [*M±95%C.I.*] more risk seeking for gains than losses [$t(151)=0.09$, $p=.93$, Cohen's $d=0.01$]. In the blocks with high and low-value gains, however, participants were 13.5±5.3% more risk seeking for high-value than low-value options [$t(151)=5.08$, $p<.001$, $d=0.41$].

Critically, risky choices for the target decisions (+20 versus +10/+30) were only 1.8±2.5% higher on Gain/Loss blocks than on High/Low blocks [$t(151)=1.37$, $p=.17$, $d=0.11$].

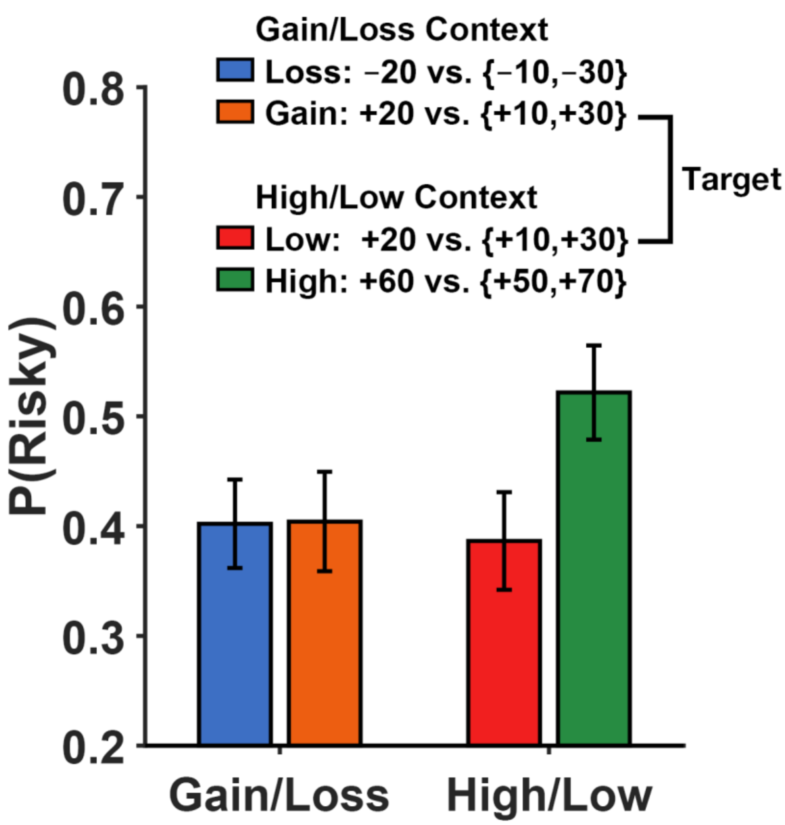

Figure S1. Proportion of risky choices for each option pair and their respective

decision context, averaged across the last block in each context for Experiment S1. Error bars represent 95% confidence intervals.

Memory tests

For the recall tests, participants were more likely to report the lowest value for the risky loss door and for the risky target door and to report the higher outcome for the risky high-value door [$\chi 2(1)$= 26.1, 7.81, and 96.6, all *ps* <.01], as shown in Figure S2. In the frequency-judgment test, participants reported that the lower-valued outcome occurred more often for the risky loss door [$t(142)=6.49$, $p<.001$, $d=0.54$] and for the risky target door [$t(142)=4.72$, $p<.001$, $d=0.40$], but there were no reliable differences in judged percent of the two outcomes for the high-value risky door [$t(146)=0.09$, $p=.93$, $d=0.01$].

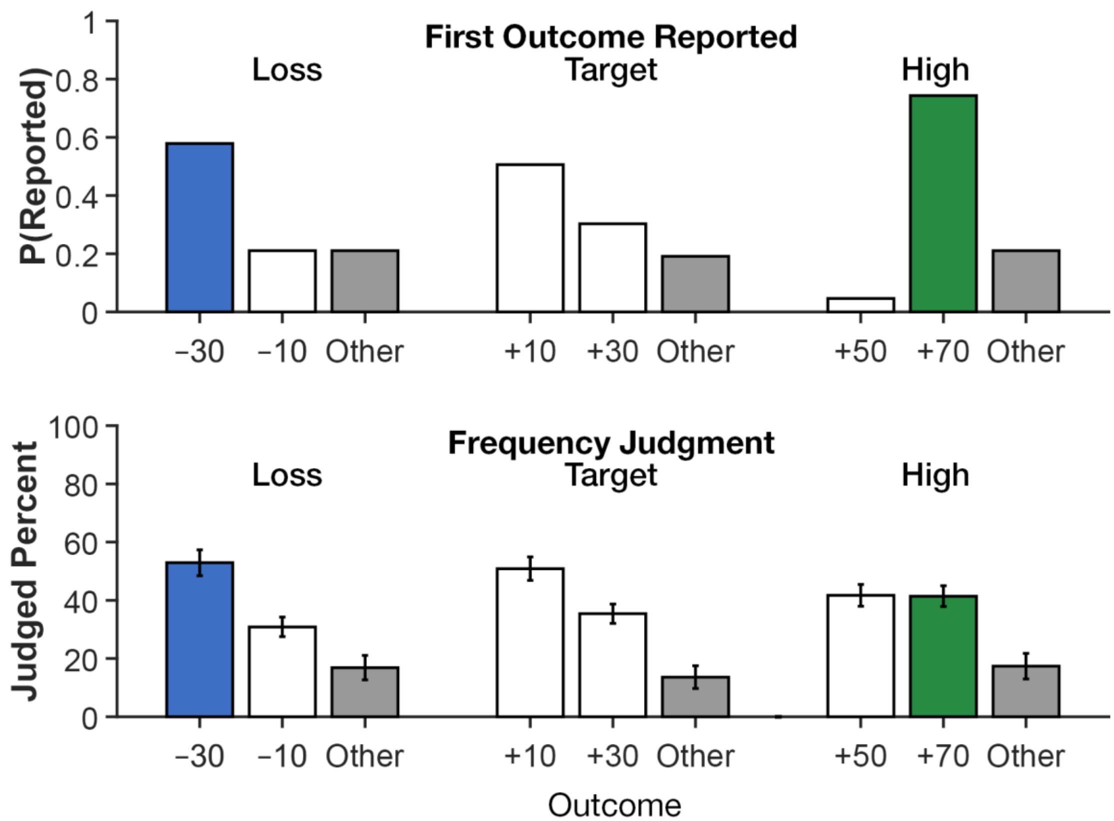

Figure S2. Results of the two memory tests for each decision in Experiment S1. Coloured bars are global extreme outcomes and white bars are non-extremes. The colour code matches the conditions in previous figures (Blue = Loss; Green = High).

Error bars represent 95% confidence intervals.

Discussion

The alternating block structure of the two decision sets was not sufficient to induce a local context for choice when no visual cues indicated the change in context. In particular, unlike in Experiment 1, risky choice on the target choices did not differ depending on whether they were presented in blocks with losses or in blocks with higher value gains. Thus, visually distinguishing the contexts, either by choice stimuli or background effects seems necessary for these context-dependent biases.

EXPERIMENT S2: ROLE OF BACKGROUND IN DETERMINING CONTEXT

Experiment S1 showed that the alternation of decision sets, without any distinctive cues to signal the context change, was not sufficient for discretization of the contexts. Here we removed the distinct background cues as in Experiment S1 but provided visually distinct choice options for the target decisions in the two decision contexts.

If visually-distinct but functionally identical choice options acquired different values as a result of grouping with other options, risk preference should show local context effects as seen in Experiments 1 and 2. If, however, the background image is required to segregate the contexts, then risk preferences for the target decisions should not differ between contexts. As in Experiment 2, probe tests were conducted in an altered context to assess whether biases were based on encoding or retrieval.

The sample size, methods, hypotheses, and analyses for the experiment were preregistered on the Open Science Framework (https://osf.io/gt4rc/).

Methods

Participants

A total of 106 participants (71 female; age [$M \pm SD$] = 25.1±4.9 years old) from the University of Warwick were recruited using the SONA online sign-up system and provided informed consent. Participants were paid an honorarium of £4 (UK pounds) along with a cash bonus for participating. Participants were paid £1 for every 200 points after the first 8000 points earned, up to a maximum bonus of £3. Participants were instructed in groups of up to 12. All participants scored more than 60% overall on the catch trials and were retained in the main analysis. Procedures were approved by the Warwick Psychology Research Ethics Committee.

Procedure

The experimental design was similar to Experiment 2, except that for all blocks, choice stimuli were presented against a uniform white background, rather than distinct images. Door images always appeared on a white background screen (Figure S3) and clicking a door led to feedback (points awarded or deducted) for one second before a button saying *NEXT* appeared. Pressing the “Next” button started an inter-trial interval which varied randomly from one to two seconds and provided a uniform white screen. The accumulated points were shown at the end of each block rather than at the end of each trial.

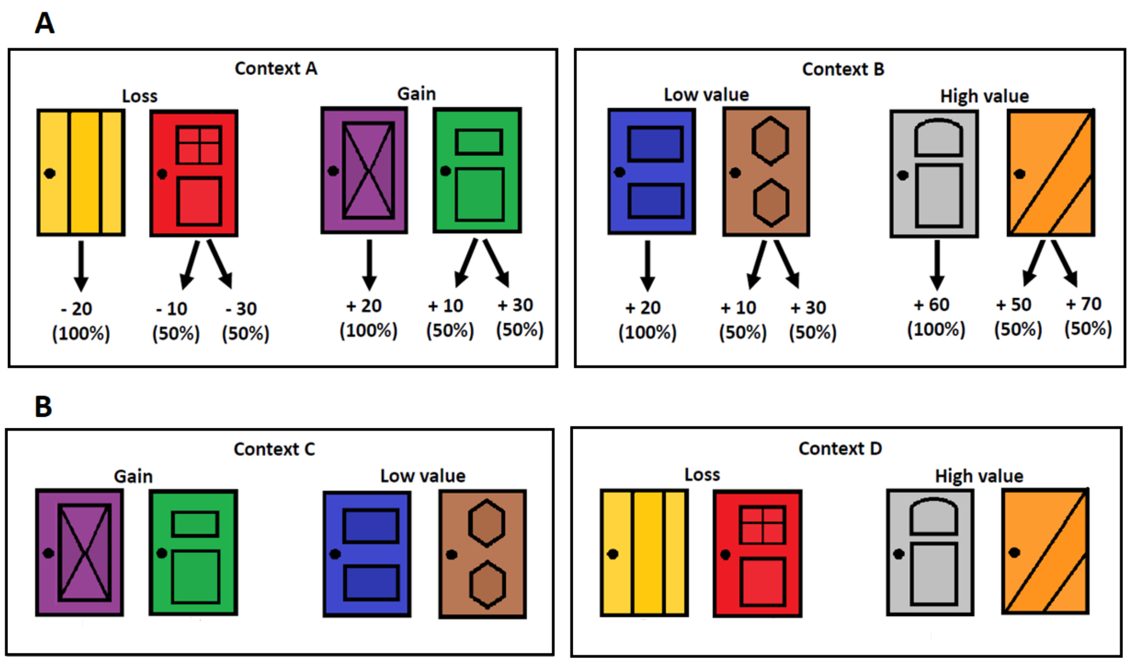

Figure S3. Choice stimuli used in Experiment S2. (A) Set of options in Context A (Gain/Loss) and Context B (High/Low) in the training phase. (B) Set of options in Context C and Context D in the test phase. The association between door and outcome was randomized across participants but remained constant within participants. Note that there were no distinct visual stimuli indicating the contexts apart from the doors/options themselves.

The eight visually distinct doors were randomly associated with the set of outcomes shown in Figure S1, but the assignment was constant for a given individual. As in Experiment 2, the training phase contained 6 blocks with alternating decision sets. The first two blocks of the training phase had 56 trials consisting of 32 *decision trials*, 16 *catch trials*, and 8 *single-choice trials*. The following 4 blocks of the training phase had 48 trials and had the same structure except that there were no s*ingle-choice trials.*

Training was followed by two blocks of test trials, each with 32 trials. In the test phase, the contexts were switched as illustrated in Figure S3 by changing which decisions were present in which context. Context C contained choices between the fixed gain (+20) and the risky gain (+10 or +30) and between the fixed low-value (+20), and the risky low-value (+10 or +30) options[1]. Context D contained choices between the fixed loss (-20) against the risky loss (-30 or -10) and between the fixed high-value (+60) and the risky high-value (+50 or +70) options. Each context consisted of one block of 32 trials (16 trials of each choice). Participants did not receive feedback after their selections in the test trials, but they were informed that the outcomes of their choices would continue to contribute to their accumulated bonuses.

After the choice task, participants completed two memory tests that were the same as in Experiments 1 and 2 with the following exceptions. For the recall test, they had to select a bullet option to indicated whether the outcome was positive or negative in addition to typing the value of the recalled outcome. An error message appeared if a bullet option was not selected or a non-numeric character was typed.

For the frequency-judgement test, each door image was shown together with a 3x3 matrix consisting of all outcomes from the experiment in ascending order. Each outcome value was paired with a blank space where participants reported their answers. Participants were instructed to type the judged percent frequency of each outcome for the displayed door image, and they were advised that all blank spaces would be considered as zero. The task only continued if the sum of their responses for a given door totaled to 100.

Hypotheses and Preregistered Data Analysis

As stated in the preregistration, the main hypotheses were:

[1] This design differed from the pre-registered plan which was to put the high-value and gain options in one context and the low-value and loss options in a second context (a full cross-over). Instead, the gain and low-value options (which have the same values) were placed in the same context. This altered design still allows testing of the core hypotheses, but is perhaps a less stringent test than initially planned. The hypotheses were adjusted slightly from the pre-registration to account for this shifted design, but the same, planned statistical tests were run.

1. The *Context-Replication Hypothesis* predicts that extreme outcomes in each context would be overweighted, leading to greater risky choice for the highest value options in each context. This hypothesis was assessed through one-tailed paired *t*-tests on risky choice for the higher value and lower value options during the final block of the training trials in each context. Risky choice for the target choice (+20 vs. +10/+30) in the two contexts was also compared. This hypothesis predicts more risk-seeking in Context A (where the target was the higher value choice in the set) than in Context B (where the target was the lower value choice in the set).

2. The *Encoding Hypothesis* predicts that context effects on choice are based on the encoding context of each option. As a result, the pattern of choice in the test phase should be the same as the pattern in the last block of the training phase for the same decision sets. This was assessed with the same 3 one-tailed paired *t*-tests on choices during the test phase.

3. The *Retrieval Hypothesis* supposes that context effects on choice are based on the retrieval context at the time of choice. As a result, the pattern of choice in the test phase (which presents options in a different context from training), should differ from the pattern of choice seen at the end of the training phase. Specifically, people should be less risk-seeking for the gain choice and more risk-seeking for the low-value choice during testing (with similar levels for both choices). This prediction was assessed with a two-way repeated-measures ANOVA (Choice [Gain vs. Low value] by Context [Training vs. Test]), with a predicted interaction between the two variables. Contrary to the encoding

hypothesis (above), the retrieval hypothesis predicts no reliable difference between gain and low-value choices, when they were both tested in the same context (i.e., Context C)[2].

4. The *Noise Hypothesis* supposes that a context shift will lead to more random choices. As a result, choice should shift toward indifference whenever doors are tested outside their training context (i.e., Context C and Context D). This was tested by calculating, for each participant, whether risky choice was closer to 50% in the novel context than at the end of training. A one-sample t-test was employed to test for reliable differences from 0%.

Results

Risky choice

Figure S4 shows risky choice for the higher and lower-value options in the four contexts. In the Gain/Loss context, participants were on average 8.7±6.4% [*M±95%C.I.*] more risk seeking for gains than for losses [$t(105)=2.65$, $p=.005$, $d=0.29$]. In the High/Low context, participants were 14.6±6.2% more risk-seeking for the high-value than for the low-value decision [$t(105)=4.60$, $p<.001$, $d=0.46$].

As in Exp 1 and Exp 2 in the main text, participants demonstrated significantly different risk preferences for the target choices in the two contexts. Participants were 9.8±6.1% more risk-seeking for the target choice in the Gain/Loss context than in the High/Low context [$t(105)=3.15$, $p=.001$, $d=0.33$], as highlighted in Figure S4. These results replicate the context effects seen in Exp. 1 and Exp. 2 and reveal that participants

[2] The pre-registration incorrectly states that the main effect of "Context" could be used to evaluate this hypothesis when this main effect actually indicates a shift in overall risk preference from training to test. It is the interaction between "Context" and "Choice" that could provide support for the Retrieval Hypothesis.

can discretize distinct decision contexts even without a background image to cue the context change.

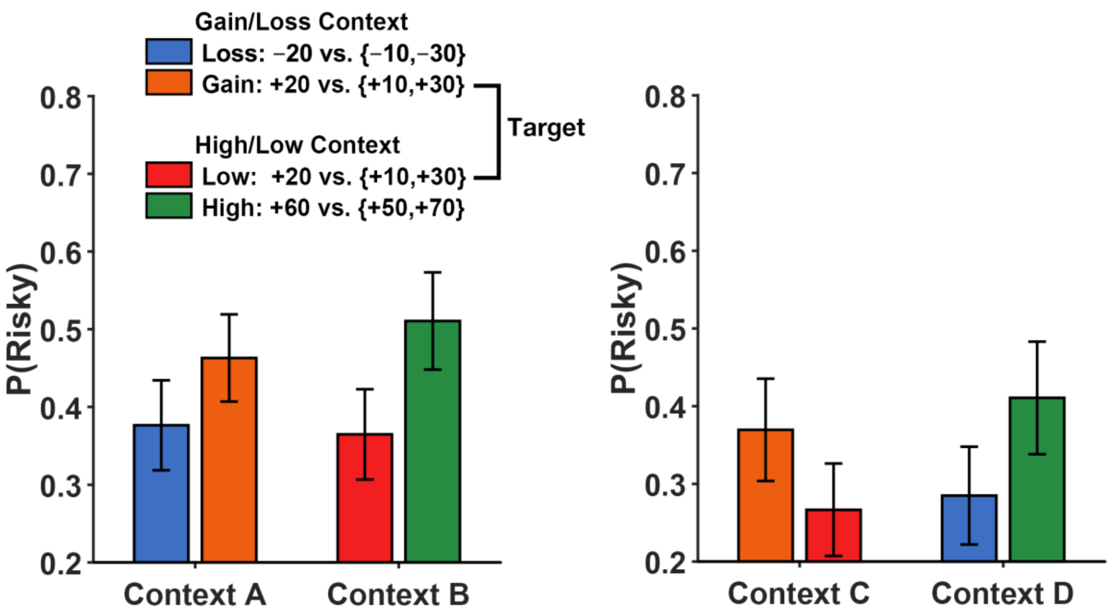

Figure S4. Proportion of risky choices for each decision set and their respective decision context, averaged across the last block in each context for Experiment S2. Error bars represent 95% confidence intervals.

Test blocks

During the test phase, the options were intermixed, and participants completed the choice task without feedback. Figure S4 shows how, for the choices trained in the Gain/Loss context, participants were still 8.5±7.9% more risk seeking for gains than for losses [$t(105)=2.14$, $p=.017$, $d=0.25$]. For the choices trained in the High/Low context, participants were still 14.4±7.6% more risk-seeking for the high-value gamble than for the low-value gamble [$t(105)=3.69$, $p<.001$, $d=0.46$]. The two target choices both appeared in Context C, yet people choose differently for each pairing despite their outcome equivalence. Similar to the training phase, they were 10.3±6.1% more risk-seeking for the target choice trained in the gain/loss context than for the target choice trained in the high/low value context [$t(105)=3.32$, $p<.001$, $d=0.31$]. These results support the Encoding Hypothesis and are inconsistent with the Retrieval Hypothesis. A two-way ANOVA confirmed a main effect of Choice [$F(1,105)=14.3$, $p<.001$, $\eta_p^2=.12$], an effect of Context [$F(1,105)=34.6$, $p<.001$, $\eta_p^2=.25$] whereby people were more risk averse overall during the test context, and no interaction [$F(1,105)=0.015$, $p=.90$, $\eta_p^2=.00$]. These data support the notion that the encoding context is more important than the retrieval context in determining choice.

We further examined if participants' average risky choices in the test phase tended towards indifference. At the individual-choice level, for the gain target choices, people were on average of 8.4±3.2% *further* from indifference during the test [$t(105)=4.01$, $p<.001$, $d=0.39$], and for the low-value target choices, people were 6.6±3.2% further from indifference [$t(105)=5.22$, $p<.001$, $d=0.51$]. These results firmly invalidate the noise hypothesis.

Memory tests

Error! Reference source not found. shows the average responses for the first outcome reported for each option. The results show context-dependent overweighting of the extreme loss and low-value outcomes. In the first-outcome-reported test, for both the loss and low-value options, participants were significantly more likely to report the worse value (−30 and +10, respectively) [Loss: $\chi^2(1,N=93)=23.75$, $p<.001$; Low: $\chi^2(1,N=77)=19.75$, $p<.001$]. The difference in reporting of outcomes was not significant for the gain option [$\chi^2(1,N=79)=0.013$, $p=.91$], but for the high-value option, the upper extreme (+70) was moderately more likely to be reported than the non-extreme high-value outcome (+50) [$\chi^2(1, N=98)=4.94$, $p=.03$].

Error! Reference source not found. also illustrates participants' mean judged frequencies for each option. The frequency-judgment test also showed a context-dependent bias in which people reported higher percentages for the worse outcome for the loss and low-value options [Loss: $t(105)=5.64$, $p<.001$, $d=0.92$; Low: $t(105)=6.72$, $p<.001$, $d=1.04$]. For the gain option, the lower value was judged as more frequent [$t(105)=3.36$, $p=.001$, $d=0.47$], but there was no reliable difference in judged percent for the outcomes of the high-value option [$t(105)=1.43$, $p=.155$, $d=0.24$].

Thus, by both memory measures, participants showed consistent biases toward the extreme lower-value outcomes experienced within each training context, but they did not show consistent biases toward the extreme higher-value outcomes.

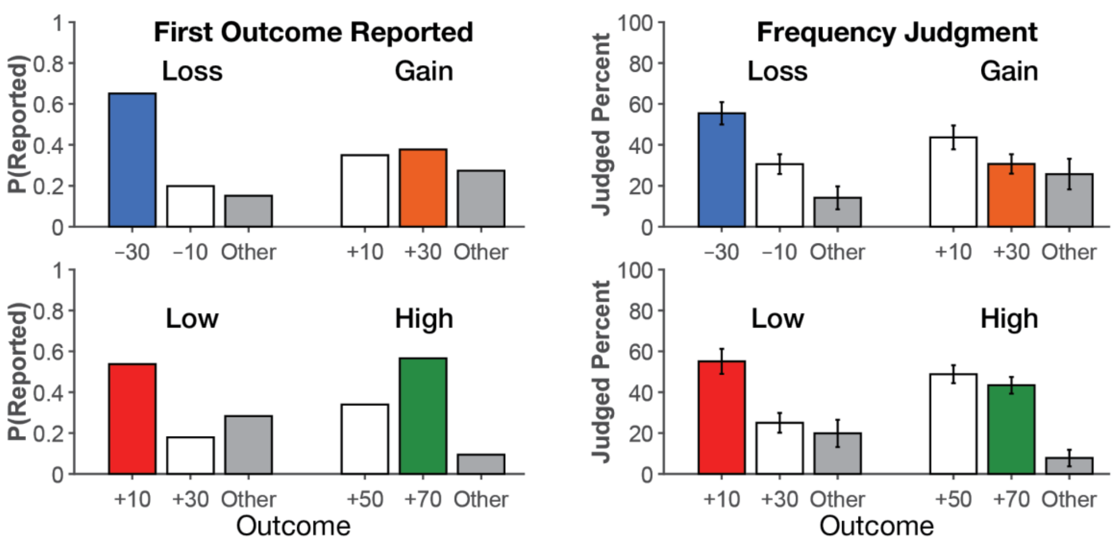

Figure S5. Results of the two memory tests for each decision context in Experiment S2. Participants were more likely to report the extreme outcomes first and judged the lowest outcome in each context as having occurred most frequently. Coloured bars are local extreme outcomes, and white bars are non-extremes. The colour code matches the conditions in previous figures (Blue = Loss; Orange = Gain; Red = Low; Green = High). Error bars represent 95% confidence intervals.

Discussion

The choice results show that options providing the same outcomes acquired different values depending on their grouping with other outcomes and that visually distinct backgrounds are not necessary for this context effect. Participants were more risk-seeking for the higher-value options from each context, and they were more risk-seeking for the target choice when it was presented with losses (Context A) than when that same target choice was presented with higher value outcomes (Context B). The recall tests also provided some support for the difference in perception of the same choice in different contexts. In particular, for the risky target option, participants were more likely to recall the +10 outcome than the +30 outcome for doors trained in Context B (where +10 was the lowest outcome in the decision set), whereas they did not show higher recall of +10 than +30 for doors trained in Context A (where it was not an extreme outcome).

The context shift test results support the conclusions of Experiment 2 that context alters choice by influencing the encoding process. During tests in which the arrangement of the options changed and feedback was unavailable, participants continued to be more risk-seeking for target options trained in contexts with other lower-value outcomes. If the effects were due to retrieval, participants would have had equal risk preferences for options with the same outcomes (+20 versus +10/+30) during the test phase because they would have remembered the outcomes from each option according to the context during retrieval.