Selective outcome reinstatement during evaluation drives heuristics in risky choice

Evan M. Russek[1,2,*], Rani Moran[1,2], Yunzhe Liu[1,2,3,4], Raymond J. Dolan[1,2], Quentin J.M. Huys[1,2,5,6]

[1] Max Planck University College London Centre for Computational Psychiatry and Ageing Research, University College London, London, UK
[2] Wellcome Centre for Human Neuroimaging, University College London, London, UK
[3] State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China.
[4] Chinese Institute for Brain Research, Beijing, China.
[5] Camden and Islington NHS Foundation Trust, London, UK
[6] Division of Psychiatry, University College London, London, UK

[*] Correspondence: e.russek@ucl.ac.uk

Abstract

A ubiquitous feature of human decision making under risk is that individuals differ from each other, as well as from normativity, in how they incorporate reward and probability information. One possible explanation for these deviations is a desire to reduce the number of potential outcomes considered during choice evaluation. Although multiple behavioral models can be invoked involving selective consideration of choice outcomes, whether differences in these tendencies underlie behavioral differences in sensitivity to reward and probability information is unknown. Here we consider neural evidence where we exploit magnetoencephalography (MEG) to decode the actual choice outcomes participants consider when they decide between a gamble and a safe outcome. We show that variability in tendencies of individual participants to reinstate neural outcome representations, based on either their probability or reward, explains variability in the extent to which their choices reflect consideration of probability and reward information. In keeping with this we also show that participants who are higher in behavioral impulsivity fail to preferentially reinstate outcomes with higher probability. Our results suggest that neural differences in the degree to which outcomes are considered shape risk taking strategy, both in decision making tasks, as well as in real life.

Introduction

The information we consider when making decisions ultimately determines our choices. This relationship between thought content and action is particularly relevant in mental health disorders, where behavioral tendencies often relate to idiosyncrasies in the potential outcomes that are entertained. For example, defensive behaviors, commonly observed in compulsive and anxiety disorders, are frequently accompanied by excessive consideration of extremely negative events that are very unlikely, while risk seeking is characterized by an inordinate focus on improbable potential gains. Although the modification of such thought biases is a cornerstone of modern psychotherapeutic interventions, whether biases in risky choice problems are in fact related to biases in what information is considered has not been directly tested.

In typical risky choice problems, individuals choose between a 'safe' option with a known, fixed outcome, and a gamble option which might lead probabilistically to one of two possible outcomes. Normative choice in such settings requires evaluating the gamble by summing the utility of each uncertain outcome, weighted by its probability, and comparing this expected utility to the utility of a known safe option (Bernoulli, 1954). However, a key feature of human decision making in this domain is that choices made by different individuals reflect differential sensitivity to rewards or probabilities (Farashahi, Donahue, Hayden, Lee, & Soltani, 2019; Gonzalez et al., 1999; Kahneman & Tversky, 1979). One possible explanation for such deviations from normativity, as well as variability, is the need for individuals to employ heuristics that reduce the computational burden entailed in this rational approach to choice (Gigerenzer & Goldstein, 2011; Lieder & Griffiths, 2019; Payne, Bettman, & Johnson, 1988). Whereas the normative choice strategy requires independent consideration of each possible task outcome, individuals can reduce the number of outcomes they consider by prioritizing use of a particular type of information for evaluation (Farashahi et al., 2019; Stewart, 2011). For example, individuals could prioritize probability information, and selectively ignore the safe outcome as well as the unlikely gamble outcome, thus deciding solely based on whether the more likely gamble outcome is attractive. Alternatively, they could prioritize reward information, and pay attention solely to outcomes useful for comparison along this dimension. Indeed, there is evidence that models which alter the extent to which reward and probability information are separately weighted in a decision variable can account for common patterns of choice variation observed between participants' risky choices (Farashahi et al., 2019; Stewart, 2011). However, whether such weights truly reflect strategies for what information is considered when deciding, is unknown.

A challenge in addressing this type of question is that consideration of disparate pieces of choice relevant information can occur automatically, unconsciously and on a very fast timescale, rendering any distinction difficult based purely upon behavioral measurement. However, recent advances in multivariate methods for Magnetoencephalography (MEG), offer a new tool with which to potentially overcome this challenge (Kurth-Nelson, Barnes, Sejdinovic, Dolan, & Dayan, 2015; Liu, Mattar, Behrens, Daw, & Dolan, 2021; Wise, Liu, Chowdhury, & Dolan, 2021). By representing task components using visual stimuli whose representations can be decoded from MEG sensor activity, these studies have identified signatures of reactivation that occur within a timeframe of 500 ms, and often as fast as 50 to 70 ms, following a cue.

Here, we extend on this work to examine the questions posed above related to individual differences in risk-taking behavior. We show that variability in which outcomes are reinstated during evaluation can explain variation in tendencies to base choices on either probability or reward information. Furthermore, reinstatement of high probability outcome representations was reduced in participants with impulsive traits, suggesting that real life risk taking includes a contribution arising out of a failure to consider probability information. Thus, our findings establish a link between a neural reinstatement of different sources of choice relevant information and the types of decision patterns individuals manifest in risky choice. The findings may have important relevance for mechanisms of aberrant risky choice in real life decisions.

Results

Decision-Making Task

Participants (n = 19) completed a decision-making task while we acquired simultaneous neural data using MEG (Fig. 1). The design involved a risky decision-making task that required consideration of potential outcomes. On each trial, participants were presented with a gamble that required an accept or reject choice (Fig. 1a). Rejecting the gamble led to collection of a safe outcome, OS. Accepting led to collection of one of two gamble outcomes, O1 or O2. The chances of encountering O1 versus O2 upon acceptance of the gamble was signaled by presentation of one of four probability stimuli (P1, P2, P3, or P4; Fig. 1b). The probabilities implied by each of these stimuli were both instructed, extensively experienced, and tested on prior to task commencement (Supplementary Fig. 1). To encourage the prospective consideration of choice outcomes on each trial, the number of points paired with each outcome changed on each trial by addition of structured noise (Fig. 1c).

Critically, in order to facilitate MEG analysis, the time course by which information was presented was structured such as to force evaluation of choice options at an identifiable timepoint. Specifically, participants were first informed of the number of points paired with each outcome (Fig. 1a, left). However, this information was insufficient to make choices as the probabilities relevant to that trial were unknown to the participant at this timepoint. Choice evaluation involving the integration of outcomes O1 and O2 with their probability and the comparison with the safe value could only start when the probability stimulus appeared on the screen following this (Fig. 1a, middle). At this point, the outcome stimuli were no longer on the screen. Hence, we aimed to decode the neural signatures of the outcome stimuli during the time when the probability stimulus was on the screen with the aim of using this to gain insights into what information was being reinstated during the choice process and how this information was related to the resulting choice.

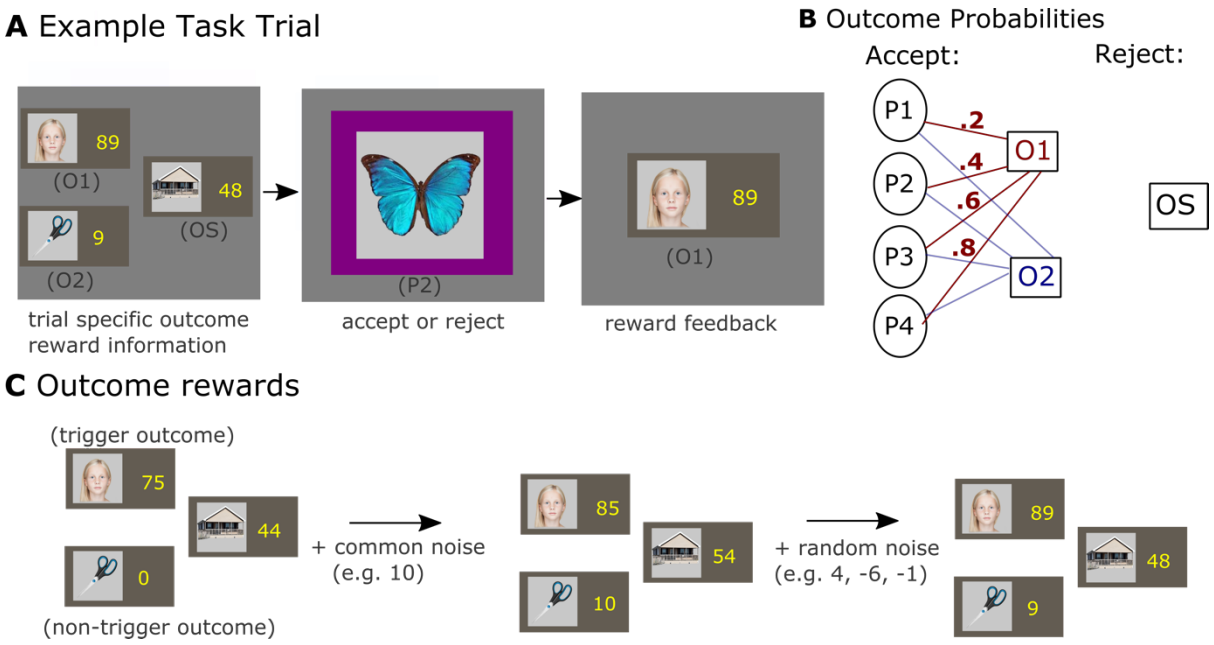


***Fig. 1. Task**. Participants (n = 19) completed a decision task to probe online integration of outcome probabilities and rewards in the MEG scanner. On each trial, participants chose between a safe stimulus (OS) or a gamble which probabilistically led to one of two outcome stimuli. The task controlled when specific computations could be performed by providing the information required for the computation in discrete stages: thus, participants first obtained information about the value of each outcome, and in a second stage about the probabilities of each outcome, O1 or O2.*

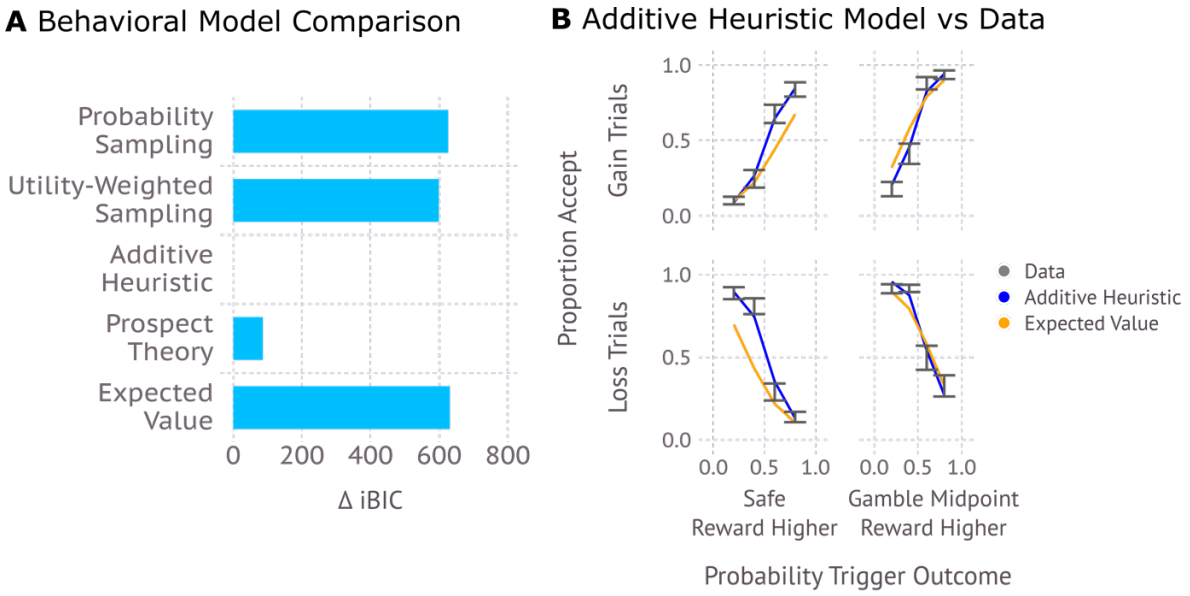
***a) Example Task Trial.** Participants were first informed of the point values for all the three outcomes OS, O1 and O2. Because they did not know the probabilities of the outcomes, they could not yet compute the expected value of the gamble. In the next step, participants were presented with one of four possible probability stimuli (P1, P2, P3 or P4) on which they had been pretrained, indicating four different probability combinations. They then decided whether to accept or reject the gamble. Rejecting led to collection of OS along with its trial-specific associated points. Accepting led to collecting either O1 or O2 along with the trial-specific associated points. All outcome and choice stimuli were represented by decodable visual stimuli. Note that in the example trial, the gamble was accepted.*

***b) Outcome Probabilities.** The chances of collecting O1 versus O2 upon accepting the gamble depended on which probability stimulus was presented. Probability of reaching O1 was .2, .4, .6, and .8 for P1, P2, P3 and P4 respectively, and p(O2) = 1 - p(O1). These probabilities were extensively pretrained. Rejecting the choice stimulus always led to collection of OS.*

***c) Outcome rewards.** On each trial either O1 or O2 was designated to be the "trigger" outcome, whose value was selected from three levels (45, 65, or 75 during gain blocks or -45 -65 or -75 on loss blocks). The non-trigger outcome was always 0. OS was selected from 4 levels (20, 32, 44, 56 during gain blocks or -20, -32, -44, -56 during loss blocks). In order to discourage habitual responding to repeated choices, a variable amount of common noise (between 0 and 20) was added to all outcomes. Finally, a random value (between -6 and 6) was added to each outcome separately.*

Behavior Reflects Additive Integration of Reward and Probability Components

We used computational models, fitted to participants' choice data, to provide the simplest possible description of their behavior (see Methods for full description of all models and fitting procedures). A baseline model fit was an expected value model, which decided based on the difference between the expected value of the gamble and safe option, and was provided with a single free parameter to capture decision noise. Comparing predictions of this model to aggregate choice data revealed that it fit the data poorly (Fig. 2b; see supplementary Fig. 6b for comparison to some individual participants). In particular, it underestimated the influence of outcome probability (Fig. 2b: note the greater slope within each panel of data compared to expected value predictions) and overestimated the influence of outcome rewards in choice (Fig. 2b: note the lesser change in data compared to expected value predictions across panel columns).



C Additive Heuristic Model

proability information component

reward information component

$$\log\left(\frac{P_{accept}}{P_{reject}}\right) \approx \beta_{gain/loss} + \beta_{prob}[P_{O_{better}} - P_{O_{worse}}] + \beta_{reward}[\frac{R^*_{O_{trig}}}{2} - R^*_{O_{Safe}}]$$

Fig. 2. Reward and probability information components are added rather than multiplied to determine choices. *Behavioral analysis suggests an additive combination of reward and probability information.*

a) Behavioral Model Comparison. *We compared the ability of a range of computational models to explain participants' choice behavior. Each bar gives iBIC (integrated Bayesian Information Criterion) relative to the best fitting model (for n = 19 participants). Standard expected value as well as sampling models provided poor fits to the data. The best-fitting model was the additive heuristic model followed by a prospect theory model.*

b) The Additive Heuristic Model captures aggregate patterns in choice data. *Comparison of Additive Heuristic model predictions and observed data, with expected value model predictions provided for reference. Each data error bar (grey) shows the across-subject mean (+/- s.e.m.) proportion acceptance for each combination of whether a trial is gain or loss (row), whether the safe reward is higher or lower than the midpoint between the two gamble*

reinforcements (column) and trigger outcome probability contingent on acceptance (x-axis). Note values reflect outcome reinforcements prior to common and other noise added. The blue line shows predictions of the additive heuristic model, at best fit parameters. Relative to predictions of the expected value model (orange) the additive heuristic model was able to capture an over-weighting of outcome probabilities in valuation.

c) The Additive Heuristic Model additively combines reinforcement and probability information. *The additive heuristic model computes two components from the gamble information. The first component, the "probability information component", measures the difference in probability between reaching the better (higher reward) versus worse gamble outcome, contingent on accepting the choice stimulus. The second component, the "reward information component", measures the difference in reward associated with the midpoint between the two gamble reward and the safe reward. Note that because of the actual reward used in the task (Fig. 1c.) this difference can be computed by considering the trigger and safe rewards without needing to refer to the non-trigger reward. R* refers to the reward after the non-trigger reward (which simply amounts to common noise along with noise specific to that outcome) has been subtracted from all rewards. Working with R*, the difference between the gamble midpoint and safe reward can be computed by dividing the trigger reward by two and subtracting the safe reward.*

The expected value model might fail if individuals differed in the utility derived from the rewards, or if instructed probabilities were distorted. Indeed, a prospect theory model, which provides parameters for both of these distortions (Kahneman & Tversky, 1979; Prelec, 1998a), provided a better fit, and was able to match aggregate patterns in the choice data (Expected Value vs Prospect Theory $\Delta iBIC = 554.89$; Fig. 2a, Supplementary Fig. 2).

Nevertheless, we next considered models which are computationally less costly, basing their evaluation on a an approximation to a full expectation. First, we tested a recent class of sampling models in which integration proceeds by drawing a number of sampled outcomes, according to a sampling distribution, and averaging the rewards of these samples. However, two versions of sampling models, one in which samples are drawn proportional to their probability (Probability Sampling; (Vul, Goodman, Griffiths, & Tenenbaum, 2014)), and one in which they are drawn based on their probability and a difference in utility from that of the safe outcome (Utility Weighted Sampling; (Lieder, Griffiths, & Hsu, 2018; Nobandegani, Castanheira, Otto, & Shultz, 2018)), provided a less parsimonious accounts of the data than a prospect theory model (Probability sampling vs Prospect Theory $\Delta iBIC = 540.17$; Utility Weighted Sampling vs Prospect Theory $\Delta iBIC$ = 512.89; Fig. 2a, Supplementary Figs. 3 and 4), and, like expected value models, both underweighted probabilities relative to rewards in choices.

Because prospect theory still requires a potentially costly integration computation, to examine computationally less costly heuristic approximations, we next turned to a class of models in which probability and reward information are combined additively, rather than integrated multiplicatively (Stewart, 2011). Such models are observed frequently in tasks where outcome probabilities are learned from experience, and are hypothesized to be beneficial toward weighting uncertainty around either comparison (Blain & Rutledge, 2020; Donahue & Lee, 2015; Farashahi et al., 2019; Rouault, Drugowitsch, & Koechlin, 2019; Stewart, 2011; Stewart, Chater, & Brown, 2006). Applied to the current task, the Additive Heuristic model decides by computing two distinct components (Fig. 2c; Methods). A probability information component

computes the relative chances that the choice stimulus will lead to the better versus worse gamble outcome. A reward component computes the reward difference between the gamble reward midpoint and the safe reward. Note that we use the term reward to refer to number of points not only gain trials, but also loss trials, where a loss can be viewed as a negative reward. Importantly, because of how rewards were structured in the task (Fig. 1c), the difference between gamble reward midpoint and safe reward could be computed by considering just the trigger reward, which had higher absolute reward value, and the safe reward. The probability information and reward information components are respectively weighted by parameters, β_{prob} and β_{reward} and then added to a frame (gain or loss) specific intercept to form a choice probability (see Supplementary Fig. 5a for analysis of which parameters should be split between gain and loss trials and Supplementary Fig. 5b for necessity of both reward and probability information components).

This latter model provided both a substantially more parsimonious account than prospect theory ($\Delta iBIC = 86.21$) and a clear algorithmic mechanism for integration that could be examined with neuroimaging. Specifically, the Additive Heuristic model provides two parameters for each participant, β^s_{prob} and β^s_{reward} which quantify the degree to which participant, s, relied on probability and reward information. Thus, we refer to these parameters as "Behavioral Probability Weight" and "Behavioral Reward Weight" and use them as an index against which to compare different strategies for MEG outcome reinstatement to investigate whether they are related. For purposes of comparison to related parameters derived from neural data, from this point we now refer to these respective parameters as $\beta^s_{prob(behavior)}$ and $\beta^s_{rew(behavior)}$.

Behavioral reliance on reward versus probability information are related to distinct patterns of prioritized outcome reactivation

At the group level, participants made use of both the reward and probability components of the Additive Heuristic model (Supplementary Fig. 5b). However, individuals differed substantially in their tendency to rely more or less on either component (Supplementary Fig. 6). We examined whether this variability was driven by tendencies to consider different information when evaluating choices. For example, one way to compute the probability component of the additive heuristic model would be to selectively consider the gamble outcome with higher probability, and then decide whether it was attractive. Note that because of reward structure in the task (Fig. 1c) such attractiveness could be determined without consideration of the other outcomes, but rather by comparison to a fixed threshold. Such a strategy could be beneficial because it could arrive at choices by forgoing consideration of both the gamble outcome with low probability, as well as the safe outcome. Conversely, the reward component could be computed by selectively considering the gamble outcome with higher absolute reward (the trigger outcome) and the safe outcome.

In order to test the hypothesis that individual variation in choice behavior was driven by differences in tendencies in which outcomes to consider, we used MEG data to decode which outcomes participants reactivated during choice deliberation. Specifically, we tested the hypothesis that individuals whose behavior reflected greater consideration of probability information, as indexed by higher Behavioral Probability Weight, would also tend to neurally reinstate gamble outcomes with higher probability. By contrast, we also tested whether

individuals whose behavior reflected greater consideration of reward information (as indexed by higher Behavioral Reward Weight) tended to reinstate gamble outcomes based on the absolute value of their rewards and also the safe outcome for comparison.

To identify neural representations of outcome stimuli, reinstated during choice evaluation, we trained classifiers on data collected prior to the decision making task (Fig. 3a) to predict the identity of each outcome stimulus from MEG sensor data recorded at a particular time-point, τ, following its presentation (Methods). Each classifier was trained to output an estimated quantity, termed "Activation Probability", reflecting the probability that the sensor data reflected reinstatement of the outcome stimulus on which it was trained (Fig. 3b). Previous research has demonstrated that different components of a stimulus representation (Kurth-Nelson et al., 2015), corresponding to activity at different timepoints following stimulus presentation, play different roles in stimulus retrieval. On this basis we trained multiple classifiers separately on data from each 10 ms time bin, τ, following the stimulus presentations. We found that classifiers trained on data from τ = 20 to τ = 500 ms obtained above chance accuracy when tested on held out data from the same timepoint (Fig. 3c). Additionally, such classifiers were selectively accurate when tested on timepoints around the time-points they were trained (Fig. 3d). This selective decoding accuracy enabled us to then investigate which aspects of an outcome's representation are reinstated during choice evaluation.

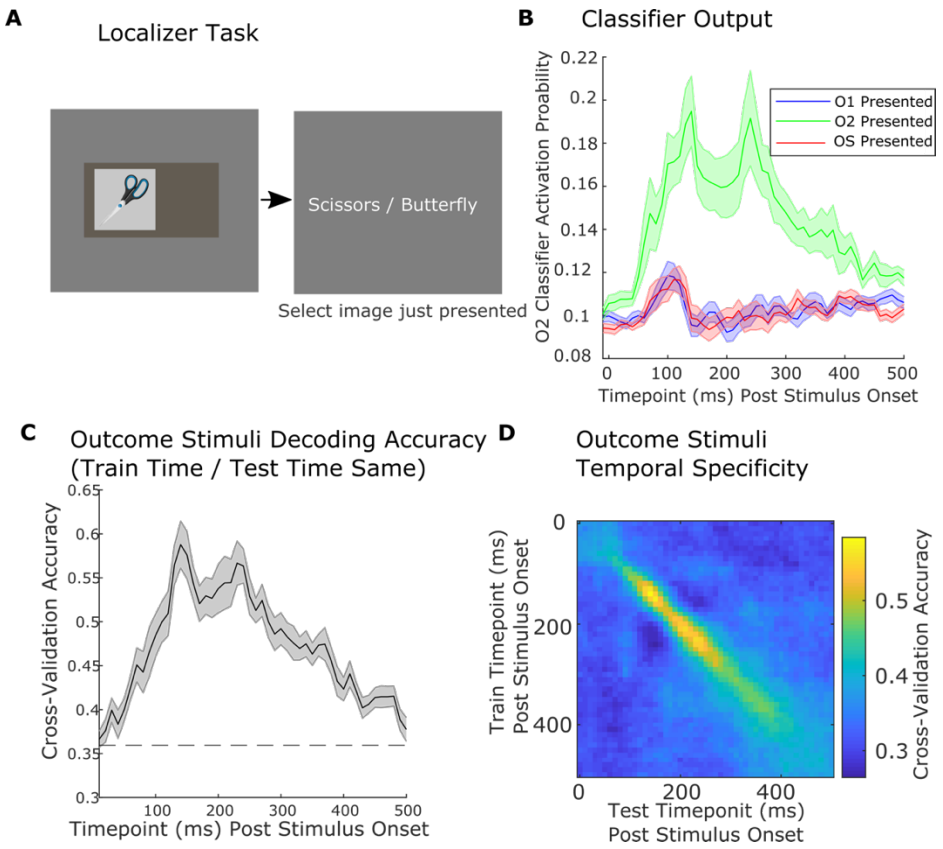


Fig. 3. Decoding outcome stimuli from MEG activity.
a) Localizer Task*. The Localizer task was completed prior to the risky decision task and prior to learning the choice-outcome probabilities. On each trial participants were shown one of the outcome or choice stimuli, and, on the next screen, they then selected a word corresponding to the stimulus they just observed.*

b) Activation Probability Measure. *For each outcome stimulus, we trained lasso-regularized logistic regression classifiers to discriminate MEG data from when that outcome stimulus was presented compared to presentation of other images and inter-trial intervals. Each classifier output an estimated probability that the corresponding stimulus was being presented (Activation Probability). A classifier for each stimulus was trained at each 10 ms bin of MEG sensor between 0 and 500 ms following stimulus presentation. As an example, lines here display the group-mean activation probability measure for the classifier corresponding to O2, for each training timepoint, applied to held out data at the same corresponding test timepoint, where the color designates the true outcome stimulus presented.*
c) Decoding accuracy. *Cross-validation accuracy is the proportion of trials for which the classifier corresponding to the presented outcome (for held-out data) had the highest activation probability. Lines denote mean accuracy (+/- s.e.m.) for each set of 10 ms time-binned outcome classifiers, applied to the same time-bin on held out examples. Dashed line designates permutation threshold corresponding to the 95 percentile peak threshold for accuracy lines generated with shuffled labels.*
d) Temporal specificity. *Classifiers trained on each 10 ms time bin were also tested on every time bin from 0 to 500 ms following presentation of stimuli from held out data. The resulting accuracy image demonstrates temporal selectivity - classifiers identify with good accuracy representations of stimuli that are specific to the timepoint on which they were trained.* ***b-d)*** *Values reflect group means across 19 participants.*

We next turned toward examining which outcome representations were reinstated during choice evaluation, and relating this to behavioral markers of consideration of either probability or reward information. For each training timepoint from 20 to 500 ms, over which we obtained above chance classification, we applied each of the three outcome classifiers to task data from each trial from 0 to 500 ms following the presentation of the probability stimulus (Fig. 4a). This produced, for each trial, for each outcome, a 2-d image (train timepoint, τ, by task/test timepoint, τ'), reflecting the probability that the corresponding outcome representation (at τ), was reactivated at τ' following probability stimulus onset.

We first asked whether participants who relied on probability information tended to prioritize reinstatement of gamble outcomes based on their probability. We computed the difference between the reactivation probability (ΔRP_O) of O1 and O2 ($\Delta_{RP_O}^{s,t,\tau,\tau'}$ for each participant s, trial t, train timepoint, τ, and task timepoint, τ'; Fig. 4b). We then fit a linear model to predict the relative reactivation measure (separately for each s, τ, and τ') as a function of the relative probability for O1 versus O2 indicated by the choice stimulus ($\Delta_{P_O}^{s,t}$; Fig. 4c). The estimate of this effect, $\beta_{prob(neural)}^{s,\tau,\tau'}$ reflects a tendency of a participant, s, to prioritize reactivation of outcome representations (elicited τ following their direct presentation) according to their probability (measured at τ' following probability stimulus presentation; Fig. 4d). We refer to $\beta_{prob(neural)}^{s,\tau,\tau'}$ as Neural Probability Prioritization.

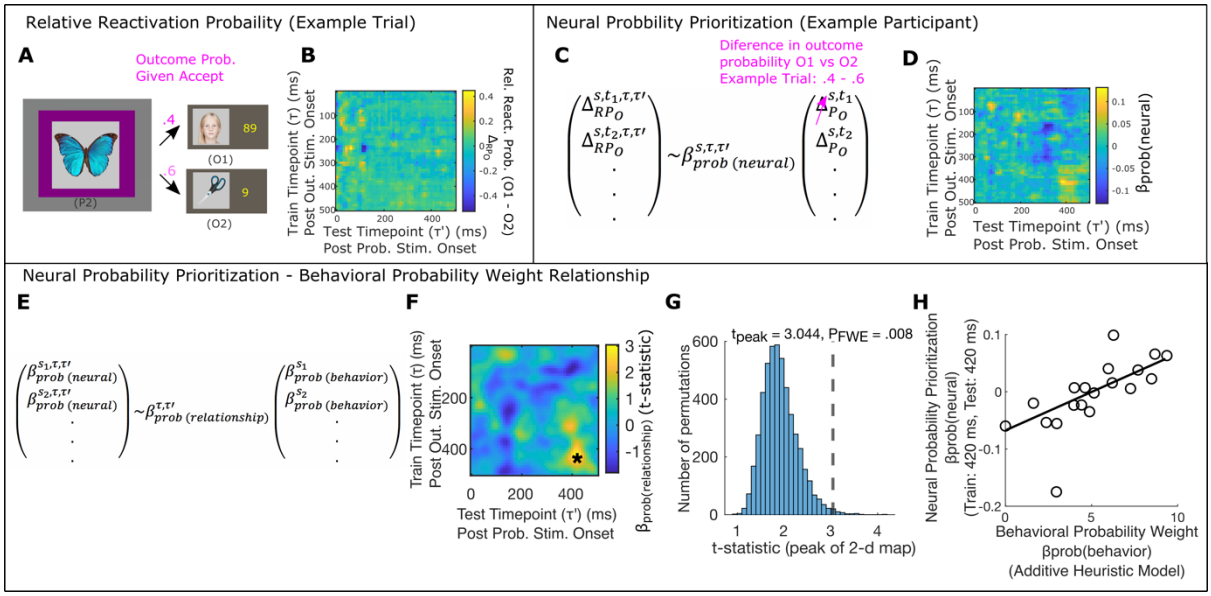


Fig. 4. Behavioral sensitivity to probability information relates to relative reactivation of more probable gamble outcomes following probability stimulus onset*. Neural Probability Prioritization, $\beta^{s,\tau,\tau'}_{prob\ (neural)}$, measures the extent to which relative reinstatement probability of O1 versus O2 changes depending on the probability of encountering O1 versus O2.*

a) Example trial to demonstrate computation of $\beta^{s,\tau,\tau'}_{prob\ (neural)}$. *In this example trial, (Trial 2 from Participant 11), P2 was presented, indicating that, if accepted, O1 would be reached with .4 probability and O2 would be reached with .6 probability.*

b) Example relative activation for O1 versus O2. *Following probability stimulus presentation for each trial, we measure reactivation probability for O1 and O2, $\Delta^{s,t,\tau,\tau'}_{RP_O}$, for τ' = 0 to τ' = 500 ms following probability stimulus onset, for each classifier trained on MEG sensor data from τ = 20 to τ = 500 ms following outcome stimulus onset in the localizer task. Image demonstrates the results of this computation for the example trial in Fig. 4a.*

c) Neural Probability Prioritization, $\beta^{s,\tau,\tau'}_{prob\ (neural)}$, measures tendency to reactivate gamble outcomes according to their probability. *Neural Probability Prioritization, $\beta^{s,\tau,\tau'}_{prob\ (neural)}$, is computed by regressing relative trial-varying reactivation probability of O1 versus O2, $\Delta^{s,t,\tau,\tau'}_{RP_O}$, onto the trial-varying probability of encountering O1 versus O2, $\Delta^{s,t}_{P_O}$ (see Methods).*

d) Neural Probability Prioritization, $\beta^{s,\tau,\tau'}_{prob\ (neural)}$ for example participant. *Image denotes $\beta^{s,\tau,\tau'}_{prob\ (neural)}$ for every classifier train timepoint,τ, following outcome stimulus onset and test timepoint, ,τ', following probability stimulus onset, for an example participant (s = 11).*

e) Measuring relationship between Behavioral Probability Weight and Neural Probability Prioritization. *Following computation of $\beta^{s,\tau,\tau'}_{prob\ (neural)}$ we measured the between-participant relationship between this and behavioral evidence for consideration of probability information, as measured by the behavioral probability weight $\beta^{s}_{prob\ (behavior)}$. This was done by regressing $\beta^{s,\tau,\tau'}_{prob\ (neural)}$ onto $\beta^{s}_{prob\ (behavior)}$, separately for each train and test timepoint, τ and,τ'.*

f-h) Behavioral Probability Weight relates to Neural Probability Prioritization. *f) Image shows t-statistic for this regression (applied to 19 participants), for each train and test*

*timepoint, smoothed with a Gaussian kernel (σ = 1.5 timebins). *P_{FWE} = .009, non-parametric permutation test on image peak. g) Histogram shows null distribution of maximum t-statistics over 5000 2-d maps, each generated by randomly shuffling $\beta^{s}_{prob\,(behavvior)}$ between participants, s. Dashed line shows true maximum t-statistic. h) Raw participant-specific measurements of Neural Probability Prioritization and Behavioral Probability Weight. Neural estimates are taken at peak training and test timepoints. Note this is shown for display purposes only, as raw estimates are biased due to maximization of t-statistic over train and test timepoints.*

To test whether the tendency to reactivate outcomes according to their probability is reflected in behavioral choice sensitivity to outcome probability information, we computed the between participant effect of Neural Probability Prioritization, $\beta^{s,\tau,\tau'}_{prob(neural)}$, on Behavioral Probability Weight, $\beta^{s}_{prob(behavior)}$ (Fig. 4e). The peak of this effect was significantly positive (Figs. 4f-h; τ = 420 ms, τ' = 420 ms; P_{FWE} = .009, non-parametric permutation test on image peak; see Methods; see Discussion for consideration of identified peak significant timepoints), providing evidence supporting the hypothesis that the more an individual's reactivation reflected differences in outcome probabilities, the more that individual showed behavioral evidence of sensitivity to probability information. Importantly, we did not observe a positive relationship between $\beta^{s,\tau,\tau'}_{prob(neural)}$ and $\beta^{s}_{rew(behavior)}$ (Supplementary Fig. 8a).

In a similar manner, we next investigated the reward component, which calls for consideration of the trigger outcome (gamble outcome with higher absolute reward) and safe outcome value (Fig. 2c). Thus, we asked whether individuals who were more sensitive to reward information preferentially reinstate these outcomes. To measure a tendency to reactivate gamble outcomes with higher absolute reward values, we measured the between-trial effect of the difference between the absolute rewards for O1 and O2, $\Delta^{s,t}_{|R_O|}$, on difference in reactivation probability for O1 and O2, $\Delta^{s,t,\tau,\tau'}_{RP_O}$ (Fig. 5a-c). This effect, $\beta^{s,\tau,\tau'}_{rew(neural)}$, Neural Reward Prioritization, measures a participant's tendency to prioritize reactivation of an outcome's representation (at specific τ and τ') based on its trial-varying absolute reward value (Fig. 5d). Regressing $\beta^{s,\tau,\tau'}_{rew(neural)}$ onto Behavioral Reward Weight ($\beta^{s}_{rew(behavior)}$; Fig. 5e), revealed a significant positive effect (Figs. 5f-h; τ = 480 ms, τ' = 110 ms; ; P_{FWE} = .024, non-parametric permutation test on image peak). This association was specific as we did not observe a positive relationship between $\beta^{s,\tau,\tau'}_{rew(neural)}$ and $\beta^{s}_{prob(behavior)}$ (Supplementary Fig. 8b).

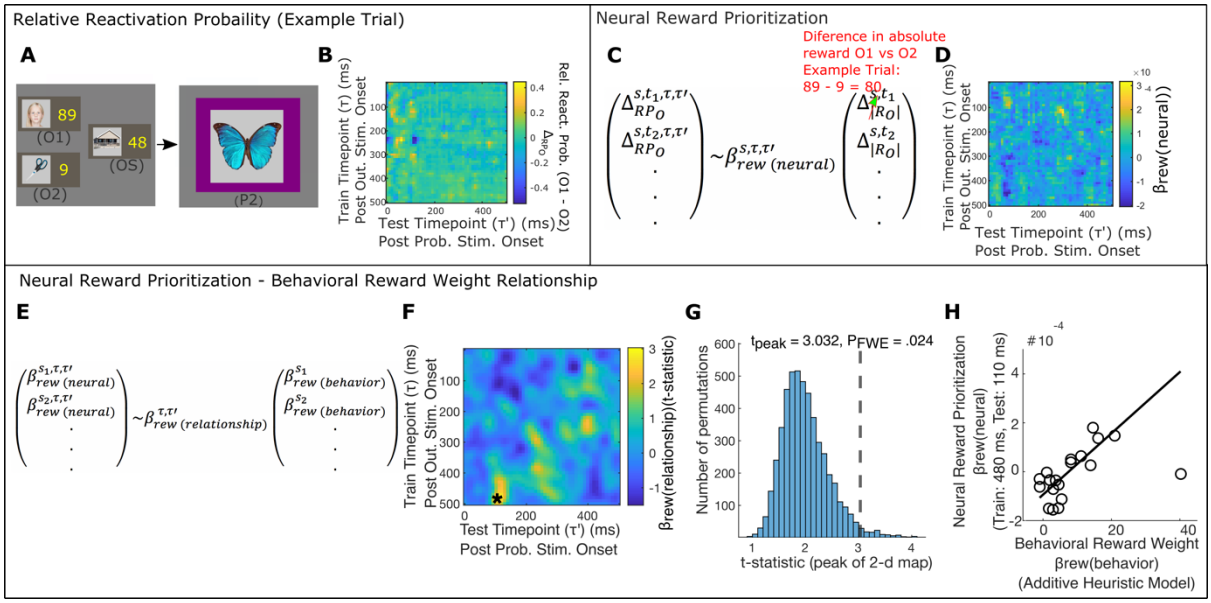


Fig. 5. Behavioral sensitivity to reward information relates to relative reactivation of higher absolute value gamble outcome representation. *Neural Reward Prioritization, $\beta^{s,\tau,\tau'}_{rew\ (neural)}$, measures the extent to which relative reinstatement probability of O1 versus O2 changes depending on the absolute reward paired with of O1 versus O2.*

a) Example trial to demonstrate computation of $\beta^{s,\tau,\tau'}_{rew\ (neural)}$. *On this trial, O1 is paired with 89 points and O2 is paired with 9 points. Note that this is the same trial as in Fig. 4a.*

b) Example relative activation for O1 versus O2, $\Delta^{s,t,\tau,\tau'}_{RP_O}$. *Image displays $\Delta^{s,t,\tau,\tau'}_{RP_O}$ for example trial in 5a. Replotted from Fig. 4b.*

c) Neural Reward Prioritization, $\beta^{s,\tau,\tau'}_{rew\ (neural)}$, measures tendency to reactivate gamble outcomes according to their absolute reward. *Neural Reward Prioritization, $\beta^{s,\tau,\tau'}_{rew\ (neural)}$, is computed by regressing relative trial-varying reactivation probability of O1 versus O2, $\Delta^{s,t,\tau,\tau'}_{RP_O}$, onto the trial-varying difference in absolute points paired with O1 versus O2, $\Delta^{s,t}_{|R_O|}$.*

d) Neural Reward Prioritization, $\beta^{s,\tau,\tau'}_{rew\ (neural)}$ for example participant. *Image denotes $\beta^{s,\tau,\tau'}_{rew\ (neural)}$ for every classifier train timepoint, τ, following outcome stimulus onset, and test time timepoint, τ', following probability stimulus onset, for an example participant (s = 11).*

e) Measuring relationship between Behavioral Reward Integration and Neural Reward Prioritization. *Following computation $\beta^{s,\tau,\tau'}_{rew\ (neural)}$, we measured the between-participant relationship between this and behavioral sensitivity to reward information, as measured by Behavioral Reward Weight, $\beta^{s}_{rew\ (behavior)}$. This was done by regressing $\beta^{p,\tau,\tau'}_{rew\ (neural)}$ onto $\beta^{s}_{rew\ (behavior)}$ separately for each τ and τ'.*

f-h) Behavioral Reward Weight relates to Neural Reward Prioritization. *f) Image shows a t-statistic for this regression (across 19 participants), for each train and task time-bin, smoothed with a Gaussian kernel (σ = 1.5 time-bins). *: P_{FWE} = .024, permutation tested. g) Histogram shows null distribution of maximum t-statistics over 5000 2-d maps, each generated by randomly shuffling $\beta^{s}_{rew\ (behavior)}$ between participants. Dashed line shows true maximum t-statistic. g) Raw participant-specific measurements of Neural Reward Prioritization and Behavioral Reward Weight, at peak train and test timepoints. Note that this is shown for display purposes only as raw estimates are biased due to maximization over train and test timepoints.*

Next, we computed participant specific tendencies to reactivate the safe outcome, $RP_{O_S}^{s,\tau,\tau'}$, as the mean reactivation probability of the safe outcome classifier across trials (Fig. 6a) and regressed this onto Behavioral Reward Weight ($\beta_{rew(behavior)}^{s}$; Fig. 6b). The peak of this effect was also significantly positive (τ = 350 ms τ' = 270 ms; P_{FWE} = .011, non-parametric permutation test on image peak; Figs. 6c-e*)* Hence, the more an individual tended to rely on a simple comparison between the rewards of gamble and safe options, the more they reactivated the high absolute reward and safe outcomes. As with the above, this association was specific as we did not observe a positive relationship between $RP_{O_S}^{s,\tau,\tau'}$ and $\beta_{prob(behavior)}^{s}$ (Supplementary Figs. 8c).

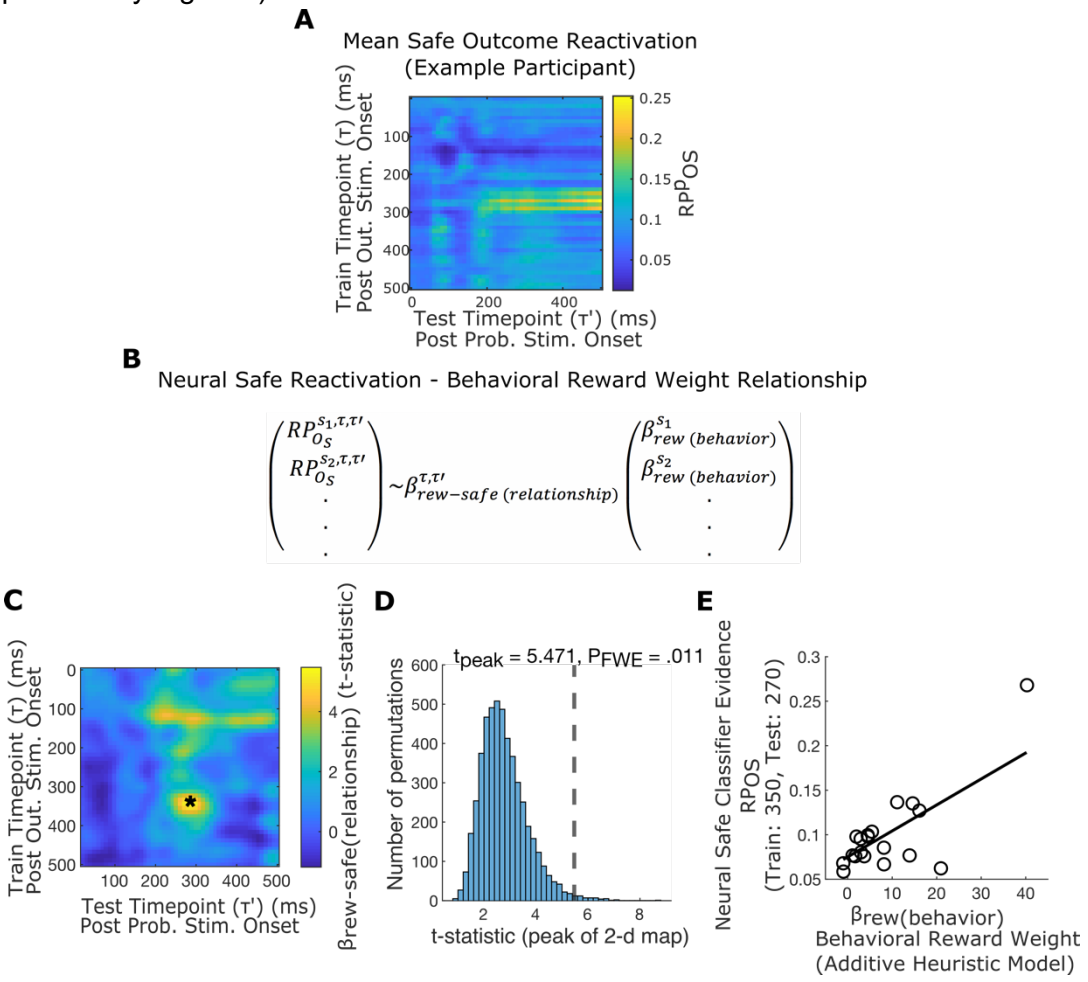


Fig. 6. Behavioral sensitivity to reward information relates to greater reactivation of safe outcome representation. *In order to measure a tendency to reactivate a safe outcome representation, we computed the mean reactivation probability of the safe outcome representation,* $RP_{O_S}^{s,\tau,\tau'}$ *for participant,* s*, train timepoint,* τ*, following outcome stimulus onset and test timepoint,* τ*', following probability stimulus onset.*

a) Safe Reactivation for Example Participant. *Image denotes safe reactivation probability,* $RP_{O_S}^{s,\tau,\tau'}$*, averaged across trials, for each train and task time-point,* τ *and* τ'*, for an example participant, s.*

b) Measuring Relationship Between Safe Outcome Reactivation and Behavioral Reward Integration *In order to measure the between-participant relationship between reactivation of the safe outcome and behavioral integration of reward information into choice, as measured*

by the reward component of the additive heuristic model ($\beta^{s}_{rew\,(behavior)}$), we regressed $\beta^{s}_{rew\,(behavior)}$ onto between participant measure of mean safe reactivation, $RP_{O_S}^{s,\tau,\tau'}$ separately for each , τ and τ'.

c-e) Safe Outcome Reactivation Relates to Behavioral Sensitivity to Reward Information. *c) Image shows a t-statistic for this regression (applied to 19 participants), for each train and task timebin, smoothed with a Gaussian kernel (σ = 1.5 timebins). *: P_{FWE} = .011, non-parametric permutation test on image peak. d) Histogram shows null distribution of maximum t-statistics over 5000 2-d maps, each generated by randomly shuffling $\beta^{s}_{rew\,(behavior)}$ between participants. Dashed line shows true maximum t-statistic. e) Raw participant-specific measurements of Safe Outcome Reactivation and Behavioral Reward Weight, at peak train and test time-points. Note that this is presented for display purposes only, as raw estimates are biased due to maximization of t-statistic.*

Altogether, these results provide evidence that individual differences in outcome representation prioritization for reinstatement drive individual differences in choices. Participants who were behaviorally influenced by probability information were also more likely to reactivate gamble outcomes based on their probability. Conversely, participants who were behaviorally influenced by reward information tended to reactivate gamble outcomes based on their absolute reward, as well as the safe outcome. Hence, whether probability and reward information influence behavior relates to prioritized neural reinstatement of the relevant information.

Prioritized reactivation of high probability outcomes relates to a real-life measure of risky decisions

Aberrant valuation and decision making particularly in risk settings are features of multiple psychiatric disorders (Amlung et al., 2019; Berwian et al., 2020; Deserno et al., 2015; Gillan, Kosinski, Whelan, Phelps, & Daw, 2016; Loewenstein, Hsee, Weber, & Welch, 2001; Mathews & MacLeod, 2005). Based upon the finding above, we hypothesized that aberrant decision making and valuation in the context of behavioral impulsivity tendencies would relate to a lack of selectivity in reinstatement of choice outcomes. Impulsivity is characterized by a predisposition toward risky behavior and a predisposition to act without adequate thought (Eysenck & Eysenck, 1977). Items on the self-report Barratt Impulsivity Scale (BIS) capture a tendency to act without thinking about the likely future consequences of the action (e.g. “I do things without thinking”, “I am more interested in the present than the future”). Impulsivity has also previously been associated with reduced neural signatures of model-based decision making (Deserno et al., 2015), while theoretical models of impulsivity suggest a relationship between it and noisy simulation of action outcomes (Gabaix & Laibson, 2017). Based on this, we specifically hypothesized that aspects of impulsivity would relate to failure to reactivate (consider) outcomes according to their probability. We thus examined the relationship between impulsivity and Neural Probability Prioritization ($\beta^{s,\tau,\tau'}_{prob(neural)}$, Fig. 7a) and identified a significant negative relationship (Figs. 7b-d, $\tau = 410$ ms, $\tau' = 370$ ms, P_{FWE} = .006, non-parametric permutation test on image minimum). This provides evidence that real-life patterns of risky decision making, as expressed through behavioral impulsivity, relate to a failure to prioritize reactivation of more likely choice outcomes.

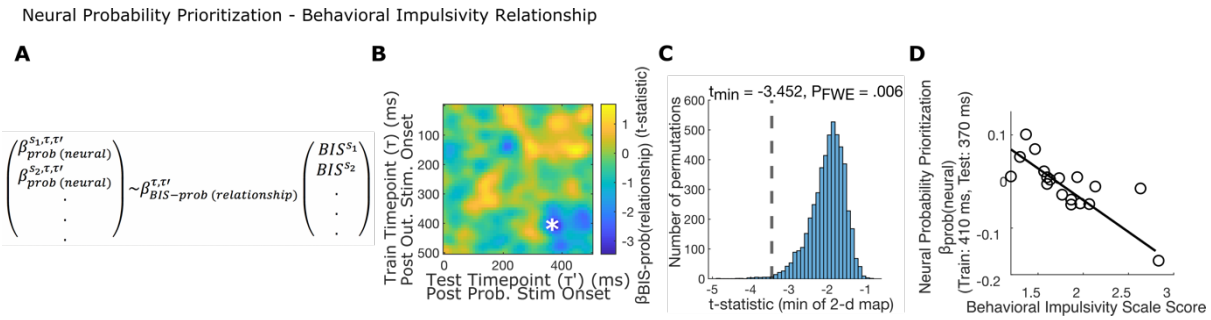


Fig. 7. Relative reinstatement of outcomes with high probability is less in individuals higher in impulsivity. a) Measuring Relationship Between Behavioral Impulsivity and Neural Probability Prioritization. *In order to measure the between-participant relationship between behavioral impulsivity and neural probability prioritization, we regressed between participant neural probability prioritization* $\beta^{p,\tau,\tau'}_{prob\ (neural)}$ *onto Behavioral Impulsivity Scale (BIS) scores, separately for each train and test timepoint (*τ *and* τ'*).*
b-d) Behavioral Impulsivity Relates to Neural Probability Prioritization. *b) Image of t-statistic of relationship (for 18 participants) between Behavioral Impulsivity Scale (BIS) score and neural probability prioritization,* $\beta_{prob(neural)}$*, computed for each train and test timepoint, smoothed with a Gaussian kernel (*σ *= 1.5 time-bins). *:* P_{FWE} *= .006, non-parametric permutation test on image minimum. c) Histogram shows null distribution of minimum t-statistics over 5000 2-d maps, each generated by randomly shuffling* BIS *scores between participants. Dashed line shows true minimum t-statistic. d) Raw participant-specific measurements of neural probability prioritization and BIS score, at minimum train and task timepoints. Presented for display purposes only, as raw estimates are biased due to minimization of t-statistic.*

Discussion

It is widely conjectured that differences in behavioral choice patterns relate to differences in what information individuals consider during evaluation. Here, we examined this phenomenon informed by neural data. Specifically, we used MEG to decode on a fast time scale the content of what information is under consideration, in the context of a risky decision-making task. Our findings are consistent with a hypothesis that underlying individual differences in integration of reward and probability information into choice, in both a laboratory task and in real life, reflect differences in the nature of the information that is considered during evaluation.

Our behavioral analysis revealed that participants adopted a strategy where outcomes were compared along separate reward and probability dimensions which were then additively integrated. However, individuals differed in the extent to which they relied on either reward versus probability components of this heuristic when deciding. By decoding outcome representations using MEG, we identified that underlying these distinct decision strategies were differences in what outcomes were neurally reinstated during evaluation. In particular, participants who decided based on a difference in probability between the better and worse gamble outcomes preferentially reactivated high probability gamble outcomes, suggesting they mainly 'thought' about probability information. Conversely, participants who decided more based on the difference in reward between outcomes preferentially reinstated the safe outcome and high absolute value gamble outcomes, suggesting they mainly considered the

relative points of safe and gamble options. Finally, we observed that individuals higher in Behavioral Impulsivity, a marker of real-world risk taking, demonstrated a relative failure to selectively reinstate outcome based on probability.

Our results fill a key gap in the literature as to what accounts for individual differences of treatment of reward and probability in risky choice. Although individual differences are ubiquitous in the literature of risky choice, the full range of factors that determine individual differences are unknown. Previous modeling approaches have demonstrated that models which selectively integrate either reward or probability information account for some aspects of commonly observed variance in risky choice (Stewart, 2011), though whether such variation is driven by differences in the types of information considered during choice evaluation has not been directly demonstrated. Here, by identifying a link between outcomes that are reinstated during choice evaluation and behavioral signatures that reflect consideration of either reward or probability information, we provide evidence that such variation is indeed driven by differences in what sources of information are considered during evaluation.

One caveat in interpreting our reactivation results is that we only analyzed choice periods of up to 500 milliseconds following choice stimulus presentation. This was necessary because participants made fast responses (Supplementary Fig. 7), thus, limiting the available time window over which activations could be averaged. However, the majority of participant's choice evaluations lasted longer than this time period, suggesting that we only were able to examine reactivation data corresponding to a fraction of possible evaluation time used by participants. One explanation for the success in identifying relationships between reactivation and behavior, despite not including the entire evaluation period, is that outcome consideration at a neural level unfolded immediately upon choice stimulus onset, possibly at stereotyped time-points, and then continued beyond that until a choice was made. Although we were limited in this study to examination of the fastest reactivation measures that cohered across participants, future studies might avail of other methods, such as identification of transitions of reactivation events between stimuli (Liu et al., 2021) so as to aggregate reactivation events across trials that may have different response times, thus availing of all data during evaluation.

It is likely that some of the specifics of our results are attributable to specific features of our task. We found that an additive heuristic, in which participants take a weighted sum of components relating to probability and reward information provided the best account of the data. Additive models of integration have previously been suggested as a hypothesis model of choice integration because, with the right setting of weights, they largely account for the same phenomena as prospect theory models and dovetail with cue combination models used in perceptual research (Stewart, 2011; Stewart et al., 2006). Such models also find use in tasks that require learning of reward probabilities from experience (Blain & Rutledge, 2020; Bongioanni et al., 2021; Donahue & Lee, 2015; Farashahi et al., 2019; Rouault et al., 2019; Stewart, 2011; Stewart et al., 2006). It has been argued that an advantage of such models is that they allow weightings to change based on certainty in either piece of information (Farashahi et al., 2019). In the present task, requirements on memory may have induced uncertainty in both reward magnitudes as well as probabilities, and thus encouraged an additive strategy in the task: the probability over outcomes given choice stimulus had to be recalled on any particular trial and similarly the reward magnitudes associated with each outcome were not shown at the timepoint the choice was made, but rather had to be remembered based on the previous screen's information. An additional advantage of additive

integration models is that they permit comparisons between reward and probability components of outcomes to be computed at distinct time-points. The ability to make distinct comparisons at different timepoints may have been specifically incentivized in this task due to the particular staging of gamble information, and our neural results with regards to when reward and probability component computations occurred suggest that participants exploited this.

Our results build on prior work investigating reinstatement of outcomes during choice evaluation. A number of studies have investigated outcome reinstatement in the context of model-based reinforcement learning algorithms (Bornstein & Daw, 2013; Castegnetti et al., 2020; Doll, Duncan, Simon, Shohamy, & Daw, 2015; Russek, Momennejad, Botvinick, Gershman, & Daw, 2021; Wimmer & Büchel, 2019; Wise et al., 2021). Typical model-based algorithms postulate that choices be evaluated by simulating potential consequent outcomes and adding the rewards of those outcomes to a running average (Sutton, 1991; Sutton & Barto, 2017). Evidence that outcome reinstatement functions to simulate outcomes in this manner comes from studies demonstrating that variation in a tendency to reactivate the deterministic outcome of a chosen action predicts a propensity for behavior to reflect model-based choice evaluation (Doll et al., 2015; Wise et al., 2021). This mechanism for outcome reinstatement has also accounted for within subject variation relating what is simulated to ultimate valuation of a choice option (Castegnetti et al., 2020; Russek et al., 2021). Our results add to this work by revealing that outcome reactivation can support functions beyond typical model-based simulation of outcomes, such as comparison of reward values between choice outcomes (as used in the reward component of the choice model identified here). Furthermore, our results show that individual variation in reactivation tendencies relate to individual differences in choice. These results thus point toward a more general flexibility in the computational function of outcome reactivation, and emphasize a close link between the processes determining reactivation and ultimate behavior.

Relatedly, a recent body of work has examined how the brain solves the meta-decision problem as to which potential outcomes of a choice should be simulated. Although standard formulations of simulation in model-based choice postulate that outcomes should be simulated proportionally to their probability (Sutton, 1991), theoretical analyses have demonstrated that in situations where the total number of simulations is limited, it is possible to arrive at more accurate estimates of choice utility by considering outcome utilities in the decision of what to simulate (Lieder et al., 2018; Nobandegani et al., 2018). In an MEG neuroimaging study, a tendency to reactivate outcomes proportionally to their absolute utility has been reported (Castegnetti et al., 2020). Importantly, however, this is based on a framework of evaluating choices by sampling outcomes according to some distribution and averaging rewards. How to prioritize which outcomes to represent in situations where choice strategies other than sampling are used, as was the case in the current study, as well as the situations the brain might rationally ‘choose’ to deploy such strategies, remains an open question for future investigation.

Our use of MEG rather than FMRI permitted an analysis of not only which outcomes were reactivated, but also the temporal structure of when such reactivations occurred as well as what temporal component of a representation, measured in terms of time following direct presentation of the stimulus, was reactivated. Such temporal structure has previously been demonstrated as important for integration of rewards with non-directly paired stimuli in a

sensory pre-conditioning task (Kurth-Nelson et al., 2015). Our findings as to when reactivation events occurred bear multiple similarities to the key results of that previous study. Notably, our identification of reinstatement related to integration of reward and probability information, occurred at two distinct timepoints (110 ms and 420 ms following choice stimulus onset), approximately resembles time-points (Kurth-Nelson et al., 2015) when a non-directly rewarded stimulus was re-activated either following a paired stimulus onset (400 ms) and a reward (70 ms). We speculate that such similarities in the timing of reinstatement events may relate to oscillatory events which, triggered by the onset of a choice stimulus, coordinate reactivation (Vidaurre, Cichy, & Woolrich, 2021). The function of such timings should be explored in future work.

Finally, we identified that participants with higher behavioral impulsivity demonstrated relatively reduced prioritized reactivation of higher probability outcome representations. This reactivation result matches nicely recent theoretical proposals that impulsive choice may result from noisy simulation of future events (Gabaix & Laibson, 2017). Given the separate, positive relationship of this pattern of reinstatement with integration of probability information, this result also points toward a mechanism underlying real-life aberrant risky choice, potentially driven by neglect of probability information (Rouault et al., 2019). More generally, our finding here opens up a line of research that certain of disorders of choice may in fact reflect disorders related to what information should be prioritized.

In summary, we demonstrate a relationship between the information that individuals tend consider during evaluation, and how they decide. Moving forward, this suggests the possibility that one could learn to make better choices by learning to change how they reinstate information. This possibility points toward further research into the treatment of mental health disorders characterized by aberrant choice.

Methods

Participants

We recruited 21 participants (mean (std) age: 23.67 (4.33), 13 female) from University College London subject databases who provided informed consent prior to beginning the study. 13 were female. The mean age was 23.67, with a range of 18 to 36. Based on consideration from prior literature, we chose a sample of 30 participants, however, due to the coronavirus pandemic and the UK lockdown, we were required to stop collecting data at 21 participants. Two participants were removed from analysis for choosing the same action on greater than 80% of trials (89% and 83%), thus leaving 19 participants included in the main analysis (Figs. 2 - 6). We additionally failed to collect questionnaire data for one participant. Thus the neural-questionnaire analysis (Fig. 7) reflects data from 18 participants. Although this number of subjects is less than intended, we note that it is within the range for similar studies in the field (Doll et al., 2015; Momennejad, Otto, Daw, & Norman, 2018; Park, Miller, & Boorman, 2021; Wimmer & Shohamy, 2012). For completing the entire study, participants were paid 40 GBP with a performance dependent bonus of up to 20 GBP. This study was approved by UCL ethics (ID: 9929/002).

Experimental Procedures

Training Session and Task Session.

The entire task took place over two consecutive days. On day 1, subjects completed the task instructions. Following this, using different stimuli than used in the actual task, participants completed the entire probability learning task, and then completed three randomly selected blocks from the risky decision making task. Following this they completed a number of Questionnaires.

On day 2, in the MEG scanner, participants completed the functional localizer task, the probability learning task, and the gamble task. Different task stimuli were used on Day 1 and Day 2.

Task overview

In the main task, subjects were required to make decisions about whether to accept or reject a gamble. Rejecting the gamble led to collecting a safe outcome (OS). Accepting, in contrast, led to collecting one of two gamble outcomes (O1 or O2). On each trial, each of the three outcomes were associated with a distinct number of points, which the participant was made aware of at the start of the trial, and which, if collected, contributed toward a bonus. The task contained four probability stimuli (P1, P2, P3 and P4). Each probability stimulus determined, whether, if accepting the gamble, the probability that O1 versus O2 would be encountered. The probability of gamble acceptance leading to O1 was were .2, .4, .6, and .8, for P1, P2, P3 and P4 respectively (Fig. 1b).

The task consisted of eight blocks, which alternated between gain and loss blocks (four of each). To construct each trial, either O1 or O2 was selected to be the trigger option. The reward value of the trigger option was selected from {47.5, 60, 75} on gain trials, or {-47.5, -60, -75} on loss trials, and the non-trigger option value was 0 (Fig. 1c). The value of the safe option was selected from {20, 40, 60, 80} on gain trials and {-20, -40, -60, -80} on loss trials. Following this, a single random value drawn from uniform(0,25) for gain trials, or uniform(-25,0) for loss trials was added to each outcome. Finally, three separate random values drawn from uniform(0,5) for gain trials and uniform(0,-5) for loss trials were added to each value separately.

Trials consisted of each combination of trigger value, and safe value, such that the absolute value of the trigger value was greater than the absolute value of the safe value, for both O1 and O2 occurring as the trigger value, for each level of P(O1|Cn). Finally, each exact trial repeated twice in the task.

Participants were instructed that their bonus would be computed by randomly selecting one trial from each block of the task and adding the points they collected on these trials. The bonus was proportional to this sum.

Functional Localizer

Each task stimulus was represented using a decodable visual stimulus. Our analysis of the task relied on decoding from MEG data what outcome stimulus was represented during choice

evaluation. In order to collect data with which to train a classifier to detect stimulus representations, participants completed a functional localizer task, consisting of three blocks. each block, the seven images representing each task state were each presented 20 times, in randomized order. For each presentation, the image was presented for 800 ms. Following a 200 ms ISI, two words appeared on the screen, one corresponding to the name of the image just presented and one corresponding to the name of a different image. Participants were given 600 ms to select the word corresponding to the image just seen.

Probability Learning Task

In order to learn the probabilities that each choice stimulus, if accepted, led to either gamble outcome stimulus, participants completed four blocks of a probability learning task. In each block, for each probability stimulus, participants were first shown a screen instructing them on the probabilities that that probability stimulus (if as part of a gamble that was accepted) would lead to either gamble outcome stimulus. Following this, the participants experienced 10 trials in which they were required to "play" that probability stimulus. For each play, the participant experienced that stimulus, followed by one of the two gamble outcomes. For the 10 trials, it was guaranteed that the number of either outcome experienced matched the instructed probability, however in randomized order (e.g. if the probability stimulus led to O1, 40% of the time, the participant experienced O1 4 out of the 10 times following the choice stimulus). In order to ensure attention, following 25% of these trials, participants were required to report either which choice stimulus, or which outcome stimulus they had just experienced. After experiencing two rounds of instructed probabilities and experienced transitions for each probability stimulus, the participant was then required to respond to number of queries about the probability that each probability stimulus led to each outcome. For each query, the participant was shown an image of one of outcome stimuli as well as two of the probability stimuli, and was required to report which of the two probability stimuli was more likely to lead to that outcome. The proportion correct for these queries across rounds is reported in Supplementary Fig. 1.

Risky Decision Making Task

On each trial of the risky decision maknig task, participants were first shown how many points would be earned if they were to encounter either of the three types of outcomes (O1, O2 or OS). This was displayed on a screen, presented for 2.5 s, containing three separate banknote-like images, with each banknote containing one of the outcomes and the number of points (Fig. 1a). The position of the two gamble outcomes was randomly counter-balanced. Following a 1.5 s ISI, participants were then presented with one of the four probability stimuli, and were required to either accept or reject the gamble. Rejecting the gamble would lead to encountering the safe stimulus and collecting the number of points associated with it for that trial. Conversely accepting the gamble would lead to encountering either O1 or O2, and collecting the number of points associated with that outcome for that trial. The probability stimulus remained on the screen until the subject made a response, up to a maximum of 6 s. Then, following a 1.5 s ISI participants observed a banknote corresponding to the outcome they received, along with the number of points they collected. In order to encourage participants to decide at the time of probability stimulus onset, on 10% of trials, participants were not presented with a probability stimulus, and were instead required to report the reward paired with one of the outcome stimuli.

Questionnaires

Participants completed the following questionnaire: The Barratt Impulsivity Scale, The State-Trait Anxiety Inventory (STAI), the Penn State Worry Questionnaire (PSWQ), and the MASQ anhedonia scale. Prior to administering the task, we expected that we would identify differences in how subjects treated loss and gain blocks of the task, and that this difference would be relevant for relating to the STAI, PSWQ. However, after failing to observe relevant behavioral differences in this regard, we focused only on the BIS measure and MASQ. We hypothesized that BIS would be related negatively probability prioritization. We additionally tested whether MASQ would relate negatively to reward prioritization, however did not observe this effect to be significant. Because these were planned comparisons, we do not present correction for multiple comparisons (across multiple tests), however, we note that the strength of the effect relating BIS to neural probability prioritization would survive Bonferroni correction for the two tests performed. Note that, due to an error in recording data, we failed to collect questionnaire data for one participant. Thus, Neural-Questionnaire analysis was examined for 18 participants.

Computational models of choice data

All behavioral analysis was implemented using the Julia (version 1.5) programming language (Bezanson, Edelman, Karpinski, & Shah, 2014). In order to gain an algorithmic description of subjects decision making we fit a number of computational models to their choices. We describe each model here.

Expected value: The expected value model decides based on the difference in expected value for accepting and rejecting the gamble,

$$\mathrm{P}_{accept} = logit^{-1}(\beta[P_{O1}R_{O1} + P_{O2}R_{O2} - R_{OS}])$$

where the inverse temperature parameter, β is a free parameter.

Prospect theory: The prospect theory model (Kahneman & Tversky, 1979) allows expectations to be taken using a probability weighting function, w, and subjective utility function, v,

$$\mathrm{P}_{accept} = logit^{-1}(\beta[w(P_{O1})v(R_{O1}) + w(P_{O2})v(R_{O2}) - v(R_{OS})])$$

We used standard utility functions, and the Prelec probability distortion functions (Prelec, 1998b), $v(x) = x^{\alpha_{gain}}$ when $x \geq 0, v(x) = -(x^{\alpha_{loss}})$ when $x < 0$, and, $w(p) = \exp[-\delta(-\ln p)^a]$. β, α_{gain}, α_{loss}, δ, and a are free parameters.

Additive Heuristic: The additive heuristic model, based on additive integration models (Farashahi et al., 2019; Stewart, 2011), yet adapted for features of this task, simply does a linear integration of two features: one related to the probability of reaching the better outcome, and one related to the difference in reward between the trigger outcome:

$$P_{accept} = logit^{-1}(\beta_0 + \beta_{prob}[P_{O_{better}} - P_{O_{worse}}] + \beta_{rew}[\frac{R^*_{O_{trig}}}{2} - R^*_{O_{Safe}}])$$

where $\beta_0 = \beta_{gain,}$ on gain trials and $\beta_0 = \beta_{loss}$ on loss trials. $\beta_{gain,}$ β_{loss} , β_{prob}, and β_{rew} are free parameters. $R^*_{O_{trig}}$is the reward of the trigger outcome, baseline corrected such that the common noise added to each item is subtracted (Fig. 1c). $R^*_{O_{safe}}$is the reward of the safe outcome, baseline corrected such that the common noise added to each item is subtracted.

Sampling models (Probability Sampling and Utility Weighted Sampling) According to our sampling models, the participant uses importance sampling to estimate the difference in utility between accepting and rejecting the gamble outcome. Both models assume first select a number of samples to take, S, which we assume is drawn from an ordered probit distribution, $OrderedProbit(S \mid n, c)$, where n, which sets the center of the distribution is a free parameter, and the scale parameter, c is set to 2. Following this, the participant draws S samples where each sample corresponds to either O_1 or O_2, from the distribution $q(O_i)$. Given S samples, the subject computes an estimate of the value difference between the gamble option and safe option.

$$\hat{E} = \frac{1}{\sum_{j=1}^{S} w_j} \sum_{i=1}^{S} w_i [v(R_{o_i}) - v(R_{o_{safe}})]$$

where w_i reflects the importance weights, $w_i = \frac{P_{O_i}}{q(o_i)}$, v is defined the same as it is for the prospect theory models, with two free parameters, α_{gain}, and α_{loss}.

The participant's probability of accepting is then 1 if $\hat{E} > 0$, 0 if $\hat{E} < 0$ and .5 if $\hat{E} = 0$. We define $\hat{E}$ as a function of the number of samples taken S, and the number of samples drawn as O_1, n_{O_1},

$$\hat{E}(n_{O_1}, S) = \frac{1}{n_{O_1} w_1 + (1 - n_{O_1}) w_2} [n_{O_1} w_{O_1} \left[v(R_{o_1}) - v(R_{o_{safe}})\right] + [1 - n_{O_1}] w_{O_2} \left[v(R_{o_2}) - v(R_{o_{safe}})\right]]$$

Then the probability of acceptance then marginalizes over the number of samples taken, S, as well as the number of samples drawn as O1 n_{O_1}:

$$P_{accept} = \sum_{S=1}^{\max S} OrderedProbit(S|n, c) \sum_{n_{O_1}=0}^{n_{O_1}=S} Binomial(n_{O_1}, s, q(O_1)) P(accept| \hat{E}(n_{O_1}, S))$$

where we took the maximum number of samples, $\max S$, to be 7. Here, we assume the number of samples taken, S, is selected from an Ordered Probit distribution, with scale parameter, c = 2, and center parameter, n, a free parameter.

We considered two sampling models, which differ with regards to the sampling distribution $q(O_i)$. For probability sampling, $q(O_i) \propto P_{O_i}$ (Vul et al., 2014). For utility weighted sampling,

$q(O_i) \propto P_{O_i}\left|v(R_{O_i}) - v(R_{Safe})\right|$ (Lieder et al., 2018). Both models have a 3 free parameters: n, α_{gain}, and α_{loss}.

Model fitting: For each participant, we estimated the free parameters of each model by maximizing the likelihood of choices, jointly with group-level distributions over the entire population using an Expectation Maximization (EM) procedure (Huys et al., 2011). Models were compared by computing the integrated Bayesian information criterion over the entire group of subjects for each model. In order to compare model predictions to data points, we computed for each trial, for each participant, the probability of acceptance under that participant's best fitting parameters.

MEG acquisition

MEG data was acquired on a CTF 275-channel axial gradiometer system (CTF Omega, VSM MedTech) sampling at 1200Hz. The task was divided into multiple scanning sessions, with each session lasting less then 10 minutes. Participants were asked to remain still during the scanning session but were able to take a rest between sessions. At the start of each scanning session, participants were asked to move back to where they were, and their head positions were registered.

MEG analysis

All MEG analyses were completed using custom Matlab (version 2019a) scripts.

Pre-processing

Preprocessing was done using OSL (OHBA Analysis Group, OHBA, Oxford, UK). Preprocessing steps included high-pass filtering, at 0.5 Hz, followed by down sampling to 100 Hz. After identification and removal of excessively noisy segments and sensors, independent component analysis (ICA) was applied to denoise the data. In each scanning session, up to 10 independent components could be excluded if they were marked as noise based on spatial topography, time course, kurtosis of the time course and frequency spectrum. Following this, all analyses were performed on filtered MEG data at the whole-brain sensor level.

Decoding Analysis Training Classifiers

Data from the functional localizer task was epoched between 0 and 500 milliseconds following stimulus onset. We trained binary classifiers on data from the functional localizer task. For each 10 ms timepoint following stimulus onset, three binary classifiers were trained, one for each outcome stimulus to discriminate between sensor data associated with that stimulus, and sensor data associated with each of the 6 other stimuli, along with null data corresponding to the intertrial interval (equal in number to 100% of training examples). The classification pipeline consisted of scaling the data by dividing by its 95th (absolute) percentile. Following this, data from all sensors for a given timepoint was used as training examples to train a lasso logistic regression classifier (using matlab function lassoglm). Figs. 3b-d were generated by doing a 7-fold cross validation, the three classifiers training on each time-point (out of 50) using 6/7 of the training data and then testing using remaining 1/7 examples on each timepoint.

The regularization hyperparameter of the logistic regression selected as the parameter which maximized the mean cross validation accuracy along the diagonal of the 2-D map in Fig. 3d (matching train and test timepoints). A given test example was considered correct if its classifier had the highest activation (out of the three). This identified .002 as the best regularization parameter, which was used for further analysis.

Outcome reactivation analysis

After choosing a lasso penalty, we trained the three classifiers on all the localizer data, on each timepoint, τ, following outcome stimulus onset. This generated three classifiers, one for each outcome, for each of 50 timepoints, τ, following outcome stimulus presentation in the localizer task. Given the task response times (Supplementary Fig. 7), we epoched the decision making task data from 0 to 500 ms following the onset of the probability stimulus in each trial. Additionally, we removed all trials that had response times faster than this so as to only examine data involved in deliberation. We then applied each outcome classifier, for each training timepoint, τ, to each task timepoint, τ', following probability stimulus onset. We use $RP_{O_x}^{p,t,\tau,\tau\prime}$ to represent the reactivation probability output by the classifier, trained to activate for stimulus OX (either O1, O2, or OS) at timepoint τ ms following its presentation, for participant s, on trial t, at timepoint τ' following presentation of the probability stimulus.

Relating reinstatement of gamble outcomes to behavioral measures of reward and probability consideration. In order to examine the question of how prioritization of reactivated outcomes relates to behavioral evidence for consideration of probability versus reward information, we used a two-stage analysis. In the first stage, we fit, separately, for each participant, s, train timepoint τ, and test timepoint τ', a linear model to predict the difference in reactivation probabilities between the two gamble outcomes, $\Delta_{RP_O}^{s,t,\tau,\tau\prime} = RP_{O_1}^{s,t,\tau,\tau\prime} - RP_{O_2}^{s,t,\tau,\tau\prime}$.

We predict this difference as a function of the participant and trial specific difference in probability, $P_{O_1}^{s,t} - P_{O_2}^{s,t}$, as well as absolute rewards, $|R_{O_1}^{s,t}| - |R_{O_2}^{s,t}|$ between the two gamble outcomes:

$$\Delta_{RP_O}^{s,t,\tau,\tau\prime} \sim \beta_0 + \beta_{prob(neural)}^{p,\tau,\tau\prime}\left[P_{O_1}^{s,t} - P_{O_2}^{s,t}\right] + \beta_{rew(neural)}^{s,\tau,\tau\prime}[|R_{O_1}^{s,t}| - |R_{O_2}^{s,t}|]$$

This provides an estimate of $\beta_{prob(neural)}^{s,\tau,\tau\prime}$, and $\beta_{rew(neural)}^{s,\tau,\tau\prime}$, for each participant, s, train timepoint, τ, and test timepoint, τ'.

In the second level, in order to investigate whether prioritization of outcomes related to behavioral consideration of reward and probability information, we relate $\beta_{prob(neural)}^{s,\tau,\tau\prime}$ and $\beta_{rew(neural)}^{s,\tau,\tau\prime}$ to fitted parameters from the best fitting behavioral model, the Additive Heuristic model, β_{prob} and β_{reward}, which we now refer to as $\beta_{prob(behavior)}^{s}$ and $\beta_{rew(behavior)}^{s}$. We predicted that behavioral consideration of probability information, indexed by $\beta_{prob(behavior)}^{s}$ would be related to selective reinstatement of more probable gamble outcomes, as indexed by $\beta_{prob(neural)}^{s,\tau,\tau\prime}$, and that behavioral consideration of reward information, indexed by $\beta_{rew(behavior)}^{s}$ would be related to selective reinstatement of outcomes with higher absolute reward, as indexed by $\beta_{rew(neural)}^{s,\tau,\tau\prime}$.

We performed two between participant regressions: one relating $\beta^{s}_{prob(behavior)}$ to $\beta^{s,\tau,\tau\prime}_{prob(neural)}$ (Fig. 4) and one relating $\beta^{s}_{rew(behavior)}$ to $\beta^{s,\tau,\tau\prime}_{rew(neural)}$ (Fig. 5). In order to mitigate the impact of potential outliers, following previous work (Eldar, Bae, Kurth-Nelson, Dayan, & Dolan, 2018), all between-subject regressions and associated t-statistics were computed using robust linear regression, (Matlab function robustfit, with default settings). Note that this approach has been shown to both increase power and reduce false positive rates in the presence of outliers (Wager, Keller, Lacey, & Jonides, 2005). Additionally note that significance (p-values) of computed t-statistics were computed by non-parametric permutation test, thus additionally ensuring appropriate false positive rates. Specifically, each between participant regression was applied separately for each train timepoint, τ and test timepoint, τ', thus providing a 2-d map (τ by τ') of t-statistics for each regression. Following (Kurth-Nelson et al., 2015), this map was then smoothed with a Gaussian kernel (σ = 1.5 time bins). Significance for each between participant regression was computed over the peak (max) t-statistic of this smoothed map by non-parametric permutation test (Kurth-Nelson et al., 2015). For this, the 2-d map was re-computed 5000 times, each time shuffling which participant was assigned to which behavioral parameter (e.g. assigning the behavioral parameter for participant 11, $\beta^{s_{11}}_{prob(behavior)}$, to participant 15) according to a random permutation. A null distribution over t-statistics was created by taking the peak of each of the 5000 t-statistic maps (over τ and τ'). Family wise error corrected p-values (P_{FWE}) were computed as the proportion of permutations less than the peak of the true observed map.

Reinstatement of safe outcome. Because the behavioral reward weight in the additive heuristic model requires comparison of the gamble outcome with higher reward absolute value to the safe outcome, we also predicted that behavioral reward consideration would be related to reinstatement of the safe outcome (Fig. 6). As a measure of safe outcome reinstatement, we computed, for each participant, s, train timepoint, τ, and test timepoint, τ', the mean reinstatement probability across trials, $RP^{s,\tau,\tau\prime}_{O_S}$. We then related this to $\beta^{s}_{rew(behavior)}$ and computed significance equivalently as was done for the above between participant regressions.

Relating neural probability prioritization to behavioral impulsivity. In order to behavioral reinstatement to tendency to reinstate outcomes based on their probability (Fig. 7), we repeated the previous between participant regression involving $\beta^{p,\tau,\tau\prime}_{prob(neural)}$, however replacing the $\beta^{s}_{prob(behavior)}$ with the BIS score of participant p. Significance of this regression was computed equivalently to the above, except here P_{FWE} value was computed as proportion of permutations less than the observed minimum (since a negative effect was predicted).

References

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., … McCabe, R. E. (2019). Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. *JAMA Psychiatry*, *76*(11), 1176–1186. https://doi.org/10.1001/JAMAPSYCHIATRY.2019.2102

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23. https://doi.org/10.2307/1909829

Berwian, I. M., Wenzel, J. G., Collins, A. G. E., Seifritz, E., Stephan, K. E., Walter, H., & Huys, Q. J. M. (2020). Computational Mechanisms of Effort and Reward Decisions in Patients with Depression and Their Association with Relapse after Antidepressant Discontinuation. *JAMA Psychiatry*, *77*(5), 513–522. https://doi.org/10.1001/jamapsychiatry.2019.4971

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2014). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. Retrieved from http://arxiv.org/abs/1411.1607

Blain, B., & Rutledge, R. B. (2020). Momentary subjective well-being depends on learning and not reward. *ELife*, *9*, 1–27. https://doi.org/10.7554/eLife.57977

Bongioanni, A., Folloni, D., Verhagen, L., Sallet, J., Klein-Flügge, M. C., & Rushworth, M. F. S. (2021). Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature*, *591*(7849), 270–274. https://doi.org/10.1038/s41586-020-03115-5

Bornstein, A. M., & Daw, N. D. (2013). Cortical and Hippocampal Correlates of Deliberation during Model-Based Decisions for Rewards in Humans. *PLoS Computational Biology*, *9*(12), e1003387. https://doi.org/10.1371/journal.pcbi.1003387

Castegnetti, G., Tzovara, A., Khemka, S., Melinščak, F., Barnes, G. R., Dolan, R. J., & Bach, D. R. (2020). Representation of probabilistic outcomes during risky decision-making. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-16202-y

Deserno, L., Wilbertz, T., Reiter, A., Horstmann, A., Neumann, J., Villringer, A., … Schlagenhauf, F. (2015). Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. *Translational Psychiatry 2015 5:10*, *5*(10), e659–e659. https://doi.org/10.1038/tp.2015.139

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*(5), 767–772. https://doi.org/10.1038/nn.3981

Donahue, C. H., & Lee, D. (2015). Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nature Neuroscience*, *18*(2), 295–301. https://doi.org/10.1038/nn.3918

Eldar, E., Bae, G. J., Kurth-Nelson, Z., Dayan, P., & Dolan, R. J. (2018, September 7). Magnetoencephalography decoding reveals structural differences within integrative decision processes. *Nature Human Behaviour*. Nature Publishing Group. https://doi.org/10.1038/s41562-018-0423-3

Eysenck, S. B. G., & Eysenck, H. J. (1977). The place of impulsiveness in a dimensional system of personality description. *British Journal of Social and Clinical Psychology*, *16*(1), 57–68. https://doi.org/10.1111/j.2044-8260.1977.tb01003.x

Farashahi, S., Donahue, C. H., Hayden, B. Y., Lee, D., & Soltani, A. (2019). Flexible combination of reward information across primates. *Nature Human Behaviour*, *3*(11), 1215–1224. https://doi.org/10.1038/s41562-019-0714-3

Gabaix, X., & Laibson, D. (2017). Myopia and Discounting. Retrieved from http://www.nber.org/papers/w23254

Gigerenzer, G., & Goldstein, D. G. (2011). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Heuristics: The Foundations of Adaptive Behavior*. https://doi.org/10.1093/acprof:oso/9780199744282.003.0002

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goaldirected control. *ELife*, *5*(MARCH2016), 1–24. https://doi.org/10.7554/eLife.11305

Gonzalez, R., Wu, G., Brenner, L., Griffin, D., Heath, C., Klayman, J., … Wakker, P. (1999).

On the Shape of the Probability Weighting Function. *Cognitive Psychology*, *38*, 129–166. Retrieved from http://www.idealibrary.comon

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Computational Biology*, *7*(4). https://doi.org/10.1371/journal.pcbi.1002028

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R., & Dayan, P. (2015). Temporal structure in associative retrieval. *ELife*, *2015*(4). https://doi.org/10.7554/eLife.04919

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X1900061X

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of Extreme Events in Decision Making Reflects Rational Use of Cognitive Resources. *Psychological Review*, *125*(1), 1–32. https://doi.org/10.1037/rev0000074

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science*, *372*(6544), eabf1357. https://doi.org/10.1126/science.abf1357

Loewenstein, G. F., Hsee, C. K., Weber, E. U., & Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Mathews, A., & MacLeod, C. (2005). Cognitive Vulnerability to Emotional Disorders. *Annual Review of Clinical Psychology*, *1*(1), 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *ELife*, *7*. https://doi.org/10.7554/eLife.32548

Nobandegani, A. S., Castanheira, K. da S., Otto, A. R., & Shultz, T. R. (2018). Over-representation of Extreme Events in Decision-Making: A Rational Metacognitive Account. In *Proc. of the 40th Annual Conference of Cognitive Science Society* (pp. 2394–2399). Retrieved from http://arxiv.org/abs/1801.09848

Park, S. A., Miller, D. S., & Boorman, E. D. (2021). Inferences on a multidimensional social hierarchy use a grid-like code. *Nature Neuroscience*, *24*(9), 1292–1301. https://doi.org/10.1038/s41593-021-00916-3

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive Strategy Selection in Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/0278-7393.14.3.534

Prelec, D. (1998a). The Probability Weighting Function. *Econometrica*, *66*(3), 497. https://doi.org/10.2307/2998573

Prelec, D. (1998b). The Probability Weighting Function. *Econometrica*, *66*(3), 497. https://doi.org/10.2307/2998573

Rouault, M., Drugowitsch, J., & Koechlin, E. (2019). Prefrontal mechanisms combining rewards and beliefs in human decision-making. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-08121-w

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2021). Neural evidence for the successor representation in choice evaluation. *BioRxiv*, 2021.08.29.458114. https://doi.org/10.1101/2021.08.29.458114

Stewart, N. (2011). Information integration in risky choice: Identification and stability. *Frontiers in Psychology*, *2*(NOV), 301. https://doi.org/10.3389/fpsyg.2011.00301

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/J.COGPSYCH.2005.10.003
Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bulletin*, *2*(4), 160–163. https://doi.org/10.1145/122344.122377
Sutton, R. S., & Barto, A. G. (2017). Reinforcement Learning : An Introduction 2nd Edition. https://doi.org/10.1109/TNN.1998.712192
Vidaurre, D., Cichy, R. M., & Woolrich, M. W. (2021). Dissociable Components of Information Encoding in Human Perception. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhab189
Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637. https://doi.org/10.1111/cogs.12101
Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. *NeuroImage*, *26*(1), 99–113. https://doi.org/10.1016/j.neuroimage.2005.01.011
Wimmer, G. E., & Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-10597-z
Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252
Wise, T., Liu, Y., Chowdhury, F., & Dolan, R. J. (2021). Model-based aversive learning in humans is supported by preferential task state reactivation. *Sci. Adv*, *7*, 9616–9644.

Acknowledgements

We thank Matt Nour, Toby Wise, Jess McFayden, Oliver Vikbladh and Rachel Bedder for helpful conversations about analysis. Additionally, we thank Daniel Bates for assistance with data collection and Nathaniel Daw for contributing code used for part of behavioral model fitting. We acknowledge funding from the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning to Y.L. and a Wellcome Trust Investigator Award (098362/Z/12/Z) to R.J.D. This work was carried out whilst R.J.D. was in receipt of a Lundbeck 20 Visiting Professorship (R290-2018-2804) to the Danish Research Centre for Magnetic Resonance. The Max Planck UCL Centre is supported by UCL and the Max Planck Society. The Wellcome Centre for Human Neuroimaging (WCHN) is supported by core funding from the Wellcome Trust (203147/Z/16/Z).

Data Availability

Data underlying all figures will be made available upon publication.

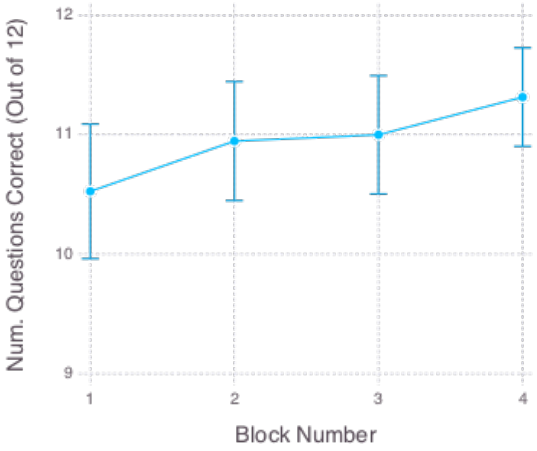
Code Availability

Analysis code underlying all figures will be made available upon publication.

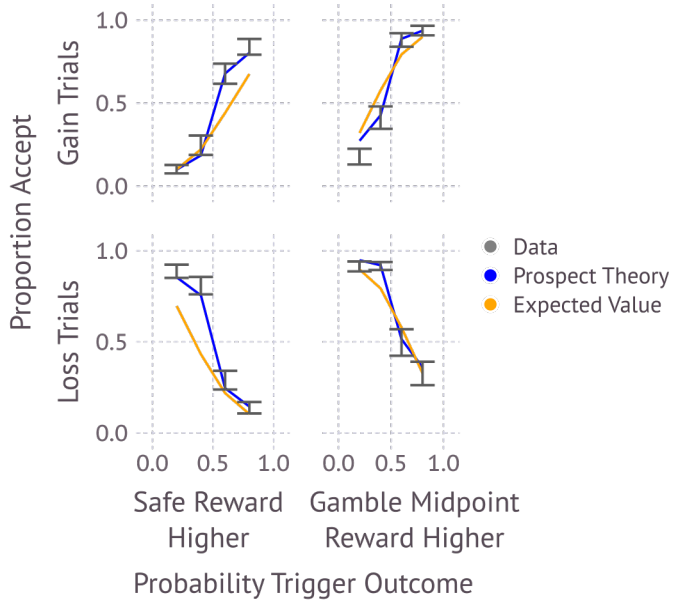
Conflict of Interest

None.

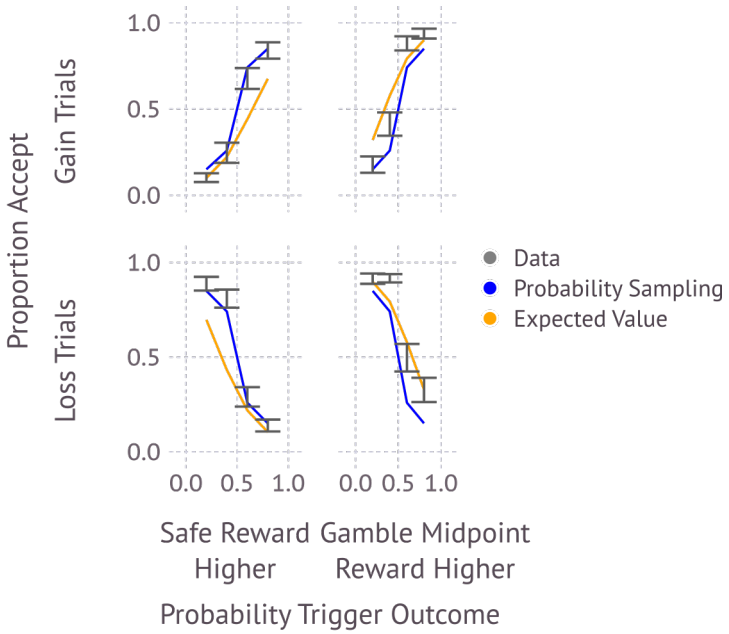
Supplementary Figures



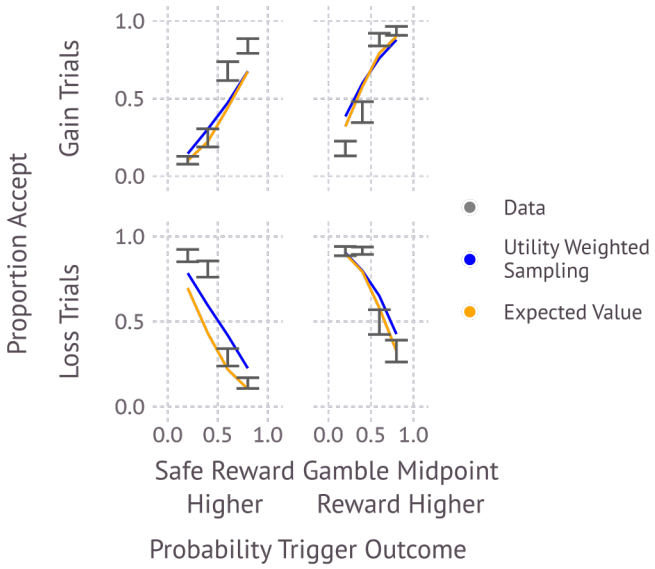
Supplementary Fig. 1. Performance on probability quiz. Prior to the decision making task, yet following the localizer task, participants were trained to learn the probability that each choice stimulus led to each outcome following acceptance (see Methods). Training consisted of 4 blocks. Each block ended with a series of 12 questions, where participants had to answer either which of two outcome stimuli were more likely to follow a choice stimulus (if accepted), or alternatively which of two choice stimuli, if accepted, were more likely to lead to a presented outcome. Line designates mean (+/- s.e.m.) number of questions correct (out of 12) on each block.



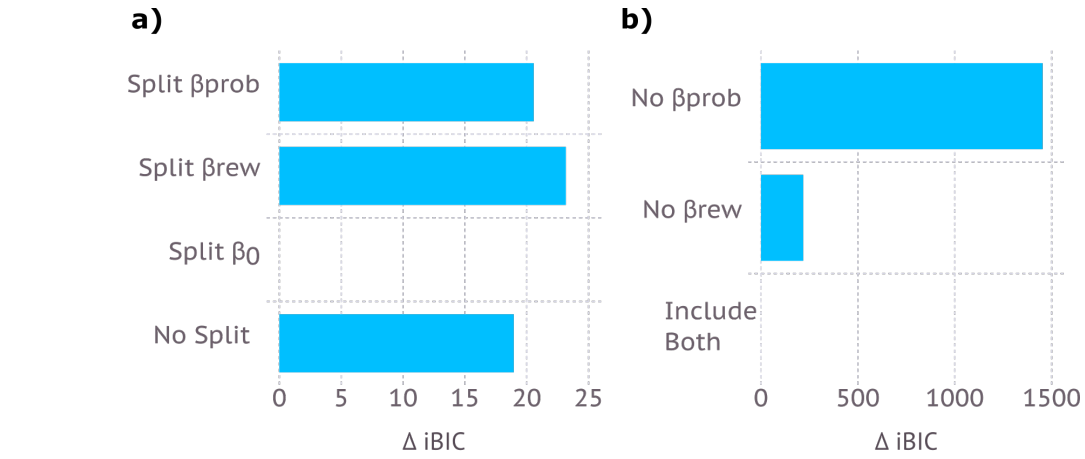
Supplementary Fig. 2. Comparison of Prospect Theory to Data. Comparison of Prospect Theory model predictions and data. Each data point (grey) shows the across-subject mean proportion acceptance for each combination of trigger-outcome value (column), safe outcome value (row) and trigger outcome probability contingent on acceptance (x-axis). Note values reflect rewards prior to common and other noise added. The blue line shows predictions of the Prospect Theory model, at best fit parameters.



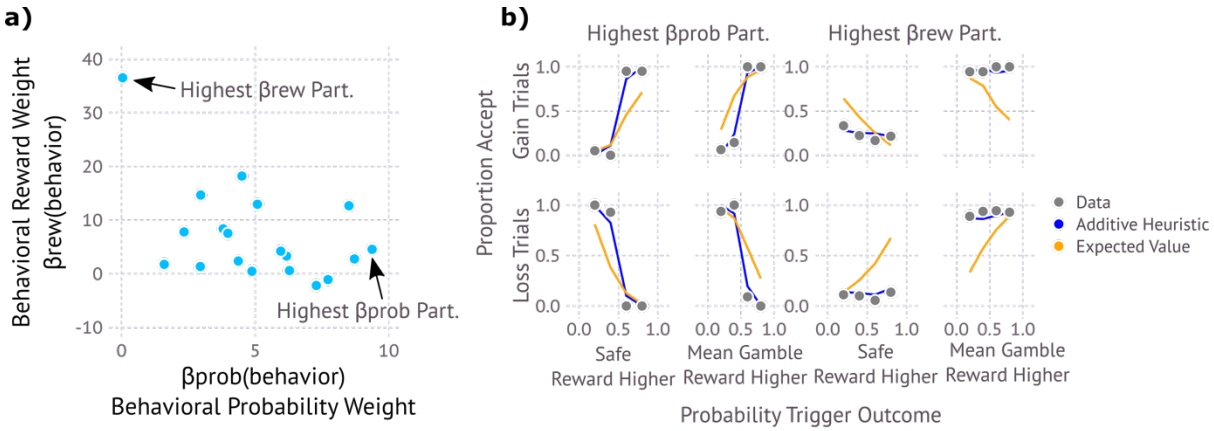
Supplementary Fig. 3. Comparison of Probability Sampling Model to Data. Comparison of Probability Sampling model predictions and data. Each data point (grey) shows the across-subject mean proportion acceptance for each combination of trigger-outcome value (column), safe outcome value (row) and trigger outcome probability contingent on acceptance (x-axis). Note values reflect rewards prior to common and other noise added. The blue line shows predictions of the Probability Sampling model, at best fit parameters.



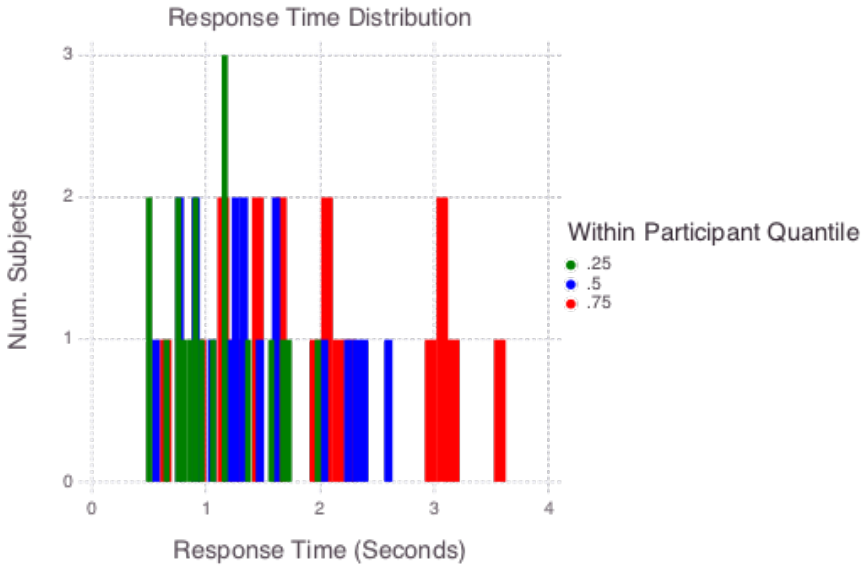
Supplementary Fig. 4. Comparison of Utility Weighted Sampling to Data. Comparison of Utility Weighted Sampling model predictions and data. Each data point (grey) shows the across-subject mean proportion acceptance for each combination of trigger-outcome value (column), safe outcome value (row) and trigger outcome probability contingent on acceptance (x-axis). Note values reflect rewards prior to common and other noise added. The blue line shows predictions of the Utility Weighted Sampling model, at best fit parameters.



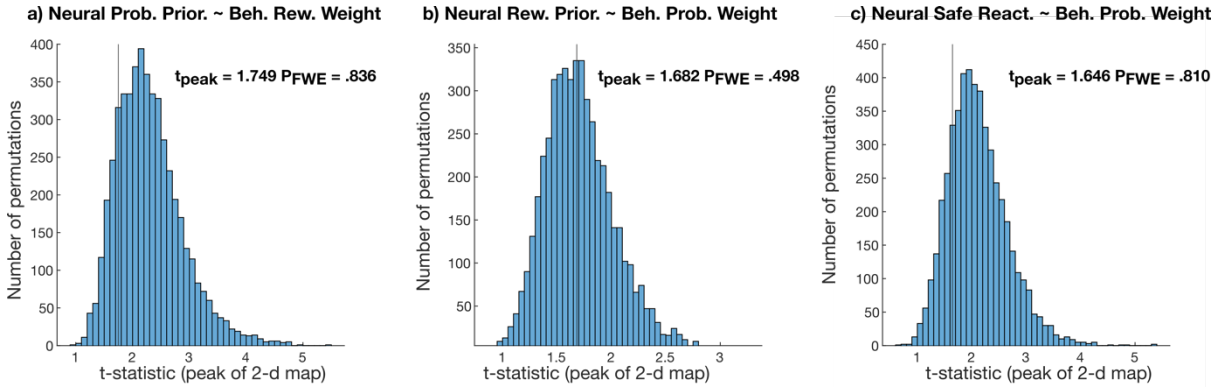
Supplementary Fig. 5. Comparing variations of Additive Heuristic Model. a) Model that splits just β_0 between gain and loss trials fits provide the best account to choice data. "No Split" model is the Additive Heuristic model (as in Fig. 2c), yet does not use separate β_0 for gain and loss trials. "Split β_0" is the Additive Heuristic model as presented in Fig. 2c. "Split β_{rew}" and "Split β_{rew}" respectively include either a separate β_{rew} or β_{prob} parameter for gain and loss trials. **b) Models that did not use either probability information, "No β_{prob}", or did not use reward information ""No β_{rew}" fit the data worse than the model that use both components, "Include Both".** b,c) Models are compare using integrated Bayesian Information Criterion (iBICO. Plots show iBIC relative to best fitting model (Additive Heuristic, Fig. 2c).



Supplementary Fig. 6. Participants varied in use of probability and reward information. a) Each point represents the best fit behavioral probability weight and behavioral reward weight, as use in the Additive Heuristic Model (Fig. 2c), for a single participant. Note that models were fit using a hierarchical approach which jointly maximizes the probability of group and individual participant parameters (Methods). b) Model-fits for individual participants with highest fitted probability information component (left) and highest fitted reward information component (right).



Supplementary Fig. 7. Distribution of participant response time quantiles. Each color designates a different within participant response time quantile. Bar heights show the number of participants at that quantile. The .25 quantile of the fastest responding participants was around 500 ms. This informed our decision to limit the analysis of MEG data to the first 500 ms following choice stimulus onset.



Supplementary Fig. 8. Relationships between MEG outcome reinstatement and alternative behavioral weights. a) We did not identify a positive relationship between neural probability prioritization and behavioral reward weight. b) We also did not identify a positive relationship between neural reward prioritization and behavioral probability weight. c) We also did not identify a positive relationship between neural safe reactivtion and behavioral probability weight.