Combinations of low-level and high-level neural processes can account for distinct patterns of context-dependent choice

Abbreviated title: Modulation of context effects by risk attitudes

Mehran Spitmaan[1*], Oihane Horno[2*], Emily Chu[1], Alireza Soltani[1]

[1] *Department of Psychological and Brain Sciences, Dartmouth College*
[2] *Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal*

Correspondence: A.S., Department of Psychological and Brain Sciences, HB 6207, Dartmouth College, Hanover, NH, 03755 (Email: soltani@dartmouth.edu)

Manuscript information: 7 Figures; 1 Table; 5 Supplementary Figures; Total 37 pages
Abstract 246 words
Significant statement 114 words

Keywords: decision making, risk aversion, neural adaptation, decoy effects, attention

Conflict of interest: The authors declare no competing interests.

Acknowledgments: We thank Daeyeol Lee and Chanc VanWinkle Orzell for helpful comments on the manuscript.

Abstract

Context effects have been explained by either high-level cognitive processes or low-level neural adjustments but not their combination. It is currently unclear how these processes interact to shape individuals' responses to context. Here, we used a large cohort of human subjects in experiments involving choice between two or three gambles in order to study the dependence of context effects on neural adaptation and individuals' risk attitudes. We found no evidence that neural adaptation on long timescales (~100 trials) contributes to context effects. However, we identified two groups of subjects with distinct patterns of responses to decoys, both of which depended on individuals' risk aversion. Subjects in the first group exhibited strong, consistent decoy effects and became more risk averse due to decoy presentation. In contrast, subjects in the second group did not show consistent decoy effects and became more risk seeking. The degree of change in risk aversion due to decoy presentation was positively correlated with the initial degrees of risk aversion. To explain these results and reveal underlying neural mechanisms, we developed a new model that incorporates both low- and high-level processes to fit individuals' choice behavior. We found that observed decoy effects can be explained by a combination of adjustments in neural representations and competitive weighting of reward attributes, both of which depend on risk aversion but in opposite directions. Altogether, our results demonstrate how a combination of low- and high-level processes shapes multi-attribute choice, modulates overall risk preference, and explains distinct behavioral phenotypes.

Significance statement

A large body of experimental work has illustrated that the introduction of a new, and often irrelevant, option can influence preference among the existing options, a phenomenon referred to as context or decoy effects. Although context effects have been explained by high-level cognitive processes—such as comparisons and competitions between attributes—or low-level adjustments of neural representations, it is unclear how these processes interact to shape individuals' responses to context. Here, we show that both high-level cognitive processes and low-level neural adjustments shift risk preference during choice between multiple options but in opposite directions. Moreover, we demonstrate that a combination of these processes can account for distinct patterns of context effects in human subjects.

Introduction

Despite the prevalent use of two-alternative choice paradigms to study value-based choice, real-life decisions often involve selecting among multiple options. The set of options, even those irrelevant to the decision maker, can strongly influence preference and alter choice processes (Tversky, 1972; Huber et al., 1982; Simonson and Tversky, 1992; Pratkanis and Farquhar, 1992; Bettman et al., 1998; Dhar and Simonson, 2003). For example, introducing a new option that is worse than one of the two existing options in all its attributes but better than the alternative option in some attributes (asymmetrically dominated option) can increase the preference for the former option (Huber et al., 1982; Dhar and Simonson, 2003; Huber and Puto, 1983; Simonson 1989). The dependence of preference on the choice set, often referred to as context or decoy effects, has been extensively studied in value-based choice and has revealed important aspects of valuation and decision processes (Bettman et al., 1998; Kahneman and Tversky, 1984; Ratneshwar et al., 1987; Payne et al., 1988; Dhar and Glazer, 1996; Wedell and Pettibone, 1996; Tsetsos et al., 2010; Trueblood, 2012).

Because of their complexity, decoy effects often have been explained by changes in subjective value due to various high-level cognitive processes, such as attentional switching to different choice attributes, menu-dependent evaluation of choice attributes, and competition between attribute processing to enhance contrast between certain attributes (Tversky, 1972; Tsetsos et al., 2010; Tversky and Simonson, 1993; Roe et al., 2001; Usher and McClelland, 2004; Johnson and Busemeyer, 2005; Usher et al., 2008; Trueblood et al., 2014). In contrast, aiming to address underlying neural mechanisms, others have attributed decoy effects to low-level adjustments or adaptation of neural representations on a short timescale to the set of options (Soltani et al., 2012; Louie et al., 2013). However, none of the current models of context-dependent choice includes both high-level and low-level processes. Furthermore, despite strong evidence for adaptation of neural response to the set of options over a block of trials or an experimental session (Louie et al., 2011; Padoa-Schioppa and Assad, 2008; Padoa-Schioppa, 2009), it is currently unknown whether this adaptation contributes to context effects.

In addition, although context effects mainly have been measured in a between-subjects design (Heath and Chatterjee, 1995; Pettibone and Wedell, 2000; Dhar and Sherman, 1996), there is a

large variability in how such effects influence different individuals (Huber et al., 1982; Huber and Puto, 1983; Soltani et al., 2012; Pan and Lehmann, 1993; Pettibone and Wedell, 2007; Ha et al., 2009) that could reflect additional mechanisms involved in the valuation of multi-attribute options. For example, during choice between multiple gambles, the presence of a few gambles with large reward magnitudes could bias the decision maker to more strongly process reward magnitude at the expense of reward probability, ultimately resulting in more risk-seeking behavior. The influence of choice set on this selective processing of information, however, could depend on each individual's degree of risk aversion. Currently, little is known about the dependence of context effects on risk attitudes.

Here, we measured risk preference and decoy effects in a large cohort of subjects to study how context effects depend on an individual's degree of risk aversion and on adaptation to the range of reward probabilities and magnitudes. Risk aversion was measured using binary choice between pairs of monetary gambles. Decoy effects were measured using a phantom-decoy design; subjects were initially presented with three monetary gambles to evaluate, one of which was subsequently removed at the time of choice. To examine the contribution of neural adaptation, we introduced a new set of gambles in 20% of trials to interrupt possible ongoing adaptation to the range of reward probabilities and magnitudes in the main set of gambles. Finally, we developed a new model of context-dependent choice that incorporates both low- and high-level processes to explain our results and reveal underlying neural mechanisms.

Results

Overall decoy effects. Each subject performed two experiments in which they evaluated and selected between two or three monetary gambles. In the first experiment, subjects selected between two gambles (estimation task; see Materials and Methods for more details). Choice behavior in this experiment was used to estimate individuals' degree of risk aversion and to tailor gambles for each subject in the second experiment. In the second experiment, subjects selected between three gambles under two different conditions (decoy task; **Figure 1**). Choice behavior in this experiment was used to study context effects.

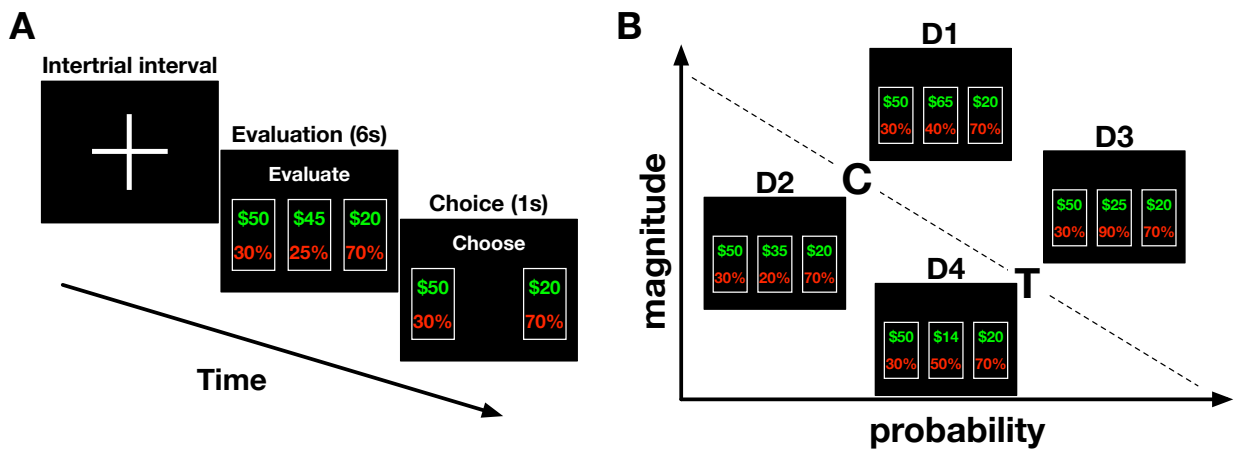


Figure 1. Experimental design. (**A**) Timeline of the decoy task used to study context effects. After the inter-trial interval and fixation cross, three monetary gambles were presented on the screen. The subjects had 6 seconds to evaluate the gambles. Subsequently, one of the three gambles was randomly removed, and the subjects had 1 second to choose between the remaining two gambles. The subjects made a choice by pressing the left or right arrow key; no feedback was provided following choice. (**B**) Four possible decoy positions relative to the target (T) and competitor (C) gambles are shown. For illustration purposes, decoy gambles are shown in the middle of the three gambles, but their arrangement was randomly determined on each trial of the actual experiments. D1 (D3) decoys, referred to as asymmetrically dominant decoys, were greater in both attributes than the competitor (target), whereas D2 (D4) decoys (asymmetrically dominated decoys) were worse than the competitor (target).

During the estimation task, subjects selected between two monetary gambles in each trial. One of the two gambles was always a fixed low-risk gamble: a gamble with a reward probability (p_R) equal to 0.7 and a reward magnitude of $20. The other gamble was selected from a set of high-risk gambles with p_R = 0.3 and different reward magnitudes, *M*. As expected, subjects chose the high-risk gambles over the low-risk gamble more often as *M* increased (**Supplementary Figure 1A**). By fitting choice behavior with a logistic function of the difference between the two gambles' magnitudes in a given trial, we estimated the indifference point for each subject. The indifference point was defined as the magnitude of a high-risk gamble that was as equally preferable as the low-risk gamble (see Materials and Methods). Thus, it reflects the initial degree of risk aversion for a given subject; a larger indifference point corresponds to increased risk aversion. Overall, we observed large variability for the indifference points across subjects (**Supplementary Figure 1B**).

During the decoy task, subjects were initially presented with three monetary gambles to evaluate, one of which was subsequently removed at the time of choice (**Figure 1A**; see decoy task in Materials and Methods). This phantom-decoy design allowed us to measure the influence of the removed gamble (the decoy) on the preference between the remaining gambles. Two out of the three gambles, which we refer to as the target (T) and competitor (C), were tailored for each subject using his/her indifference point estimated from the estimation task. The decoys were positioned in the attribute space relative to the target and competitor gambles in one of the four possible positions with some jitters (**Figure 1B**; see Materials and Methods). In the control condition of the decoy task, jittered versions of the target, competitor, and decoy gambles were presented throughout the experiment. In the range-manipulation condition, however, we introduced a new set of gambles with large probabilities and magnitudes in 20% of trials in order to interrupt possible adaptation to the range of reward probabilities and magnitudes in the main set of gambles.

Consistent with our previous results (Soltani et al., 2012), we found significant decoy effects for all decoy locations except D2 in the control condition (**Figure 2A–B**). Asymmetrically dominant decoys decreased the preference for the gamble next to them in the attribute space (competitor for D1 and target for D3; two-sided Wilcoxon signed-test; D1: $p = 0.023$, $d = 0.72$; D3: $p = 0.042$, $d = -0.73$). In contrast, asymmetrically dominated decoys increased the preference for the gamble next to them (competitor for D2 and target for D4), but this effect was only significant for D4 (two-sided Wilcoxon signed-test; D2: $p = 0.17$, $d = -0.35$; D4: $p = 0.030$, $d = 0.59$). This phenomenon is often referred to as the attraction effect (Huber et al., 1982). Additionally, we measured the effect of decoys on preference using decoy "efficacy," which quantifies the tendency to choose the target due to the presentation of decoys at a given location after subtracting the overall tendency to choose the target (Materials and Methods). Finally, the decoy efficacies for the dominant and dominated decoys were anti-correlated (Pearson correlation; D1 and D3, $r = -0.39$, $p = 0.039$; D2 and D4, $p = -0.33$, $p = 0.027$), indicating that similar mechanisms underlie their generation.

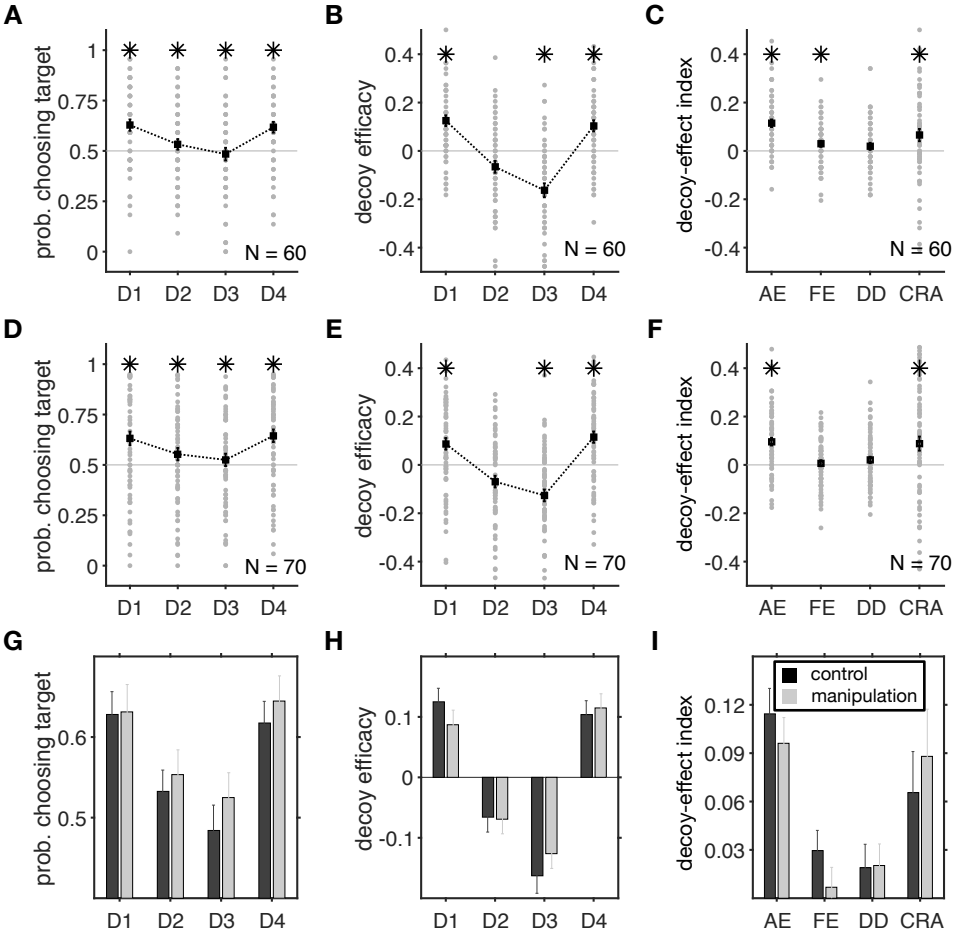


Figure 2. Preference was similarly influenced by decoys in the control and range-manipulation conditions. (**A**) Probability of selecting the target for different decoy locations during the control condition. Each gray circle shows the average probability that an individual subject selected the target for a given decoy location, and black squares indicate the average across all subjects. Error bars show the s.e.m., and an asterisk shows that the median of choice probability across subjects for a given decoy location is significantly different from 0.5 (two-sided Wilcoxon signed-test, $p < 0.05$). (**B**) Decoy efficacies measure the tendency to choose the target due to the presentation of decoys at a given location after subtracting the overall tendency to choose the target. An asterisk shows that the median of a given decoy efficacy across subjects is significantly different from zero (two-sided Wilcoxon signed-test, $p < 0.05$). Presentation of decoys resulted in change in preference in all locations. (**C**) Plot shows four measures for quantifying different effects of decoys on preference (AE: attraction effect; FE: frequency effect; DD: dominant vs. dominated; and CRA: change in risk aversion). Other conventions are similar to those in panel B. (**D–F**) The same as in A–C but for the range-manipulation condition. (**G–I**) Comparisons between the average target probabilities (G), decoy efficacies (H), and our four indices (I) in the two experimental conditions. Overall, there were no significant differences between the two experimental conditions based on choice probability, decoy efficacies, or any of our four measures (two-sided Wilcoxon ranksum test, $p > 0.05$).

In order to fully characterize the change in preference due to the introduction of a decoy to the choice set, we defined four "decoy-effect" indices that measure different aspects of decoy effects. Briefly, the overall attraction effect (AE) measures the average relative decoy effects over all locations. The overall frequency effect (FE) quantifies the overall tendency to choose the gamble next to the decoy in terms of both reward probability and magnitude (i.e., the competitor for D1 and D2 decoys, and the target for D3 and D4 decoys). The dominant vs. dominated (DD) measures the overall difference between the effects of dominant and dominated decoys. The fourth index quantifies the overall change in risk aversion (CRA) due to the decoy presentation (see Materials and Methods for more details).

Calculating these indices for each subject, we found a significant overall attraction effect (two-sided Wilcoxon signed-test, $p = 0.011$, $d = 0.94$; **Figure 2C**). We also found a small but significant frequency effect (two-sided Wilcoxon signed-test, $p = 0.046$, $d = 0.31$), indicating an overall increase in the selection of the gamble next to the decoys. There was no significant difference in the effects of dominant and dominated decoys as measured by the DD (two-sided Wilcoxon signed-test, $p = 0.22$, $d = 0.17$). Finally, because of the symmetry in decoy presentation as well as the tailoring of the target and competitor gambles for individual subjects, each subject should choose equally between the target and competitor in absence of any decoy effects. Therefore, any overall bias toward either gamble could be attributed to changes in risk preference due to the presentation of decoys, which is quantified by the CRA. Interestingly, we found a significant increase in risk aversion due to decoys across all subjects (two-sided Wilcoxon signed-test, $p = 0.035$, $d = 0.33$).

Contribution of neural adaptation to decoy effects. In order to test the contribution of neural adaptation to context effects, we next examined choices in the range-manipulation condition, in which we introduced a new set of gambles with large probabilities and magnitudes in 20% of trials (see Materials and Methods). These additional gambles were intended to interrupt possible adaptation to the range of reward probabilities and magnitudes in the main set of gambles and were not included in the analysis.

Despite the large number of subjects in our experiment, we did not find any significant difference between subjects' behavior in the two experimental conditions with regards to the

average decoy efficacies (two-sided Wilcoxon ranksum test; D1: $p = 0.51$; D2: $p = 0.92$; D3: $p = 0.35$; D4: $p = 0.68$; **Figure 2H**) or decoy-effect indices (two-sided Wilcoxon ranksum test; AE: $p = 0.31$; FE: $p = 0.11$; DD: $p = 0.42$; CRA: $p = 0.26$; **Figure 2I**). Similar to the control condition, subjects in the range-manipulation condition exhibited significant decoy effects for D1, D3, and D4 decoys (two-sided Wilcoxon signed-test; $p = 0.039, 0.33, 0.031, 0.028$; $d = 0.43, -0.34, -0.62, 0.59$, for D1 to D4, respectively; **Figure 2E**) and a significant overall attraction effect and change in risk aversion due to decoy presentation (two-sided Wilcoxon signed-test; AE: $p = 0.014$, $d = 0.72$; FE: $p = 0.61$, $d = 0.07$; DD: $p = 0.73$, $d = 0.18$; CRA: $p = 0.044$, $d = 0.36$; **Figure 2F**).

We also examined the time course of decoy-effect indices in the two experimental conditions by computing the running average of AE and CRA over time, using a sliding window with a length of 40 trials and steps of 20 trials (**Figure 3**). However, we did not observe any significant changes in any of the decoy-effect indices over time (two-sided t-test; AE (control), $p = 0.96$; AE (range-manipulation), $p = 0.75$; CRA (control), $p = 0.81$; CRA (range-manipulation), $p = 0.46$). Finally, there was no significant differences in the slopes of the AE or CRA as a function of time (within a session) between the two experimental conditions (AE (control), $b = 0.0004$, 95% CI [-0.0146, 0.0154]; AE (range-manipulation), $b = -0.0027$, 95% CI [-0.0191, 0.0137]; CRA (control), $b = 0.0021$, 95% CI [-0.0149, 0.0191]; CRA (range-manipulation), $b = 0.0052$, 95% CI [-0.0144, 0.0248]). Together, we did not find any evidence of changes in decoy effects due to range manipulation, which was intended to interrupt possible adaptation to the range of reward probabilities and magnitudes. Therefore, our experiments did not reveal any evidence that neural adaptation on the order of ~100 trials—present in the control but not in the range-manipulation condition—contributes to context effects.

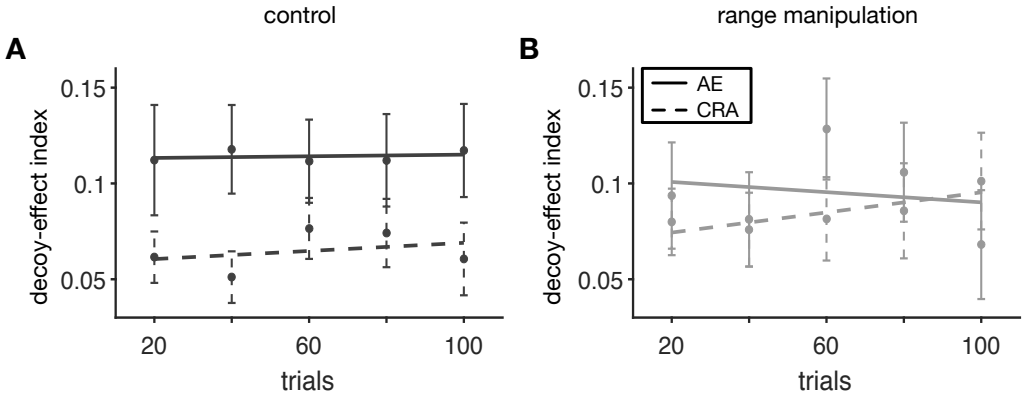


Figure 3. Time course of the overall attraction effect (AE) and the overall change in risk aversion (CRA) in the control (**A**) and range-manipulation conditions (**B**). Each point shows the average value of the respective measure during 40-trial bins around that point across all subjects. The error bar shows the s.e.m. The solid and dashed lines show the regression lines. Moreover, there was no significant difference between the control and range-manipulation conditions.

Dependence of decoy effects on individuals' risk attitudes. We used a strict clustering method to isolate distinct patterns of decoy effects in our large cohort of subjects. We adopted the k-means clustering method for different sets of features and numbers of clusters. The sets of features included all the possible exclusive combinations of four decoy-effect indices and the cluster sizes ranged from 2 to 5. We found the best clustering (based on silhouette values) is achieved using all decoy-effect indices and two clusters (**Supplementary Figure 2**). We used the clusters with the best separation to divide the subjects into two groups (Group 1: N = 72, and Group 2: N = 58).

We found very different patterns of responses to the presentation of decoys in the two groups of subjects (**Figure 4**). More specifically, subjects in Group 1 exhibited an overall increase in risk aversion during the decoy task by selecting the target with a probability larger than 0.5 (over all decoy locations), whereas subjects in the Group 2 showed a decrease in risk aversion by choosing the target with a probability smaller than 0.5 (**Figure 4A–B**). In addition, subjects in the first group showed strong and consistent (i.e. across all decoy types) decoy effects (two-sided Wilcoxon ranksum test; D1: $p = 0.001$, $d = 0.99$; D2: $p = 0.008$, $d = -0.62$; D3: $p = 0.0001$, $d = -0.67$; D4: $p = 0.0003$, $d = 0.60$; **Figure 4C**), whereas decoy effects in the second group were inconsistent and limited to decoys next to the less risky gamble (two-sided Wilcoxon ranksum

test; D1: $p = 0.42$, $d = 0.18$; D2: $p = 0.79$, $d = -0.03$; D3: $p = 0.0001$, $d = -0.67$; D4: $p = 0.0022$, $d = 0.58$; **Figure 4D**).

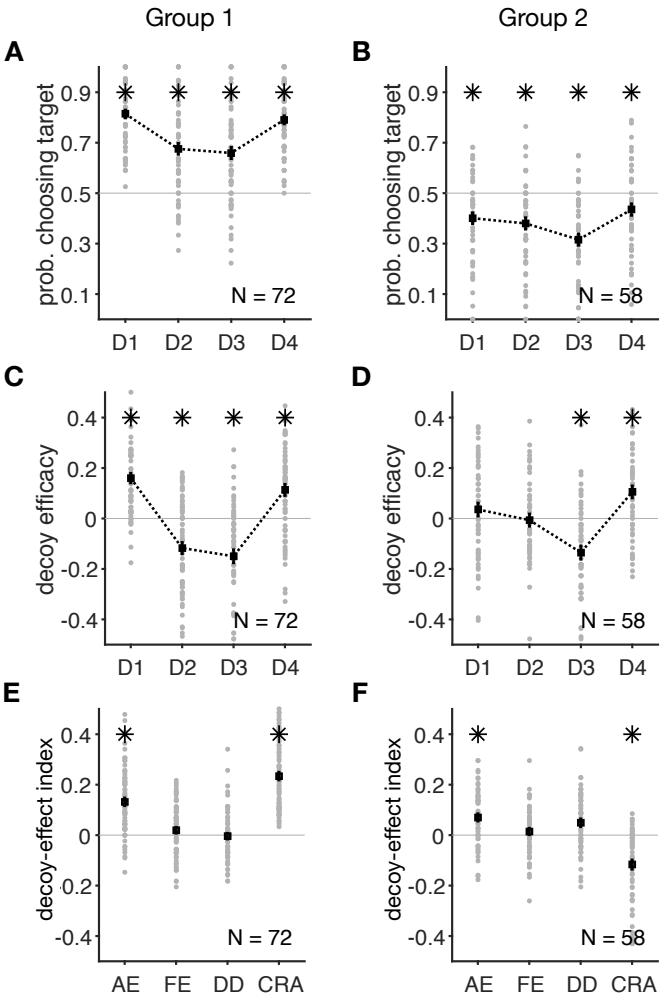


Figure 4. Distinct decoy effects in two groups of subjects identified by clustering based on all decoy-effect indices. (**A–B**) Probability of selecting the target for different decoy locations in the two groups of subjects. Conventions are the same as in **Figure 2**. Subjects in the first group ($N = 72$) selected the target with a probability larger than 0.5 for all decoy locations (increased risk aversion), whereas subjects in the second group ($N = 58$) selected the target with a probability smaller than 0.5 (decreased risk aversion). An asterisk shows that the median of choice probability across subjects for a given decoy location is significantly different from 0.5 (two-sided Wilcoxon signed-test, $p < 0.05$). (**C–D**) Decoy efficacies in the two groups of subjects. Subjects in Group 1 exhibited strong, consistent decoy effects (C), whereas the decoy effects were inconsistent and limited to decoys next to the less risky gamble (target) in Group 2 (D). (**E–F**) Plot shows decoy-effect indices for individuals in the two groups of subjects. The first group showed a strong attraction effect and an overall increase in risk aversion. The second group showed a significant attraction effect and a decrease in risk aversion.

Finally, both groups showed positive attraction effects (two-sided Wilcoxon signed-test; Group 1: $p = 0.026$, $d = 1.00$; Group 2: $p = 0.031$, $d = 0.62$), but there were no significant frequency effects (two-sided Wilcoxon signed-test; Group 1: $p = 0.71$, $d = 0.19$; Group 2: $p = 0.51$, $d = 0.15$) or differences between dominant and dominated decoys (two-sided Wilcoxon signed-test; Group 1: $p = 0.64$, $d = -0.04$; Group 2; $p = 0.33$, $d = 0.41$). Importantly, subjects in Group 1 exhibited a significant positive change in risk aversion (CRA; two-sided Wilcoxon signed-test, $p = 0.011$, $d = 1.77$; **Figure 4E**), whereas this change was significantly negative for Group 2 (two-sided Wilcoxon signed-test, $p = 0.042$, $d = -0.82$; **Figure 4F**). In other words, Group 1 became more risk aversive due to decoy presentation while Group 2 became more risk seeking.

To further reveal the differences in decoy effects between the two groups, we directly compared the four decoy-effect indices between the two groups of subjects. We found a larger overall attraction effect in Group 1 (two-sided Wilcoxon ranksum test, $p = 0.021$, $d = 0.50$; **Figure 5A**), reflecting more consistent decoy effects in this group of subjects. In contrast, the DD was slightly but significantly larger in Group 2 (two-sided Wilcoxon ranksum test, $p = 0.037$, $d = -0.48$), but this effect was not significant in either group (**Figure 4E-F**). Also, there was no difference between the two groups in terms of the overall frequency effect (two-sided Wilcoxon ranksum test, $p = 0.49$, $d = 0.05$). Finally, the change in risk aversion was much larger in Group 1 than Group 2 (two-sided Wilcoxon ranksum test, $p = 0.0007$, $d = 2.55$).

To test the relationship between decoy effects and risk attitudes, we examined the initial risk aversion in the two groups of subjects identified in the decoy task using the indifference points from the estimation task (note that indifference points from the estimation task were not used for clustering of decoy effects). Interestingly, we found that indifference points of subjects in Group 1 were significantly larger than what was predicted by expected value ($46.70), and thus, these subjects were risk averse (two-sided Wilcoxon signed-rank test; $p = 3.2\times10^{-6}$, $d = 0.65$; **Figure 5B**). In contrast, there was no evidence to support that the indifference points of subjects in Group 2 were different from risk neutral (two-sided Wilcoxon signed-rank test; $p = 0.17$, $d = 0.23$). In addition, the indifference points for Group 1 were significantly larger than Group 2 (two-sided Wilcoxon ranksum test, $p = 0.0013$, $d = 0.59$). These results were completely unexpected because we had not considered indifference points from the estimation task for our clustering.

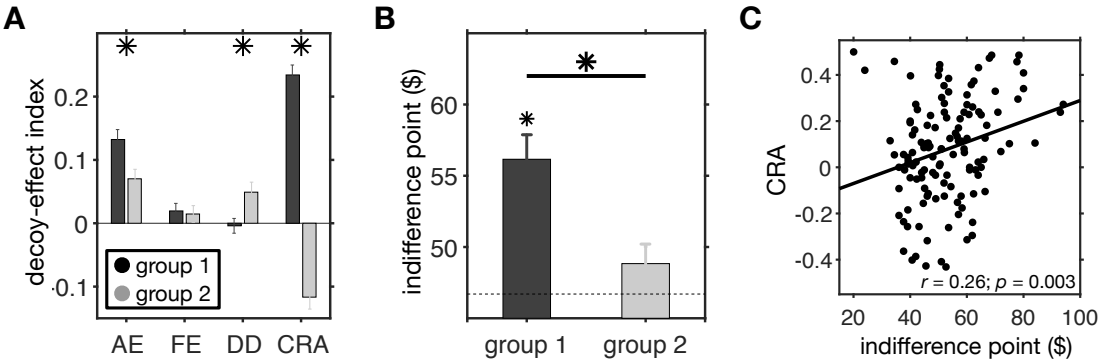


Figure 5. Comparison of the average decoy effects and initial risk preference between the two groups of subjects identified by clustering based on all four decoy-effect indices. (**A**) Plot shows the mean (±std) of each measure separately for the two groups of subjects. An asterisk indicates that the difference between a given measure for the two groups is significant (two-sided Wilcoxon ranksum test, $p < 0.05$). The two groups were different in terms of the overall attraction effect and the overall change in risk preference. (**B**) Comparison between the indifference points for the two groups of subjects. Plot shows the mean (±std) of the indifference point separately for the two groups of subjects and the dashed line indicates an indifference point of $46.70, corresponding to risk-neutrality. The asterisk above the bar shows a significant difference from $46.70 (two-sided Wilcoxon signed-rank test, $p < 0.05$), and the asterisk above the horizontal line indicates a significant difference between the two groups (two-sided Wilcoxon ranksum test, $p < 0.05$). (**C**) Plotted is the change in risk aversion in the decoy task as a function of the indifference point in the estimation task. There was a significant correlation between these quantities indicating that individuals' risk aversion in the binary choice influences their response to decoys.

Finally, considering that the most contrasting differences between the two groups were the change in risk aversion and the initial indifference points, we examined the relationship between these two quantities within individual subjects. Interestingly, we found a significant positive correlation between the CRA and the indifference point (Pearson correlation; $r = 0.26$, $p = 0.003$; **Figure 5C**), such that subjects with larger indifference points exhibited larger changes in risk aversion due to the decoy presentation. This result indicates that subjects who were more risk averse during binary choice became even more risk averse when considering three gambles.

New computational model to reveal the underlying mechanisms. In order to explain the observed pattern of choice behavior, we extended our range-normalization model of context-dependent choice to fit choice behavior of subjects in our experiments (**Supplementary Figure 3**). We compared 16 alternative models that differ in terms of how representations of values are modulated by risk aversion and competitive weighting to fit the choice behavior of subjects (in terms of decoy-effect indices) in each group in order to determine the best model using cross validation (see Materials and Methods for more details). We then used the best model parameters

to examine relationships between risk preference, adjustments in representations of reward attributes (probability and magnitude), and competitive weighting of these two attributes.

We found that the model with similar location-dependent biases for decoys next to the target and competitor (Eq. 10) and separate representation factors (f_p and f_m) provided the best fit to subjects' data (**Table 1**). The fit based on this model closely captured the observed pattern of decoy effects in the two groups of subjects (**Figure 6**). In contrast, the models with only low-level adjustments of neural representations (**Supplementary Figure 4**) or high-level cognitive adjustments (**Supplementary Figure 5**) failed to capture all aspects of our data.

Number of parameters	Model parameters	Cross-validation prediction error
7	$f_p, f_m, b_1, b_2, b_3, b_4, \sigma$	0.3716
5	$b_1, b_2, b_3, b_4, \sigma$	0.0456
7	$f_{p12}, f_{m12}, f_{p34}, f_{m34}, b_{12}, b_{34}, \sigma$	0.0371
5	$f_p, f_m, b_{12}, b_{34}, \sigma$	0.0344
5	$f_p, f_m, b_{13}, b_{24}, \sigma$	0.0433
5	$f_p, f_m, b_{14}, b_{23}, \sigma$	0.0476
6	$f_p, f_m, b_{12}, b_{34}, b_0, \sigma$	0.0359
6	$f_p, f_m, b_{13}, b_{24}, b_0, \sigma$	0.0472
6	$f_p, f_m, b_{14}, b_{23}, b_0, \sigma$	0.0493
5	$b_0, f_0, b_{12}, b_{34}, \sigma$	0.0521
4	$f_0, b_{12}, b_{34}, \sigma$	0.0538
4	b_0, f_p, f_m, σ	0.4007
4	$b_0, b_{12}, b_{34}, \sigma$	0.0402
3	f_p, f_m, σ	0.0382
3	b_{12}, b_{34}, σ	0.0459
3	b_0, f_0, σ	0.0502
2	b_0, σ	0.0461

Table 1. Comparison of different models' abilities in capturing each group of subjects' patterns of decoy efficacies based on cross-validation. Reported are cross-validation prediction errors (i.e., the absolute difference between the predicted and actual) for each model with different sets of parameters. The orange shading indicates the best model. As parameters, w_0 indicates the fixed bias, w_C and w_T show competitor and target biases, $w_k; k = \{1, 2, 3, 4\}$ stands for independent bias assigned for decoy k, f_0 is the single neural representation factor, and f_p and f_m are independent neural representation factors for probability and magnitude, respectively.

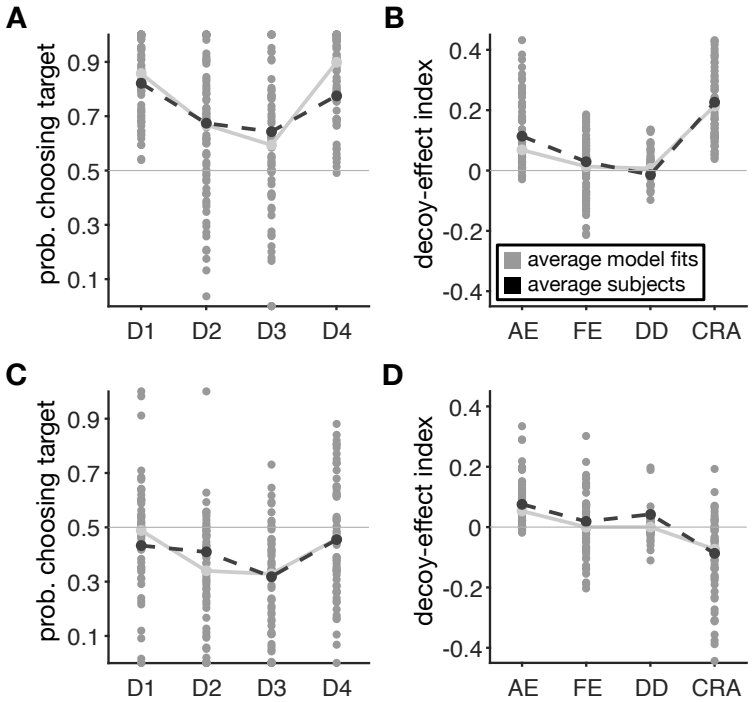


Figure 6. The model with both adjustments of value representation and competitive weighting of reward attributes can capture the experimental data. (**A, C**) Each gray circle shows the predicted probability of selecting the target for different decoy locations for individual subjects in the first (A) and second (C) groups using the best overall model. For comparison, the average values across subjects and their predicted values are plotted as well. (**B, D**) Plots show predicted decoy-effect indices for individual subjects in the first (B) and second (D) groups of subjects using the best overall model (AE: attraction effect; FE: frequency effect; DD: dominant vs. dominated; and CRA: change in risk aversion).

Considering that the best model included both low-level and high-level neural processes, we then investigated the individual effects of adjustments of neural representation and competitive weighting of reward attributes in order to explain the observed pattern of CRA in each group of subjects. First, we calculated the difference in CRA between the models with and without adjustments of neural representation to decoy presentation. We found that for subjects in Group 1, the overall change in CRA due to inclusion of low-level neural adjustments was significantly larger than zero (two-sided Wilcoxon signed-rank test; median = 0.17, $p = 6.6 \times 10^{-12}$; **Figure 7A**). In contrast, for subjects in Group 2, the change in CRA due to inclusion of low-level neural adjustments was significantly smaller than zero (two-sided Wilcoxon signed-rank test; median =

-0.06, p = 0.047). These results are consistent with the observed pattern of CRA in the two groups, indicating that adjustments of neural representation to decoy presentation can account for most of the observed changes in risk preference (increase and decrease in risk aversion in Groups 1 and 2, respectively).

We also examined the relationship between the estimated representation factors (f_p and f_m) based on the best model fit and individuals' degree of risk aversion (indifference points). The representation factor for a given reward attribute determines how that reward attribute is represented and adjusts to decoy presentation. We found opposite relationships between indifference points and estimated representation factors for probability and magnitude, reflecting competitive adjustments in how the two reward attributes were processed in the decoy task (data not shown). Importantly, we found an overall positive correlation between the difference in representation factors ($\Delta f = f_p - f_m$) and the indifference points in both groups of subjects (Pearson correlation; $r = 0.56, p = 3.5\times10^{-5}$; $r = 0.68, p = 4.0\times10^{-6}$, for Groups 1 and 2, respectively; **Figure 7C**), and this relationship was stronger for Group 2 (two-sided t-test, p = $= 7.3 \times 10^{-6}$). We focused on the difference between the two representation factors because this difference determines relative processing of the two attributes and thus, the final choice.

These results indicate that in both groups, more risk-averse subjects (subjects with larger indifference points) exhibited more sensitive processing of reward magnitude than reward probability. These results dovetail with the correlation between the CRA and the indifference points (**Figure 5C**), indicating that adjustments of neural representation was the main factor for driving changes in risk aversion due to decoy presentation. In addition, we found that the differences in the representation factors (Δf) were overall positive in Group 2 (two-sided Wilcoxon signrank test, p = 0.03; **Figure 7C** inset) and negative in Group 1 (two-sided Wilcoxon signrank test, p = 0.008). Note that a larger value for Δf translates to shallower representation for reward probability than reward magnitude, and thus, our results can explain the overall risk attitude in the two groups of subjects.

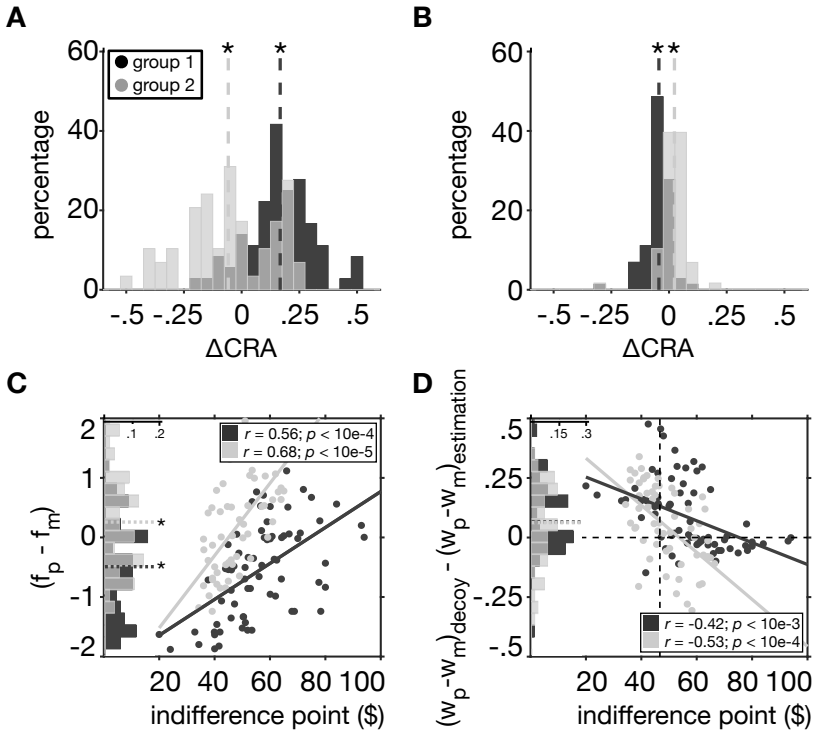


Figure 7. Adjustments of neural representation to decoy presentation and competitive weighting of reward attributes account for opposite patterns of change in risk preference. (**A**) Plotted are the distributions of the difference in CRA between the models with and without adjustments of neural representation (i.e. explained CRA due to neural adjustments) separately for each group of subjects. The dashed lines show the medians, and an asterisk shows a significant difference from zero (two-sided Wilcoxon signed-rank test, $p < 0.05$) (**B**) Plotted are the distributions of the difference in CRA between the models with and without competitive weighting of reward attributes (i.e. explained CRA due to competitive weighting) separately for each group of subjects. Both these models include adjustments of neural representations. (**C**) Plotted is the difference of estimated representation factor for probability and magnitude as a function of the indifference point. The black and gray inset histograms plot the fractions of subjects (black: Group 1; gray: Group 2) with certain values of Δf. (**D**) Plotted is the change in the differential weighting of reward probability and magnitude between the estimation and decoy tasks as a function of the indifference point. Adjustments of neural representation to decoy presentation and competitive weighting of reward attributes were both strongly correlated with the initial degree of risk aversion. The vertical sashed line indicates the indifference point of $46.70, corresponding to risk-neutrality.

Second, we calculated the difference in CRA between the models with and without competitive weighting of reward attributes in order to determine how much of the CRA was influenced by this high-level process (both these models include adjustments of neural representations). We found that for subjects in Group 1, the overall change in CRA due to inclusion of competitive

weighting was significantly smaller than zero (two-sided Wilcoxon signed-rank test; median = -0.04, $p = 4.8 \times 10^{-7}$; **Figure 7B**). In contrast, for subjects in Group 2, the change in CRA due to inclusion of competitive weighting was significantly greater than zero (two-sided Wilcoxon signed-rank test; median = 0.02, $p = 2.0 \times 10^{-6}$). These results indicate that competitive weighting gives rise to a decrease (increase) in risk aversion in Group 1 (respectively, Group 2), which is the opposite of the effect of adjustments in neural representation. Therefore, competitive weighting of reward attributes, which closely resembles selective attentional modulation, can partially compensate for the divergence of risk preference in the two groups of subjects (**Figure 7A**) due to adjustments in value representation.

Interestingly, competitive weighting of reward attributes was also strongly correlated with the initial degree of risk aversion (indifferent point). Specifically, we found significant negative correlations between the changes in the differential weighting of reward probability and magnitude between the estimation and decoy task ($[\mathrm{w_p} - \mathrm{w_m}]_{\mathrm{decoy}} - [\mathrm{w_p} - \mathrm{w_m}]_{\mathrm{estimation}}$) and the indifference points (Pearson correlation; $r = -0.42, p = 8.2 \times 10^{-4}$; $r = -0.53, p = 7.8 \times 10^{-5}$, for Groups 1 and 2, respectively; **Figure 7D**). This relationship was stronger for Group 2 (two-sided t-test, $p = 0.003$) indicating that similar to low-level neural adjustments, high-level changes in reward attribute weights were also more strongly correlated with initial risk aversion in this group. Considering that the differential weighting of reward probability and magnitude ($[w_p - w_m]$) measures the "initial" degree of risk aversion for each individual, a decrease in the differential weighting of reward probability and magnitude between the estimation and decoy tasks ($[\mathrm{w_p} - \mathrm{w_m}]_{\mathrm{decoy}} - [\mathrm{w_p} - \mathrm{w_m}]_{\mathrm{estimation}}) < 0$) would give rise to more risk-seeking behavior and vice versa. The negative correlation between changes in differential weighting of reward probability and magnitude and the initial degree of risk aversion shows that high-level adjustments to decoy presentation (due to competitive weighting of reward attributes) can push risk preference toward risk neutrality.

Together, results of fitting based on our model reveal how two separate mechanisms drive subjects' risk behavior during multi-attribute choice and give rise to opposite changes in preference. On the one hand, low-level adjustments of value representation cause more risk aversion proportional to individuals' initial degree of risk aversion. On the other hand, high-level competitive weighting of reward attributes gives rise to more risk-seeking behavior by increasing

the relative weight for probability and magnitude inversely proportional to individuals' initial degree of risk aversion.

Discussion

Previous studies on choice between multiple options have illustrated that context effects are general features of human choice behavior (Tversky, 1972; Huber et al., 1982; Bettman et al., 1998; Dhar and Simonson, 2003; Simonson 1989; Bettman et al., 1998; Kahneman and Tversky, 1984; Doyle et al., 1999) and moreover, revealed great individual variability across different types of choice (Huber and Puto, 1983; Wedell and Pettibone, 1996; Trueblood, 2012; Soltani et al., 2012; Pan and Lehmann, 1993; Pettibone and Wedell, 2007; Ha et al., 2009), indicating that multiple processes are involved in context-dependent choice. Here, we used choice between multiple risky options to study how context effects are influenced by individuals' risk attitudes and neural adaptation on different timescales. With a large cohort of subjects and a new computational model, we explored individual variability in how reward information is processed and how it influences context effects.

To our surprise, we did not observe any evidence for the effects of neural adaptation on the order of 100 trials on decoy effects. Although this null result could be due to many factors, previous studies suggest that value representations adapt to the range of reward values available in a given condition (e.g., block of trials). For example, Padoa-Schioppa (Padoa-Schioppa, 2009) has shown that the sensitivity of orbitofrontal cortex (OFC) neurons encoding reward value is inversely proportional to the range of values for the available options, and this adaptation can result in optimal coding of reward value (Rustichini et al., 2017). Moreover, neural adaptation is prevalent in different brain areas (Albright and Stoner, 2002; Tobler et al., 2005; Sallet et al., 2007; Carandini and Heeger, 2012; Summerfield and Lange, 2014) and could strongly influence context effects. Interestingly, two recent studies have shown that OFC neurons signal context-dependent reward value only during free choice (Yamada et al. 2018) and contribute to decision making in different contexts (Xie and Padoa-Schioppa, 2016), indicating that neural adaptation in OFC could contribute to value-based decision making. Our finding that neural adaptation on a long timescale may not contribute to context effect suggests that certain neural computations are

performed only when they can result in more flexible behavior (Soltani et al., 2016; Farashahi et al., 2018; Spitmaan et al., 2018).

We also found that context effects strongly depend on individuals' degree of risk aversion. Specifically, using a clustering method, we identified two groups of subjects with different patterns of decoy effects. Subjects in the one group exhibited consistently significant decoy effects and increased risk aversion due to decoy presentation. In contrast, subjects in the other group did not show consistent decoy effects and became more risk seeking due to decoy presentation. We found that subjects in the first group were more risk averse when choosing between two gambles. Thus, for the first time, to the best of our knowledge, we were able to identify distinct behavioral "phenotypes" in terms of context effects and the influence of risk attitudes on these effects. Together, our results suggest that understanding individual variability could be very informative about high-level cognitive functions and their interactions with low-level neural processes.

Although extant models of context-dependent choice assume that context effects occur due to either high-level cognitive or low-level neural processes, we found that both processes are necessary to account for the complex pattern of decoy effects. For example, it has been shown that adjustments of reward value representations can contribute to context effects via different neural mechanisms such as range normalization (Soltani et al., 2012; Louie et al., 2013) or divisive normalization (Louie et al., 2011; Carandini and Heeger, 2012; Heeger, 1992). We found that these adjustments can account for some but not all of the observed context effects. High-level models rely on different cognitive processes to account for context effects, including attentional switching to different choice attributes, menu-dependent evaluation of choice attributes, competition between attribute processing to enhance contrast between certain attributes (Tversky, 1972; Tsetsos et al., 2010; Tversky and Simonson, 1993; Roe et al., 2001; Usher and McClelland, 2004; Johnson and Busemeyer, 2005; Usher et al., 2008; Trueblood et al., 2014; Choplin and Hummel, 2005), and a more recent model based on rank-dependent weighting of information according to the salience of the sampled information (Tsetsos et al., 2012). In addition to adjustments to the set of options on each trial, we found that the model that best fit the experimental data includes competitive weighting of reward probability and magnitude. Therefore, the additional high-level mechanism included in our model shares many

general features of previous high-level models. We note that we did not perform model comparison with existing models because our goal here was to link context effects to risk aversion. Therefore, although able to reveal this link, our proposed model may not be the only model that can capture our experimental results.

Finally, our model revealed the relationship between context effects and risk preference within individuals. Specifically, we found that both low-level adjustments in value representation due to decoy presentation and high-level adjustments in weighting of reward attributes, perhaps via attentional modulations, were correlated with the initial degree of risk aversion within individuals but in the opposite directions. The opposite modulations of low-level and high-level processes shifted risk preference toward risk-aversion and risk-seeking behavior, respectively, and created a distinct pattern of choice behavior in a more complex task.

Materials and Methods

Ethics statement. The study was approved by the Institutional Review Board of Dartmouth College. A consent form was obtained from each subject prior to participating the experiment.

Subjects. In total, 108 healthy Dartmouth College students with normal or corrected to normal vision participated in the study (50 females and 58 males): 38 exclusively in the control condition, 48 exclusively in the range-manipulation condition, and 22 in both conditions, resulting in 130 sets of data (overall, 60 and 70 subjects for the control and range-manipulation conditions, respectively). Data from all subjects were included in the data analyses. Subjects were compensated with a combination of money and "t-points," that are extra-credit points for classes within the department of Psychological and Brain Sciences at Dartmouth College. More specifically, in addition to the base rate of $10/hour or one t-point/hour, subjects were compensated by up to $10/hour depending on the gambles they chose during the tasks (see below).

Experimental paradigm. Subjects performed two gambling tasks consisting of an estimation and a decoy task. Moreover, there were two conditions within the decoy task: control and range-manipulation. In both tasks, subjects were told to select the gamble they believe would result in more reward considering both options' reward probabilities and magnitudes. We told the

subjects that they would only be paid based on the outcome of one of their randomly selected gambles in each task; therefore, every trial was equally relevant to their final payoff. The estimation task was performed first followed by the control and range-manipulation conditions of the decoy task. The order of the control and range-manipulation conditions was randomized between subjects.

In the estimation task (70 trials), subjects had to choose between two gambles defined by the probability p of earning monetary reward m (p, m). The two reward attributes (probability and magnitude) were represented on the screen using two different colors (red and green). The target (T) gamble was low-risk and had a low reward magnitude ($p = 0.7$, $m = \$20 \pm 2$, while the competitor gamble (C) was high-risk and had a variable reward magnitude ($p = 0.3$, $m = \$30 - \$80$). On each trial of this task, two gambles appeared on the screen with the message "Evaluate" on the top of the screen. Subjects had 4 seconds to evaluate both gambles and decide their preferred gamble. When the 4 seconds were over, the message changed to "Choose," and subjects had 1 second to make their selection using the keyboard. Once the selection was made, they could see their choice for one second. Subsequently, a fixation cross appeared on the screen indicating that the next trial was about to start. The choice data of the subject in this task was used to determine the subject's attitudes toward risk. This information was then used to tailor equally preferable target and competitor gambles (for each subject) in the decoy task.

The decoy task was used to determine how the presence of a third gamble changed the preference between the target and competitor gambles. During each trial of the decoy task, three gambles (presented on the same horizontal plane) and a message on top of the screen saying "Evaluate" appeared for 6 seconds (**Figure 1A**). After the evaluation period, the message changed to "Choose" and one of the gambles was simultaneously removed. Subjects had 1 second to choose their preferred gamble from the remaining two using the keyboard. We allowed a short amount of time for submitting a response in order to avoid re-evaluation of the remaining options. After making a choice, their selected gamble appeared for 1 second, followed by a fixation cross indicating that the next trial was going to start. The target and the competitor gambles were tailored for each subject to be equally preferable using the subjective indifference point computed from the estimation task (see below). The third gamble (decoy) could be in one of the following four positions in the attribute space (**Figure 1B**): D1 decoys, which were better

than the competitor gamble in terms of both reward probability and magnitude but were better than the target only in terms of magnitude (asymmetrically dominant); D2 decoys, which were worse than the competitor in both probability and magnitude but were worse than the target only in terms of probability (asymmetrically dominated); D3 decoys, which were better than the target in both probability and magnitude but were better than the competitor only in terms of probability (asymmetrically dominant); and D4 decoys, which were worse than the target in both probability and magnitude but were better than the competitor only in terms of probability (asymmetrically dominated). Therefore, decoys could be asymmetrically dominant or asymmetrically dominated. Four decoy types were presented equally (30 times each), and the decoy type was pseudo-randomly assigned in every trial.

Overall, each subject performed 120 trials in the control condition and 150 trials in the range-manipulation condition of the decoy task. In 2/3 of the trials, the decoy gamble disappeared, whereas in the remaining 1/3 of the trials, either the target or the competitor was randomly selected to disappear in order to make decoy identification impractical. Removing the decoy before choice created the so-called phantom decoy effect, as the decoy was not available in the choosing portion of the trial. This design allowed us to measure the effect of asymmetrically dominant decoys that could not be studied otherwise. The trials in which either the target or the competitor disappeared were not considered in the analyses. Because we only allowed only one second for submitting a response, subjects could miss a significant percentage of trials (median of missed trials was equal to 12%). To avoid losing too many trials, we added any missed trials back into the subject's set such that each subject completed the same number of total trials. However, due to trial replacement, it took longer for subjects (that missed many trials) to perform the experiment, and consequently, some subjects were unable to perform the task in the range-manipulation condition.

In the control condition, the target was low-risk and had a low reward magnitude ($p = 0.7 \pm 0.05$, $m = \$20 \pm 2$), and the competitor was high-risk and had a variable reward magnitude ($p = 0.3 \pm 0.05$, $m = \$m_0 \pm 2$), where m_0 is the indifference point (see Eq. 1 below) obtained in the estimation experiment. Specifically, the magnitude value of the competitor was tailored for each subject based on their data from the estimation task such that the competitor and the target were equally preferable for every subject. For the decoy gambles, probability values ranged from

0.2 to 0.85, and magnitude values ranged from \$15 to \$70. The range-manipulation condition was similar to the control condition except that 20% of the trials presented a set of "high-range" gambles. The reward probability and magnitude of the high-range gambles were the following: $(p = 0.9 \pm 0.05, m = \$10 \pm 2), (p = 0.1 \pm 0.05,\ m = \$90 \pm 2)$, and $(p = 0.5 \pm 0.05,\ m = \$20 \pm 2)$. The high-range trials were removed from the subsequent analysis, as their only purpose was to expand the range of gamble attributes and disrupt neural adaptation.

Data analysis. Subjects' choice behavior in the estimation task was fit using a sigmoid function. More specifically, the probability of choosing the low-risk gamble (the gamble with reward probability $p_R = 0.7$ and a reward magnitude of $m = \$20$), $p_{low-risk}$, was fit as a function of the reward magnitude of the high-risk gambles, $m_{high-risk}$:

$$p_{low-risk} = \frac{1}{1+e^{\left(\frac{-(m_{high-risk}-\mathrm{m}_0)}{\sigma}\right)}} \qquad \text{(Eq. 1)}$$

where m_0 corresponds to a reward value at which the high-risk gamble was equally preferable to the low-risk gamble (i.e., the indifference point), and σ measures the stochasticity in choice.
In the decoy task, we computed the decoy efficacy (DE_k) for each of four locations in the attribute space that a decoy could appear as follows:

$$DE_k = P_T(D_k) - P_T \qquad \text{(Eq. 2)}$$

where $P_T(D_k)$ is the probability of choosing the target given the decoy k ($k = \{1,2,3,4\}$) and P_T is the probability of choosing the target gamble for a given subject. The decoy efficacy allowed us to measure decoy effects for each location relative to the overall subject preference for the target gamble (P_T).

In order to fully characterize the change in preference due to the introduction of a decoy to the choice set, we defined four additional quantities ("decoy-effect" indices) based on DE_k or $P_T(D_k)$ values as follows:

$$AE = \frac{+DE_1 - DE_2 - DE_3 + DE_4}{4} = \frac{+P_T(D_1) - P_T(D_2) - P_T(D_3) + P_T(D_4)}{4}$$
$$FE = \frac{-DE_1 - DE_2 + DE_3 + DE_4}{4} = \frac{-P_T(D_1) - P_T(D_2) + P_T(D_3) + P_T(D_4)}{4}$$
$$DD = \frac{+DE_1 - DE_2 + DE_3 - DE_4}{4} = \frac{+P_T(D_1) - P_T(D_2) + P_T(D_3) - P_T(D_4)}{4} \qquad \text{(Eq. 3)}$$
$$CRA = \sum_i \frac{P_T(D_i)}{4} - 0.5$$

The first index, the overall attraction effect (AE), is computed by averaging decoy efficacies over all decoy locations and considering reversed signs for the asymmetrically dominant and dominated decoys for the target and competitor ([D1-D2] versus [-D3+D4]). On the one hand, presentation of D1 and D4 decoys can increase the preference for the target by asymmetrically dominating the competitor or being dominated by the target, respectively. On the other hand, D2 and D3 decoys can reduce the preference for the target by being asymmetrically dominated by the competitor and dominating the target, respectively. Therefore, the AE measures the overall attraction effect for both dominant and dominated decoys, or more specifically, how much the presentation of the decoy increases preference for the target relative to the competitor. The second index, the overall frequency effect (FE), quantifies the overall tendency to choose the gamble next to the decoy in terms of both reward probability and magnitude (i.e., the competitor for D1 and D2 decoys, and the target for D3 and D4 decoys). The third index, the dominant vs. dominated (DD), quantifies the overall difference between the effects of dominant and dominated decoys. The fourth index measures the overall change in risk aversion (CRA) due to the decoy presentation. Note that due to the symmetry in decoy presentation and because the target and competitor gambles were set based on the indifference point in the estimation task, these gambles should be selected equally (with probability equal to 0.5) in the absence of any changes in risk aversion due to decoy presentation. These four decoy-effect indices are linear transformations of decoy efficacies and could take a value between -0.5 and 0.5. Importantly, these indices are orthogonal to each other.

For analysis of the time course of decoy effects, we used a moving window with a length of 40 trials and a step size of 20 trials in order to calculate changes in decoy-effect indices and the probability of choosing the target over time. The statistical test used for each comparison is reported where it is mentioned, and the reported effect sizes are Cohen's d values.

Clustering analysis. To identify distinct patterns of decoy effects, we separated subjects into different clusters using the k-means clustering method based on various sets of decoy-effect indices as features and for different numbers of clusters. Specifically, we tried all possible exclusive combinations of the four decoy-effect indices to cluster subjects. The "silhouette" values were calculated for clustering data with different features and numbers of clusters. The silhouette values range from -1 to 1, where a high value indicates that the data point is well matched to its own cluster and poorly matched to neighboring clusters. We found that clustering based on all four decoy-effect indices and into two clusters provides the best separation between data. We then used the clustering indices based on this model to quantify subjects' behavior in each cluster/group.

Computational model. Here, we extended our previous model of context-dependent choice (Soltani et al., 2012) to include the influence of high-level cognitive processes. In this model, we assumed that the values of gambles in a given attribute dimension (reward magnitude and probability) are encoded by a neural population selective to that attribute (attribute-encoding neurons). The output of attribute-encoding neurons, in turn, projects to value-encoding neurons that compute the overall value of each gamble (**Supplementary Figure 3**).

More specifically, we assumed that the response of neurons encoding attribute i ($i = \{m, p\}$), r_i, is a linear function of the attribute value v_i as follows:

$$\begin{aligned} r_i(v_i, f_i) &= 0 \quad if\ v_i < c_i^t(f_i) \\ r_i(v_i, f_i) &= k(v_i - c_i^t(f_i)) \quad if\ c_i^t(f_i) \leq v_i \leq c_i^s(f_i) \\ r_i(v_i, f_i) &= 1 \quad if\ v_i > c_i^s(f_i) \end{aligned} \qquad \text{(Eq. 4)}$$

where k is the slope, $c_i^t(f_i)$ and $c_i^s(f_i)$ are the threshold and saturation points, respectively, and f_i is the representation factor that determines the dynamic range (i.e., the interval between threshold and saturation) of neural response for attribute *i*. Specifically, the threshold and saturation points depend on the representation factor and the attribute values of the gambles in the choice set as follows:

$$\begin{cases} c_i^t(f_i) = \frac{v_{i,min} - f_i v_{i,nmin}}{1 - f_i} & if\ f_i \geq 0 \\ c_i^t(f_i) = v_{i,min} - f_i(v_{i,nmin} - v_{i,min}) & if\ f_i < 0 \end{cases}$$
$$\begin{cases} c_i^s(f_i) = \frac{v_{i,max} - f_i v_{i,nmax}}{1 - f_i} & if\ f_i \geq 0 \\ c_i^s(f_i) = v_{i,max} + f_i(v_{i,max} - v_{i,nmax}) & if\ f_i < 0 \end{cases} \quad \text{(Eq. 5)}$$

where $v_{i,min}$ and $v_{i,nmin}$ are the minimum and next to the minimum values of the presented gamble in a given attribute, respectively; and $v_{i,max}$ and $v_{i,nmax}$ are the maximum and the next to the maximum values of the presented gamble, respectively. Positive values of f_i allow the dynamic range to extend beyond the minimum and maximum attribute values, whereas the negative values limit the dynamic range between the minimum and maximum values (Soltani et al., 2012). Therefore, f_i's sign and magnitude determine the dynamic range of neural response.

The overall value of each gamble, *V*(*X*), is computed by a weighted sum of outputs from the attribute-encoding neurons as follows:

$$V(X) = \mathrm{w_p}(D_k)\, r_p(X_p, f_p) + \mathrm{w_m}(D_k)\, r_m(X_m, f_m) \quad \text{(Eq. 6)}$$

where $w_i(D_k)$ is the weight of the connection from the attribute-encoding population i ($i = \{m, p\}$) to the option-value population when decoy D_k was presented ($k = \{1,2,3,4\}$).

For some model simulations and fitting, we assumed a single neural representation factor f_0 to capture neural adjustments for both reward probability and magnitude:

$$f_p = f_0;\ f_m = -f_0 \quad \text{(Eq. 7)}$$

where f_p and f_m are neural representation factors for reward probability and magnitude, respectively. For other simulations and fitting, we used separate representation factors (f_p and f_m) to allow more flexible representation.

To determine the weights of the connections between the attribute-encoding and option-value encoding neurons ($w_i(D_k)$), we assumed a competitive weighting mechanism as follows:

$$\begin{cases} \mathrm{w_p}(\mathrm{D_k}) = 0.5 + (b_0 + b_k) \\ \mathrm{w_m}(\mathrm{D_k}) = 0.5 - (b_0 + b_k) \end{cases} \quad \text{(Eq. 8)}$$

where b_0 is a constant bias and b_k is the location-dependent bias for decoy D_k. The constant bias, b_0, could be a free parameter or could be set based on the choice behavior in the estimation task as follows:

$$\mathrm{b}_0 = \frac{w_p - w_m}{2} \qquad \text{(Eq. 9)}$$

where w_p and w_m are estimated weights for the reward probability and magnitude in the estimation task, respectively (using Eq. 6). Note that the differential weighting of reward probability and magnitude ($[w_p - w_m]$) in the estimation task was strongly correlated with the indifference points (Pearson correlation; $r = 0.92$, $p = 2.8 \times 10^{-8}$), indicating that the former quantity measures the "initial" degree of risk aversion for each subject.

The location-dependent biases, b_k, could be different for the four decoy locations or coupled between certain decoys as follows:

$$\mathrm{b}_1 = \mathrm{b}_2 = \mathrm{b}_{12}, \mathrm{b}_3 = \mathrm{b}_4 = \mathrm{b}_{34} \qquad \text{(Eq. 10)}$$
$$\mathrm{b}_1 = \mathrm{b}_3 = \mathrm{b}_{13}, \mathrm{b}_2 = \mathrm{b}_4 = \mathrm{b}_{24} \qquad \text{(Eq. 11)}$$
$$\mathrm{b}_1 = \mathrm{b}_4 = \mathrm{b}_{14}, \mathrm{b}_2 = \mathrm{b}_3 = \mathrm{b}_{23} \qquad \text{(Eq. 12)}$$

Although we considered all the above possibilities to determine the weights, only coupling based on Eq. 10 does not require identification of decoys and is plausible. Finally, we also assumed separate location-dependent biases for probability and magnitude in order to construct a more-flexible model.

Overall, we generated 16 alternative models using all logical combinations of representational and weighting mechanisms to fit individuals' choice data in terms of decoy-effect indices. We performed cross validation to compare the prediction error of decoy-effect indices to determine the best model. Specifically, we used 70% of data (randomly chosen) to estimate the best-fitting parameters and then calculated the model prediction errors over the remaining 30% of data. We found that the model with separate representation factors (f_p and f_m) and similar location-dependent biases for decoys next to the target and competitor (Eq. 10) provides the best fit. Therefore, we chose this model to fit the pattern of decoy-effect indices for each subject separately in order to estimate the representation factors and weights of connections for each subject.

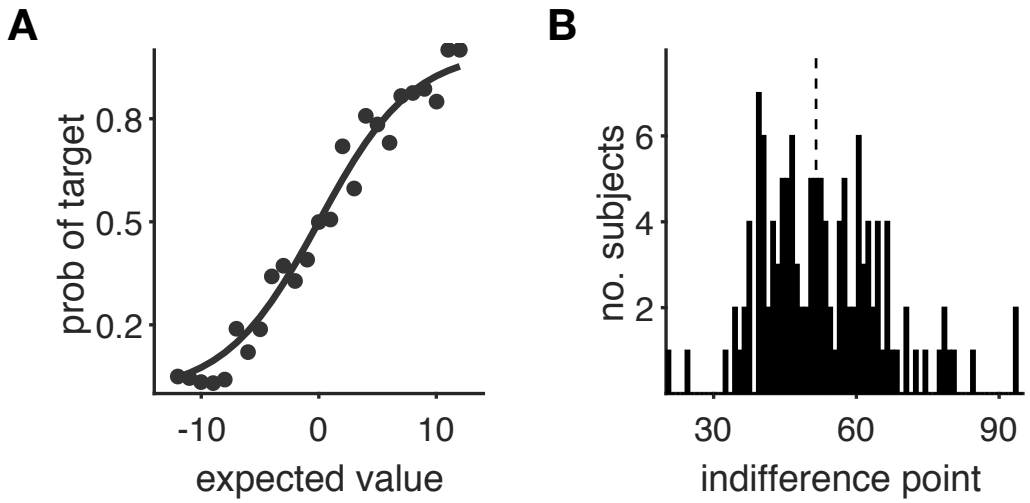
References

1. Albright T, Stoner G (2002) Contextual influences on visual processing. *Annual review of neuroscience* 25:339–379.
2. Bettman J, Luce M, Payne J (1998) Constructive consumer choice processes. *Journal of consumer research* 25:187–217.
3. Carandini M, Heeger, D (2012) Normalization as a canonical neural computation. *Nature Reviews Neuroscience* 13:51–62.
4. Choplin J, Hummel J (2005) Comparison-induced decoy effects. *Memory & cognition* 33:332–343.
5. Dhar R, Glazer R (1996) Similarity in context: Cognitive representation and violation of preference and perceptual invariance in consumer choice. *Organizational Behavior and Human Decision Processes* 67:280–293.
6. Dhar R, Sherman S (1996) The effect of common and unique features in consumer choice. *Journal of Consumer Research* 23:193–203.
7. Dhar R, Simonson I (2003) The effect of forced choice on choice. *Journal of marketing research* 40:146–160.
8. Doyle J, O'Connor D, Reynolds G, Bottomley P (1999) The robustness of the asymmetrically dominated effect: buying frames, phantom alternatives, and in-store purchases. *Psychology & Marketing* 16:225–243.
9. Farashahi S, Azab H, Hayden B, Soltani A (2018) On the flexibility of basic risk attitudes in monkeys. *Journal of Neuroscience* 38:4383–98.
10. Ha Y, Park S, Ahn H (2009) The influence of categorical attributes on choice context effects. *Journal of Consumer Research* 36:463–477.
11. Heath, T. B., & Chatterjee, S (1995) Asymmetric decoy effects on lower-quality versus higher-quality brands: *Meta-analytic and experimental evidence. Journal of Consumer Research* 22:268–284.
12. Heeger D (1992) Normalization of cell responses in cat striate cortex. *Visual neuroscience* 9:181–197.

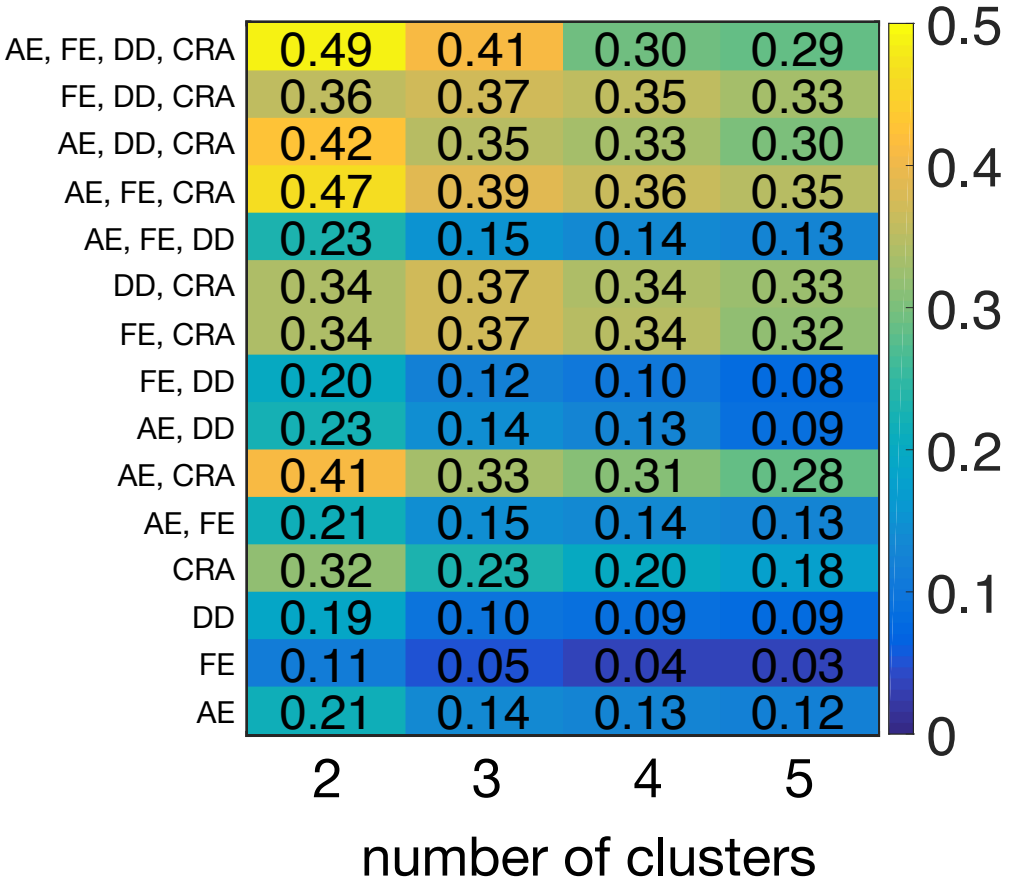
13. Huber J, Payne J, Puto C (1982) Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research* 9:90–98.
14. Huber J, Puto C (1983) Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research* 10:31–44.
15. Johnson J, Busemeyer J (2005) A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological review* 112:841–861.
16. Kahneman D, Tversky A (1984) Choices, values, and frames. *American psychologist* 39:341–350.
17. Louie K, Grattan L, Glimcher P (2011) Reward value-based gain control: divisive normalization in parietal cortex. *Journal of Neuroscience* 31:10627–39.
18. Louie K, Khaw M, Glimcher P (2013) Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences* 110:6139–44.
19. Padoa-Schioppa C, Assad J (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature neuroscience* 11:95–102.
20. Padoa-Schioppa C (2009) Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience* 29:14004–14.
21. Pan Y, Lehmann D (1993) The influence of new brand entry on subjective brand judgments. *Journal of Consumer Research* 20:76–86.
22. Payne J, Bettman J, Johnson E (1988) Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:534–552.
23. Pettibone J, Wedell, D (2000) Examining models of nondominated decoy effects across judgment and choice. *Organizational behavior and human decision processes* 81:300–328.
24. Pettibone J, Wedell D (2007) Testing alternative explanations of phantom decoy effects. *Journal of Behavioral Decision Making* 20:323–341.
25. Pratkanis A, Farquhar P (1992) A brief history of research on phantom alternatives: Evidence for seven empirical generalizations about phantoms. *Basic and Applied Social Psychology* 13:103–122.
26. Ratneshwar S, Shocker A, Stewart D (1987) Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research* 13:520–533.

27. Roe R, Busemeyer J, Townsend J (2001) Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review* 108:370–392.
28. Rustichini A, Conen K, Cai X, Padoa-Schioppa C (2017) Optimal coding and neuronal adaptation in economic decisions. *Nature Communications* 8:1208. https://doi.org/10.1038/s41467-017-01373-y.
29. Sallet J, Quilodran R, Rothé M, Vezoli J, Joseph J, Procyk E (2007) Expectations, gains, and losses in the anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience* 7:327–336.
30. Simonson I (1989) Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research* 16:158–174.
31. Simonson I, Tversky A (1992) Choice in context: Tradeoff contrast and extremeness aversion. *Journal of marketing research* 29:281–95.
32. Soltani A, De Martino B, Camerer C (2012) A range-normalization model of context-dependent choice: a new model and evidence. *PLoS computational biology* 8:e1002607. https://doi.org/10.1371/journal.pcbi.1002607.
33. Soltani A, Khorsand P, Guo C, Farashahi S, Liu J (2016). Neural substrates of cognitive biases during probabilistic inference. *Nature Communications*, 7, 11393. https://doi.org/10.1038/ncomms11393
34. Spitmaan M, Chu E, Soltani A (2018). Salience-driven value construction for adaptive choice under risk. *bioRxiv*, 506832
35. Summerfield C, De Lange F (2014) Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience* 15:745–756.
36. Tobler P, Fiorillo C, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–1645.
37. Trueblood J (2012) Multialternative context effects obtained using an inference task. *Psychonomic bulletin & review* 19:962–968.
38. Trueblood J, Brown S, Heathcote A (2014) The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review* 121:179–205.
39. Tsetsos K, Usher M, Chater N (2010) Preference reversal in multiattribute choice. *Psychological review* 117:1275–1293.

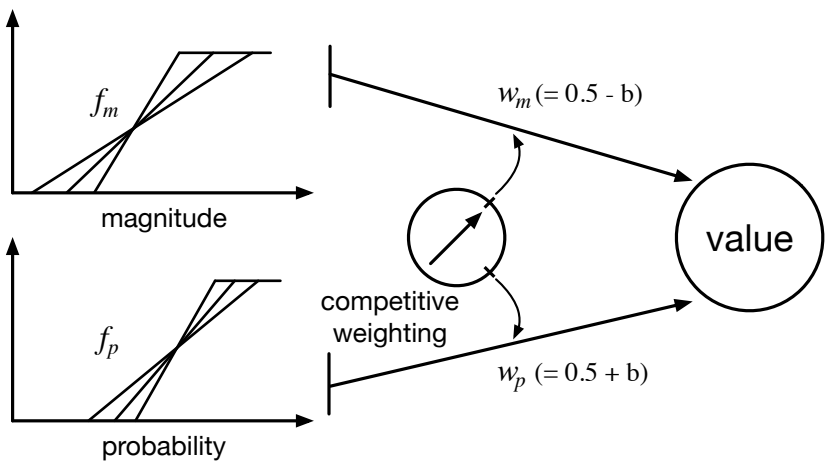
40. Tsetsos K, Chater N, Usher M (2012) Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences* 109:9659-64.
41. Tversky A (1972) Elimination by aspects: A theory of choice. *Psychological review* 79:281–99.
42. Tversky A, Simonson I (1993) Context-dependent preferences. *Management science* 39:1179–89.
43. Usher M, McClelland J (2004) Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review* 111:757–69.
44. Usher M, Elhalal A, McClelland J (2008) The neurodynamics of choice, value-based decisions, and preference reversal. *The probabilistic mind: Prospects for Bayesian cognitive science*, eds Chater N, Oaksford M (Oxford University Press), pp 277–300.
45. Wedell D, Pettibone J (1996) Using judgments to understand decoy effects in choice. *Organizational Behavior and Human Decision Processes* 67:326–344.
46. Xie J, Padoa-Schioppa C (2016) Neuronal remapping and circuit persistence in economic decisions. *Nature Neuroscience* 19:855–61.
47. Yamada H, Louie K, Tymula A, Glimcher P (2018) Free choice shapes normalized value signals in medial orbitofrontal cortex. *Nature communications* 9:162-173.



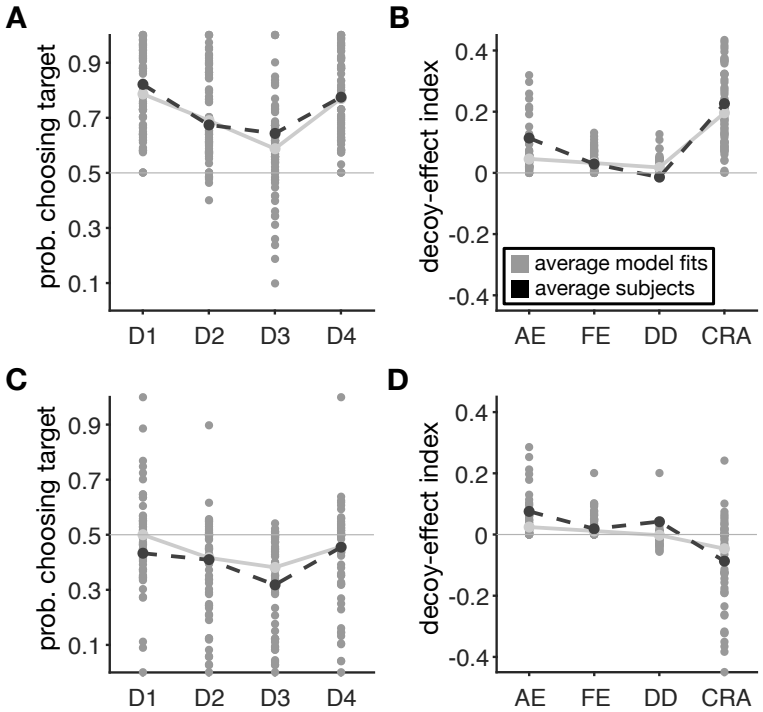
Supplementary Figure 1. Behavior of subjects during the estimation task. **(A)** Plotted is the probability of choosing the low-risk gamble (target) as a function of the reward magnitude of the high-risk gambles for an individual subject. The solid curve shows the fit using a sigmoid function. Each black dot represents an individual subject's data. **(B)** Distribution of estimated indifference points in estimation task across subjects. The indifference point is defined as the magnitude of a high-risk gamble that was as equally preferred as the low-risk gamble. The dashed line indicates the median. Overall, we observed large variability for the indifference point across subjects.



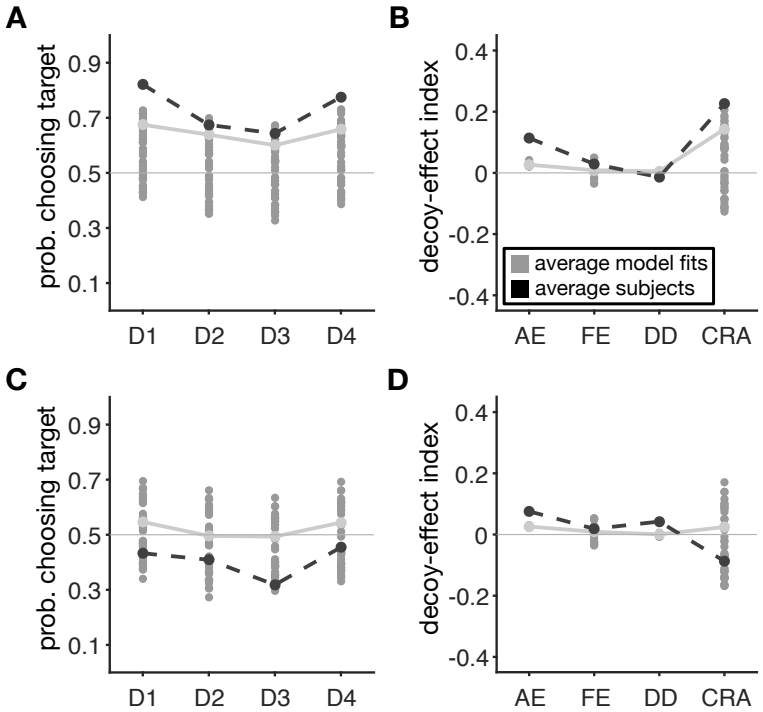
Supplementary Figure 2. Quality of clustering based on various combinations of decoy-effect indices (noted on the y-axis) and number of clusters (x-axis). Reported are silhouette values for a given number of clusters and combination of decoy-effect indices (AE: attraction effect; FE: frequency effect; DD: dominant vs. dominated; CRA: change in risk aversion). The silhouette can take any values between -1 and 1; higher values indicate that each data point (a measure or set of measures) is closely matched to its own cluster and poorly matched to neighboring clusters. Best clustering results are achieved using all decoy-effect indices and two clusters.



Supplementary Figure 3. Schematic of the extended model with both low-level and high-level mechanisms. Values of gambles in a given attribute (reward magnitude and probability) are encoded by the corresponding attribute-encoding population of neurons. We assumed that the response of attribute-encoding neurons (response curve) to be a linear function of the attribute value. Representation factor f_i ($i = \{m, p\}$) determines dynamic range (threshold and saturation points) of neural response as well as the slope of the response curve for each attribute. We assumed that an additional competitive mechanism could modulate the weight of each attribute (w_m and w_p) on the overall value of each gamble.



Supplementary Figure 4. A model with only adjustments of value representations cannot capture all aspects of the experimental data. Plotted are fit of individual subjects based on a model that has adjustments to decoy presentation similar to our original model (Soltani et al., 2012). (**A, C**) Each gray circle shows the probability of selecting the target for different decoy locations based on the model fit for individual subjects in the first (A) and second (C) groups of subjects. For comparison, the average value across subjects (black dashed lines) and their fits (gray solid lines) are plotted as well. (**B, D**) Plots show four measures for quantifying different effects of decoys on preference based on the cited model fit for individual subjects in the first (B) and second (D) groups of subjects (AE: attraction effect; FE: frequency effect; DD: dominant vs. dominated; and CRA: change in risk aversion).



Supplementary Figure 5. A model with only competitive weighting of reward attributes cannot capture all aspects of the experimental data. Plotted are fit of individual subjects based on a model that only assumes a competitive mechanism to modulate the weight of each attribute (w_m and w_p) on the overall value of each gamble. (**A, C**) Each gray circle shows the probability of selecting the target for different decoy locations based on the model fit for individual subjects in the first (A) and second (C) groups of subjects. For comparison, the average value across subjects (black dashed line) and their fits (gray solid lines) are plotted as well. (**B, D**) Plots show four measures for quantifying different effects of decoys on preference based on the cited model fit for individual subjects in the first (B) and second (D) groups of subjects (AE: attraction effect; FE: frequency effect; DD: dominant vs. dominated; and CRA: change in risk aversion).