Heuristics in risky decision-making relate to preferential representation of information

Evan M. Russek[1,2,*], Rani Moran[1,2], Yunzhe Liu[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, Queen Square Institute of Neurology, London, UK
[2] Wellcome Centre for Human Neuroimaging, University College London, Queen Square Institute of Neurology, 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: evrussek@gmail.com
[*] Current affiliation: Departments of Computer Science and Psychology, Princeton University, Princeton, NJ

Abstract

When making choices people differ from each other, as well as from normativity, in how they weigh different types of information. One explanation for this deviance relates to selective prioritization of what information is considered during choice evaluation. To formally test this, we employed a risky decision-making paradigm to examine the relationship between individual differences in neural representation of information and behavior. Specifically, we quantified the extent to which individual participants relied behaviorally on probability versus reward information and related this to how stimuli most informative for making probability and reward comparisons were neurally represented during the risky choice evaluation. Individual differences in a tendency to neurally represent reward- versus probability-informative stimuli explained differences in weighting of either information type in choices. We validated these results in a second behavioral experiment where outcome representation was indexed using a combination of priming and perceptual detection. Our overall results suggest that differences in the information individuals consider during choice shape their risk-taking tendencies.

Introduction

When faced with a choice among actions that can lead to multiple outcomes, decision theory postulates that individuals should compute choice values by taking the expectation over outcomes, each weighted by their probability (Bernoulli, 1954; Edwards, 1954; Savage, 1972). However, psychologists have long shown that instead of deploying this strategy subjects instead exploit several heuristics, including inappropriately weighting either reward or probability information (Allais, 1953; Einhorn & Hogarth, 1986; Ellsberg, 1961; Thaler, 1980). Although some models have offered parameterizations of heuristic reliance on either type of information (Farashahi et al., 2019; Gonzalez et al., 1999; Kahneman & Tversky, 1979; Stewart, 2011), the precise neurocognitive mechanisms underlie individual use of these heuristics remains unknown. In this work, we exploit novel advances in magnetoencephelopgraphy (MEG) to test a specific hypothesis – namely, that underlying heuristic reliance on either source of information reflects a preferential representation of stimuli that are most informative for using such information during evaluation.

In a typical risky choice task, individuals choose between a 'safe' option with a known, fixed outcome, and a gamble option which can 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; Edwards, 1954; Savage, 1972). One explanation for 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; Krueger et al., 2022; Lieder & Griffiths, 2019; Payne et al., 1988). Whereas the normative choice strategy requires independent consideration of each possible task outcome, individuals can reduce the number of outcomes they consider through preferential reliance on a particular type of information during 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, leading to a decision solely based on whether the more likely gamble outcome is attractive. Alternatively, they could prioritize reward information, and solely represent outcomes useful for comparison along this dimension.

We hypothesized that prioritization of distinct types of information during choice evaluation – and more specifically preferential representation of outcome stimuli relevant for comparing choices alongside the type of information prioritized – would explain heuristic weightings of probability and reward information. We leveraged individual differences in heuristic reliance on reward or probability information in choice behavior and examined whether this variability related to inter-participant variability in a disposition to represent outcomes which support a prioritization of a one or the other type of information. If heuristic reliance on probability or reward information in behavior is driven by prioritization of probability or reward information during choice evaluation, then we would expect individuals that weigh probability or reward information more in choice to preferentially represent outcome stimuli useful for comparing choices according to that information dimension. At a higher level, we sought to determine whether the outcomes that an individual tends to 'think of' when deciding underpin the type of information their choices reflect a heuristic reliance upon.

We present affirmative evidence by examining relationships between choice behavior and markers of outcome representation from two modalities: namely magnetoencephalography and behavior. Specifically, we use both magnetoencephalography and behavior to determine which outcomes are preferentially represented during choice and use choice behavior to

measure risk-taking heuristics. In our primary experiment, we utilize recent advances in multivariate methods for magnetoencephalography (MEG) (Kurth-Nelson et al., 2015; Liu, Dolan, et al., 2021; Liu, Mattar, et al., 2021; Wise et al., 2021) to decode which outcome stimuli participants represent while they make a risky choice. Briefly, this involves, first, the identification of MEG signatures of visual stimuli associated with different outcomes; and, second, the examination of these signatures during choice. This data show that individual differences in outcomes representation during decisions are systematically related to the individual differences in observed behavior. A second experiment validates this finding using a behavioral priming manipulation involving interruption of the choice evaluation period with a perceptual detection task (c.f. (Bornstein & Daw, 2013; Garvert et al., 2017)). Consistent with our MEG experiment, this shows that faster detection of a stimulus related to increased behavioral weighting of the information supported by representation of that stimulus. Finally, we find that the neural prioritization was related to real-world self-reported behavioral trait of impulsivity.

In summary, individual differences in heuristic weightings of probability and reward information during choice relate to differential tendencies related to which outcomes are prioritized for representation during option evaluation. The findings establish a link between a representation of different sources of choice relevant information and the types of decision patterns individuals manifest in risky choice.

Results

MEG Decision-Making Task

Participants (n = 19) completed a risky decision-making task while we acquired simultaneous neural data using MEG (Fig. 1). 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 prior to task commencement (Supplementary Fig. 1). On each trial, the points paired with each outcome changed and participants were notified of the reward paired with each outcome at the start (Fig. 1C). Note that reward changes occurred in a structured manner, such that one of the two outcomes, referred to as the trigger outcome (which was counterbalanced between being O1 and O2) had a reward with high absolute value, while the other gamble outcome had points close to 0. We use the term ‘safe outcome’, OS, to refer to a certain outcome whose value lay between O1 and O2. Note that for blocks, involving loss trials, the safe option is paired with negative points.

Critically, in order to facilitate MEG analysis, the time course by which information was presented was structured such as to enforce evaluation of choice options at an identifiable timepoint. 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 at this timepoint. Choice evaluation involving the integration of outcomes O1 and O2 with their probability, where comparison with the safe value could only start when the probability stimulus appeared on the screen following this (Fig. 1A, middle). Note that 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. This task structure enables us to determine what information was represented during the choice process and how this information related to the ensuing choice.

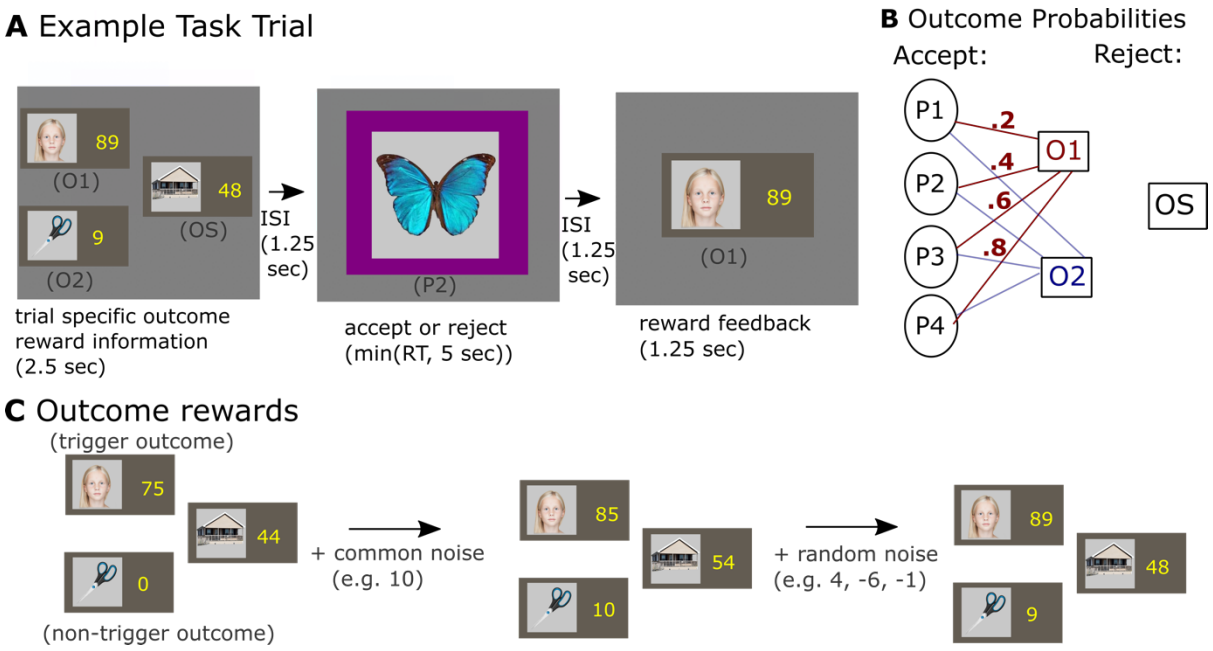


***Fig. 1. Task**. Participants (n = 19) completed a decision task to probe online integration of outcome probabilities and rewards while undergoing MEG. On each trial, participants chose between a safe stimulus (OS) or a gamble which probabilistically led to one of two outcome stimuli. The task controlled 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 (P) 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 on 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.

A parameterization of heuristic reliance on probability and reward information in choice

We hypothesized that heuristic reliance on reward and probability information reflects different approaches for deciding which information to represent during evaluation. Testing this hypothesis required us to parametrize, for each participant, the extent to which choices reflected a heuristic reliance on probability versus reward information. We obtained such a parameterization using a model inspired by prior additive models fit to choices (Farashahi et al., 2019; Stewart, 2011), which we refer to as the Additive Heuristic model (Fig. 2A). Applied to the current task, the Additive Heuristic model decides by computing two distinct components (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 for gain trials, but also for loss trials, where a loss can be viewed as a negative reward. Importantly, because of how rewards were structured in the task (Fig. 1C), following baseline subtraction of the number of points closest to zero (so that all outcome's points are the distance from the lowest absolute point number), 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. 2A for analysis of which parameters should be split between gain and loss trials and Supplementary Fig. 2B for necessity of both reward and probability information components).

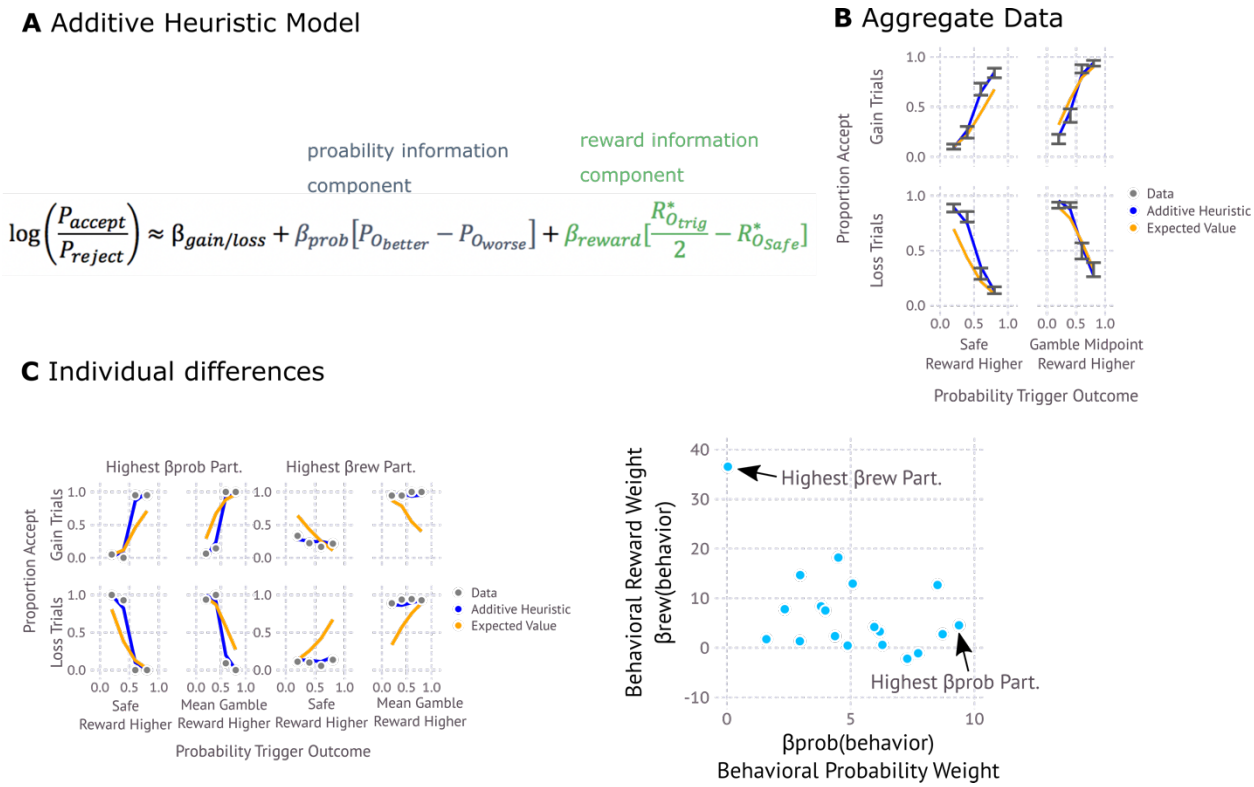


Figure 2: Additive heuristic model. ***A) The Additive Heuristic Model additively combines reinforcement and probability information.*** *The "probability information component", measures the difference in probability between reaching the better (higher reward) versus worse gamble outcome, contingent on accepting the choice stimulus. The "reward information component", measures the difference in reward associated with the midpoint between the gamble and 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 is computed by dividing the trigger reward by two and subtracting the safe reward.*
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 where trigger outcome probability contingent on acceptance (x-axis). Note values reflect outcome reinforcements prior to adding common and other noise. 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) Additive Heuristic model indexes individual differences in heuristic reliance on probability or reward information. *Left) Model-fits for individual participants that either rely exclusively on probability information (left) or reward information (right). Right) The Additive Heuristic parameterization of use of reward and probability information (β_{Prob} and β_{rew}) place these two participants on either end of a continuous which smoothly parameterizes use of either of these two strategies.*

The additive heuristic model captured participants' aggregate choices in the task, and in particular deviations from a model that decided by computing expected values (Fig. 2B). More relevantly, we found that the additive heuristic model captured the extent to which individual participants relied on either probability or reward information in choice (Fig. 2C). Specifically, the Additive Heuristic model provides two parameters for each participant, β^{s}_{prob} and β^{s}_{reward}, which measure choice reliance on probability versus reward information respectively. Henceforth, we refer to these parameters as "Behavioral Probability Weight" and "Behavioral Reward Weight" parameters $\beta^{s}_{prob(behaviour)}$ and $\beta^{s}_{reward(behaviour)}$.

Note that although we use the additive heuristic model to parameterize heuristic reliance of probability and reward information in choice, we do not consider this model itself provides for a strong claim that valuation is additive rather than multiplicative. The choice to use the additive heuristic model is solely based on it providing a more parsimonious fit to behavior (Supplementary Fig. 3) and superior parameter identifiability (Supplementary Tables 1 and 2) compared to alternative models (see Supplementary section Justification for use of Additive Heuristic Model to Parameterize Heuristic Use of Reward and Probability Information).

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. 2B). However, individuals differed substantially in their tendency to rely on one or the other of these components (Fig. 2C). We hypothesized this variability reflected tendencies to consider different classes of information when evaluating choices. The additive heuristic model suggests one way this might occur. For example, one means to compute the probability component of the additive heuristic model is 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.

To test a hypothesis that individual variation in choice behavior was driven by differences in tendencies in relation to which outcomes were considered, we used decoding of MEG data to decode during choice deliberation. Here we conjectured that individual whose behavior reflected a greater reliance on probability information, as indexed by higher Behavioral Probability Weight, would also tend to represent neurally gamble outcomes with higher probability. By contrast, individuals whose behavior reflected greater reliance of reward information (as indexed by higher Behavioral Reward Weight) would tend to represent gamble outcomes based on the absolute value of rewards and the safe outcome for comparison.

To identify neural representations of outcome stimuli, we trained classifiers on data collected prior to the decision-making task (Fig. 3A) (see Methods). Each classifier outputted an "Activation Probability", reflecting the probability that the sensor data reflected reactivation 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, reflect distinct aspects of a stimulus at retrieval. On this basis we trained multiple classifiers separately on data from each 10 ms time bin, τ, following the stimulus presentations. Cross-validation accuracy was quantified as the proportion of trials for which the classifier corresponding to the presented outcome (for held-out data) had the highest activation probability. 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 enabled us to then investigate which aspect of an outcome's representation are reinstated during choice evaluation. Note that for task analysis, we rely on the activation probability measure, as it is it is a more sensitive metric than discrete accuracy, and can identify changes in representation even if an item is not judged to be the most likely.

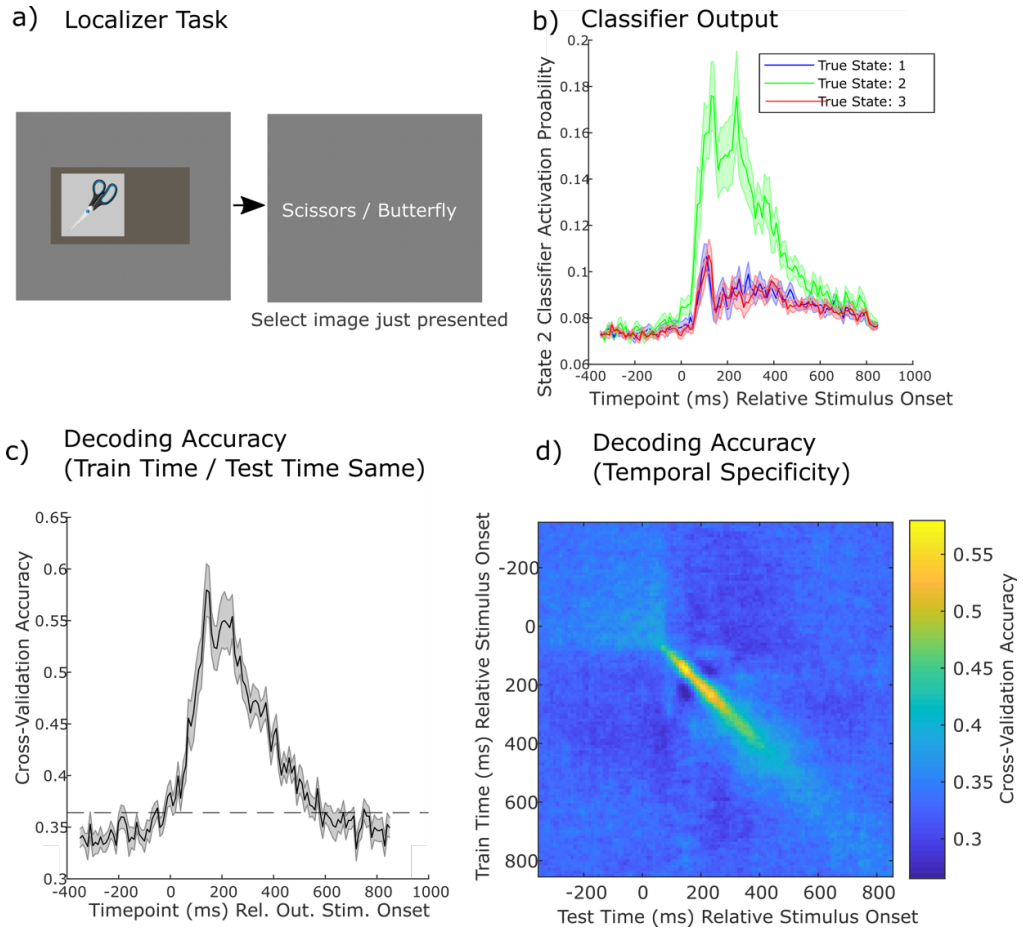


Fig. 3. Decoding from MEG activity. *A)* ***Localizer Task****. The Localizer task was completed prior to the risky decision task and prior to learning choice-outcome probabilities. On each trial participants were shown an outcome or choice stimuli, and, on the next screen, selected a word corresponding to the stimulus they had just observed.* ***B) Activation Probability Measure.*** *For each outcome stimulus, we trained lasso-regularized logistic regression classifiers to discriminate MEG data from when an outcome stimulus was presented, compared to data from presentation of all other images and inter-trial intervals. Each classifier output is an estimated probability that the corresponding stimulus was being presented (Activation Probability). A classifier for each stimulus was trained at successive 10 ms bins of MEG sensor, between -350 and 800 ms, following stimulus presentation. In the example, lines 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, and 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 -350 to 800 ms following presentation of stimuli from held out data. The resulting accuracy image demonstrates temporal selectivity - classifiers identify with good accuracy representations of stimuli specific to the timepoint on which they were trained.* ***B-D)*** *Values reflect group means across 19 participants.*

We next asked which outcome representations were reinstated during choice evaluation, and related this to behavioral markers reflecting 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, and 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 prioritized reactivation 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.

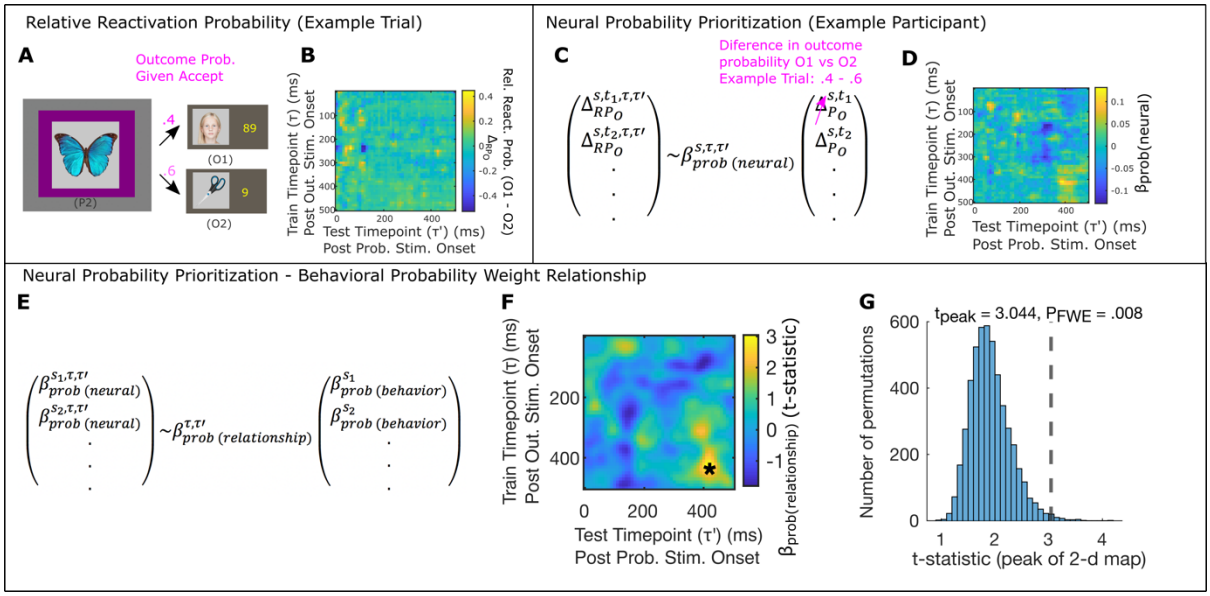


Fig. 4. Behavioral sensitivity to probability information relates to relative reactivation of more probable gamble outcomes following probability stimulus onset*. Neural Probability Prioritization, $\beta_{prob\ (neural)}^{s,\tau,\tau'}$, measures the extent to which relative reactivation probability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

A) Example trial to demonstrate computation of $\beta_{prob\ (neural)}^{s,\tau,\tau'}$*. 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 probability 0.6.*

B) Example relative activation for O1 versus O2. *Following probability stimulus presentation for each trial, we measure reactivation probability for O1 and O2, $\Delta_{RP_O}^{s,t,\tau,\tau'}$, 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, $\beta_{prob\ (neural)}^{s,\tau,\tau'}$, measures tendency to reactivate gamble outcomes according to their probability. *Neural Probability Prioritization, $\beta_{prob\ (neural)}^{s,\tau,\tau'}$, is computed by regressing relative trial-varying reactivation probability of O1 versus O2, $\Delta_{RP_O}^{s,t,\tau,\tau'}$, onto the trial-varying probability of encountering O1 versus O2, $\Delta_{P_O}^{s,t}$ (see Methods).*

D) Neural Probability Prioritization, $\beta_{prob\ (neural)}^{s,\tau,\tau'}$ for example participant. *Image denotes $\beta_{prob\ (neural)}^{s,\tau,\tau'}$ 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 $\beta_{prob\ (neural)}^{s,\tau,\tau'}$ we measured the between-participant relationship between this and behavioral evidence for consideration of probability information, as measured by the behavioral probability weight $\beta_{prob\ (behavior)}^{s}$. This was done by regressing $\beta_{prob\ (neural)}^{s,\tau,\tau'}$ onto $\beta_{prob\ (behavior)}^{s}$, separately for each train and test timepoint, τ and,τ'.*

F-G) 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 $\beta_{prob\ (behavvior)}^{s}$ between participants, s. Dashed line shows true maximum t-statistic.*

To test whether a tendency to reactivate outcomes according to their probability is reflected in a behavioral choice sensitivity to outcome probability information, we computed the between-participant relationship between Neural Probability Prioritization, $\beta_{prob(neural)}^{s,\tau,\tau'}$ and Behavioral Probability Weight, $\beta_{prob(behavior)}^{s}$ (Fig. 4E). The peak of this effect was significantly positive (Figs. 4f-g 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; see Supplementary section “Estimating Behavioral-Neural Correlations”, Supplementary Fig. 9A for estimation of unbiased behavioral-neural correlations), 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 $\beta_{prob(neural)}^{s,\tau,\tau'}$ and $\beta_{rew(behavior)}^{s}$ (Supplementary Fig. 7A).

In a similar manner, we investigated the reward component, which calls for consideration of the trigger outcome (gamble outcome with higher absolute reward) and safe outcome value (Fig. 2A). 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-g τ = 480 ms, τ' = 110 ms; ; P_{FWE} = .024, non-parametric permutation test on image peak; see supplementary Fig. 9B for estimation of unbiased behavior-neural correlation). 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).

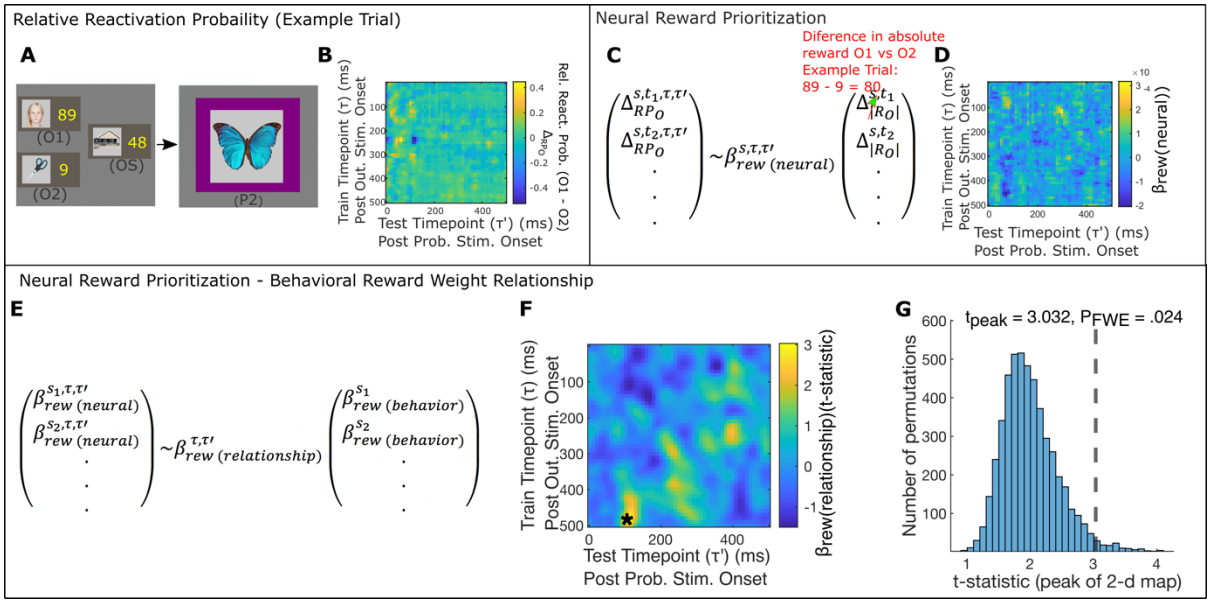


Fig. 5. Behavioral sensitivity to reward information relates to relative reactivation of higher absolute value gamble outcome representation*. Neural Reward Prioritization, $\beta^{s,\tau,\tau'}_{rew\ (neural)}$, measures the extent to which relative reactivation probability of O1 versus O2 changes depending 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, $\Delta^{s,t}_{|R_O|}$. Note that when O2 has greater absolute reward than O1, this quantity is negative.*

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-G) 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.*

We additionally computed participant specific tendencies to reactivate the safe outcome OS, $RP^{s,\tau,\tau'}_{O_S}$, as the mean reactivation probability of the safe outcome classifier across trials (Supplementary Fig. 6A) and regressed this onto Behavioral Reward Weight ($\beta^{s}_{rew(behavior)}$; Supplementary Fig. 6B). Although 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*),* we found that this effect was dependent on a single participant and thus warrants caution in interpretation (supplementary Fig. 9C).

The effects here suggest the more an individual relied on a simple comparison between the rewards from a gamble and safe options, the more they reactivated the high absolute reward outcome. Additionally, we find weak evidence that these individuals also activate the safe outcome, as would be expected for a comparison. As with the above, these associations were specific as we did not observe a positive relationship between $RP^{s,\tau,\tau'}_{O_S}$ and $\beta^{s}_{prob(behavior)}$ (Supplementary Figs. 7C).

Altogether, these results support the idea that individual differences in outcome reactivation prioritization relate to individual differences in choices. Participants who were behaviorally reliant on probability information were also more likely to reactivate gamble outcomes based on their probability. Conversely, participants who were behaviorally reliant on reward information tended to reactivate gamble outcomes based on their absolute reward. Hence, whether probability and reward information influenced behavior related to prioritized neural reactivation for the relevant dimension of information.

Behavioral priming perceptual discrimination task

We next sought to conceptually replicate these findings using a robust behavioral measure. Inspired by work on priming, as well as work which has used response times to identify prospective representations of stimuli, we reasoned that the active representation of stimuli ((as identified in the MEG study) should influence the speed at which those stimuli are perceptually detected. Based on this reasoning we devised a new version of the task where we used priming and perceptual discrimination manipulation to index representation.

The task was equivalent to the decision task used in the MEG study, except for the use of perceptual discrimination rather than MEG decoding to index which outcomes were

represented during choice. The key departure was that on one-third of trials participants performed a perceptual discrimination task rather than a choice task (Fig. 6A). Specifically, following presentation of the probability stimulus, participants were presented with a screen showing the three outcome stimuli. One of the stimuli contained a probe - an arrowhead symbol – and participants were required to report, as quickly as possible, the direction of the arrow.

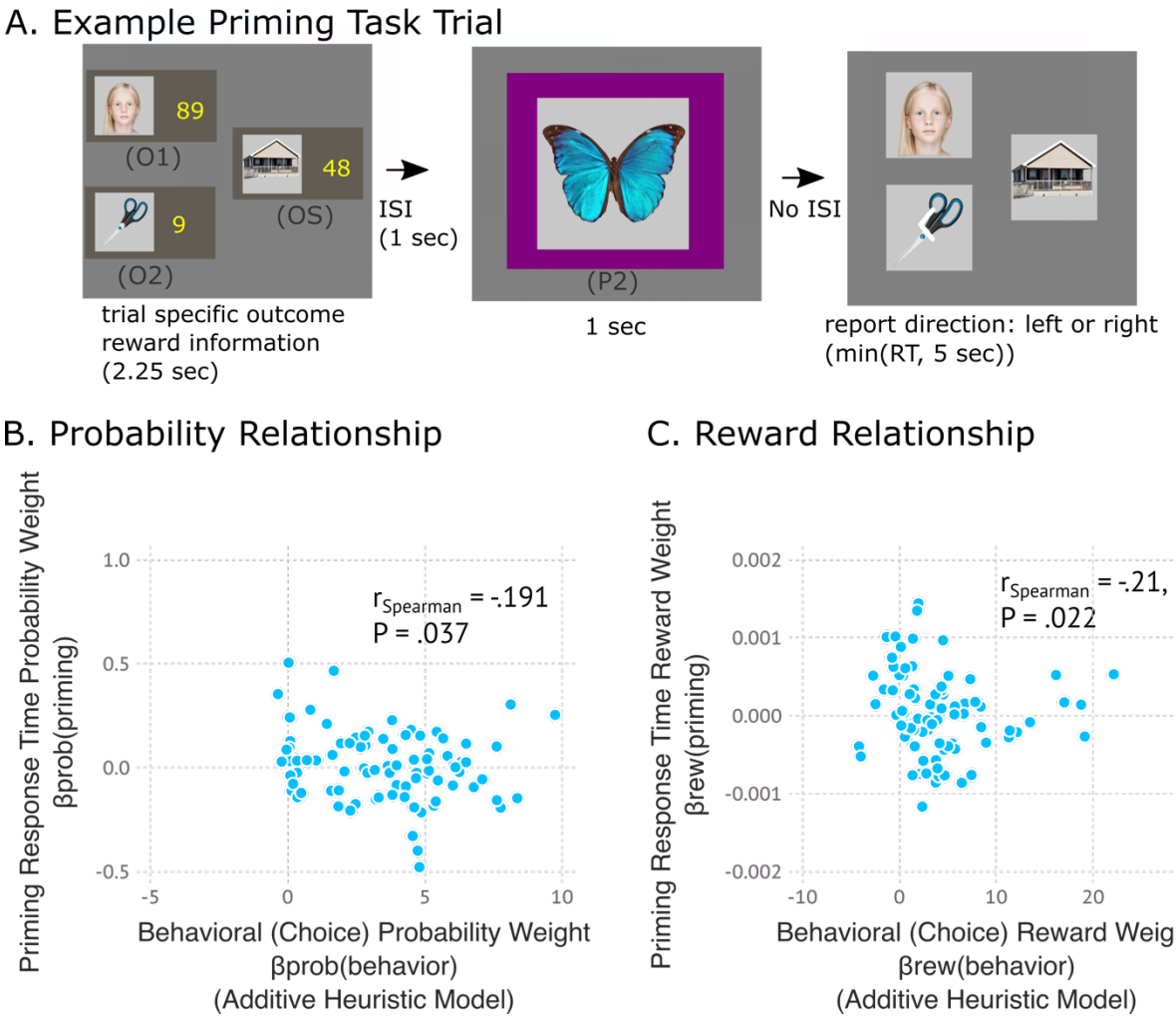


Fig. 6. Behavioral priming task provides conceptual replication of key findings. A)*. In the behavioral priming task, two thirds of trials were equivalent to trials in the MEG decision making task (Fig. 1a). In one third of trials, following presentation of the probability stimulus, the probability stimulus was removed (after 750 ms) and the participant was presented with three outcome stimuli. One of the these (the probed stimulus) had an arrow placed upon it and participants were required to respond as quickly as possible to report the arrow direction.* ***B)*** *Priming Response Time Probability Weight, $\beta_{prob(priming)}$, measures the effect of the probe stimulus's probability (conditioned on accepting the gamble presented earlier in the trial) on the participant's response time in reporting the arrow's direction. Negative values for $\beta_{prob(priming)}$ indicate faster responses when the probed stimulus is more probable, indicative of the more probable outcome being represented during the Probability stimulus presentation. We observed a negative relationship between $\beta_{prob(priming)}$ and $\beta_{prob(behavior)}$, the extent to which probability information guided choice. This is consistent with a hypothesis that the more participants used probability information in choice, the more they tended to represent outcome stimuli based on their probability. This provides a conceptual replication of the MEG findings in Fig. 4.* ***C)*** *Priming Response Time Reward Weight, $\beta_{rew(priming)}$, measures the effect of the*

probe stimulus's absolute reward (relative to the other gamble outcome) on the participant's response time to report the arrow probe. We observed a negative relationship between $\beta_{rew(priming)}$ *and* $\beta_{rew(behavior)}$*, indexing the extent to which reward information guided choice. This indicates that participants who used reward information in choice tended to prioritize which stimuli to represent based on their absolute reward, providing a conceptual replication of MEG findings from Fig. 5.*

As in MEG decoding, the discrimination of the arrowhead direction amongst the stimuli offers an opportunity to ascertain the extent to which the probed stimulus was being actively represented during presentation of the Probability stimulus. If participants were actively representing a stimulus, then processing of that stimulus would be prioritized and this would result in faster discrimination of the probe direction. In effect, we were interested in whether a tendency to use probability versus reward information in choice was accompanied by a tendency to prioritize outcomes for reactivation based on either their probability or absolute reward.

Analogous to our MEG analysis (Figs. 4 and 5) we measured the extent to which the probed stimulus's probability and absolute reward affected response times in reporting the probe ($\beta_{prob(priming)}$ and $\beta_{rew(priming)}$ respectively). More negative values of $\beta_{prob(priming)}$ and $\beta_{rew(priming)}$ indicate that participants were more inclined to represent outcome stimuli when they had higher probability, or higher absolute reward respectively. We then tested whether these tendencies were related to use of probability versus reward information in forming decision variables.

We found a greater tendency to use probability information in choice (measured as utilizing a greater Behavioral Probability Weight, $\beta_{prob(behavior)}$) related to a greater tendency to represent outcomes based on their probability (measured as lower $\beta_{prob(priming)}$ reflecting faster responses for more probable outcome stimuli $r_{spearman}$ = -.19, $t(86) = -1.8$, $P = .037$; Fig. 6B). This provides a conceptual replication of the MEG results presented in Fig. 4.

Analogously for reward, a greater tendency to use reward information in choice (measured as utilizing a greater Behavioral Reward Weight, $\beta_{rew(behavior)}$) related to a greater tendency to represent outcomes based on their reward (measured as lower $\beta_{rew(priming)}$ reflecting faster responses for outcome stimuli with higher absolute reward; Fig 6C; $r_{spearman}$ = -.21, $t(86) = -2.0$, $P = .022$). This provides a conceptual replication of the MEG results presented in Fig. 5.

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 et al., 2016; Loewenstein et al., 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 reactivation 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 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-C, $\tau = 410$ ms, $\tau' = 370$ ms, P_{FWE} = .006, non-parametric permutation test on image minimum; see supplementary Fig. 9D for estimation of unbiased behavior-neural correlation).

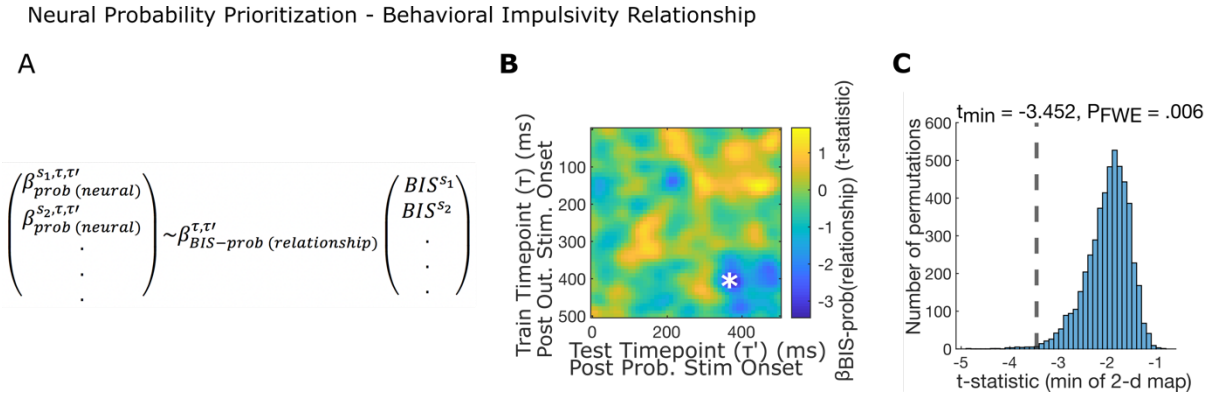


Fig. 7. Relative reactivation of outcomes with high probability is less in individuals higher in impulsivity. a) Measuring Relationship Between Behavioral Impulsivity and Neural Probability Prioritization. *In order to measure the between-participant relationship between behavioral impulsivity and neural 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-c) 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.*

In relation to this result we caution that because probability and reward information are always presented in the same order, we cannot entirely rule out that a reduced representation of high probability outcomes in individuals with higher impulsivity might in fact reflect it being presented as the second piece of information, rather than the first. Additionally, we did not identify a similar relationship in the perceptual discrimination task. Specifically, there was no relationship between (slower) response times for perceptual discrimination for higher probability probe items and participant self-reported BIS score ($r_{spearman}$ = -.084, $t(86) = -.78,\ P = .79$). This null result may reflect lower sensitivity of the behavioral measure compared to MEG. Equally, it suggests the MEG result should be interpreted with caution.

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 question behaviorally and with neural data. 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 prioritized during evaluation.

Our behavioral analysis revealed that participants differed in the extent to which they relied on either reward versus probability comparisons when deciding. By decoding outcome representations using MEG, we show these distinct decision strategies reflected 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 primarily 'thought' about probability information. Conversely, participants who decided more based on the difference in reward between outcomes preferentially reinstated the high absolute value gamble outcomes, suggesting they mainly considered the relative value of gamble options with the safe option.

Our results address a gap in the literature as to what accounts for individual differences in the treatment of reward and probability during 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 preferentially integrate either reward or probability information account for some aspects of commonly observed variance in risky choice (Stewart, 2011), though whether such variation is explained by differences in the types of information considered during choice evaluation has not been shown. Here, by identifying a link between outcomes that are represented during choice evaluation and behavioral signatures that reflect consideration of either reward or probability information, we provide evidence that this variation is related to the types of information prioritized during evaluation.

One caveat to 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. 5), limiting the available time window over which activations could be averaged. However, most participant's choice evaluations lasted longer than this time period, suggesting that we only examined reactivation data corresponding to a fraction of the possible evaluation time used by participants. One explanation for an apparent 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 to examination of the fastest reactivation measures that cohered across participants, future studies might avail of other methods. For example, identification of transitions in reactivation events between stimuli (Liu, Mattar, et al., 2021) that enables aggregation of reactivation events across trials that may have different response times, thus availing of all evaluation data.

Although our results support a hypothesis that heuristic reliance on probability versus reward information are driven by which outcomes are represented during choice, a major caveat is

that our evidence is correlational and does not support a causal conclusion. Future work could assess the latter by causally manipulating which outcomes are represented during choice, perhaps by priming participants to attend to one or other outcome by including additional outcome features.

An additional aspect of our design is that probability versus reward information was always presented in the same order. This does not impact interpretation of our results, because we did not seek to determine whether probability versus reward information is represented to a greater degree in general. Instead, our goal was to ascertain whether individual differences in representation of such information relates to individual differences in use of either source of information at choice. One potential exception to this is our finding that self-reported impulsivity relates to a lesser representation of outcomes based on their probability. We acknowledge that a reduced representation of high probability outcomes in individuals with greater BIS scores could be explained by it being presented as the second piece of information, rather than the first, if individuals with greater BIS scores preferentially represented earlier compared to later information.

A key aspect of our design was its inclusion of only two gamble outcomes. This was motivated by two considerations. Firstly, we wanted to render our task directly comparable to prior work which has characterized choice biases using two outcomes (Farashahi et al., 2019; Gonzalez et al., 1999; Kahneman & Tversky, 1979; Stewart, 2011). Secondly, we wanted to enable the simplest possible decoding analysis of outcome representation reactivations, such that two gamble outcomes can be compared. Although including two outcomes was beneficial for the decoding analysis, future work might utilize tasks with additional outcomes to enable a more fine-grained examination of how outcomes are prioritized for representation, and how this relates to choice heuristics.

Several previous studies have investigated outcome reactivation in the context of model-based reinforcement learning algorithms (Bornstein & Daw, 2013; Castegnetti et al., 2020; Doll et al., 2015; Russek et al., 2021; Wimmer & Büchel, 2019). Typical model-based algorithms postulate that choices are evaluated by simulating potential consequential outcomes and by adding rewards form these outcomes to a running average (Sutton, 1991; Sutton & Barto, 2017). Evidence that outcome reactivation 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 reactivation also accounts for within subject variation of 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, 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. The results 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 a consideration of 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 also been reported (Castegnetti et al., 2020).

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 occur and what temporal component of a representation, 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 occurr bears similarities to that previous study. Notably, our identification of reactivation related to integration of reward and probability information, occurred at two distinct timepoints (110 ms and 420 ms following choice stimulus onset), approximately resembling time-points (Kurth-Nelson et al., 2015) when a non-direct rewarded stimulus was re-activated following a paired stimulus onset (400 ms) or a reward (70 ms). Relatedly, we found that activated representations were those corresponding to classifiers trained around 400 ms following stimulus onset. Previous work (Kurth-Nelson et al., 2015) has identified such classification time-points corresponded to representations that load on temporal cortex topographies, suggesting these areas may support decision-relevant outcome representations.

Finally, we identified that participants with higher behavioral impulsivity demonstrated relatively reduced prioritized reactivation of higher probability outcome representations. This reactivation result matches recent theoretical proposals that impulsive choice may result from a noisy simulation of future events (Gabaix & Laibson, 2017). Given the separate, positive relationship of this pattern of reactivation with integration of probability information, this points toward a potential mechanism to explain real-life aberrant risky choice, potentially a neglect of probability information (Rouault et al., 2019). More generally, our finding here opens line of research that disorders of choice may relate to what information should be prioritized. However, we note again that our failure to replicate the latter result in a a perceptual discrimination study suggests that it needs to be further investigated.

In summary, we demonstrate a relationship between the nature of the information individuals tend to consider during evaluation, and how they actually decide., This implies that one could learn to make better choices by learning to change what information is prioritized for consideration and points toward a research direction for treatment of mental health disorders characterized by aberrant choice.

Methods

MEG Study

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 et al., 2018; Park et al., 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, participants 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. Note that participants completed day 1 from their own personal computers and that behavior from practice trials day 1 was not analyzed, and not reported here.

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. The full MEG session lasted about 90 minutes and consisted 13 runs of scanning sessions. This included 3 runs of the localizer task (each lasting about 5 minutes), 2 runs of probability learning task (each less than 5 minutes, not analyzed), and 8 runs of the decision-making task (each lasting about 7 minutes).

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. 2, .4, .6, and .8, for P1, P2, P3 and P4 respectively (Fig. 1b).

Note that our decision to include two potential gamble outcomes in the task was based on two reasons. The first was make our task directly comparable to prior work which has characterized choice biases in tasks using two outcomes (Farashahi et al., 2019; Gonzalez

et al., 1999; Kahneman & Tversky, 1979; Stewart, 2011), The second was to enable the simplest possible decoding analysis of representation reactivations, such that activations of two gamble outcomes can simply be compared. Although this decision to include two outcomes was beneficial toward making decoding analysis simpler, future work might utilize tasks with additional outcomes so as to study in a more fine-grained manner how outcomes are prioritized for representation and how this relates to choice heuristics.

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.

Additionally in relation to the non-manipulated ordering of probability versus reward information, it was not our goal to determine whether participants were more likely to use either probability or reward information in choice. Rather our goal was to test for a relationship between heuristic weightings of either type of information and a disposition for outcomes to be reinstated at choice.

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-making 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 et al., 2014). In order to gain an algorithmic description of subjects decision making we fit a number of computational models to their choices. The following models are compared in Fig. 3.

For each model, we describe how it determines the probability of accepting an offer based on trial information along with model-specific free parameters.

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

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

Here, P_{O1} and P_{O2} are the respective probabilities of O1 and O2 being received conditioned on accepting the gamble. R_{O1}, R_{O2} and R_{OS} are the number of points paired with O1, O2 and OS for that trial. $logit^{-1}$ is the standard sigmoid logistic sigmoid function. β is a free parameter, th inverse temperature, and controls decision noise.

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 and the safe outcome:

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

$\beta_0 = \beta_{gain,}$ on gain trials and $\beta_0 = \beta_{loss}$ on loss trials and controls baseline tendencies to accept or reject gambles independently of trial information. $\beta_{gain,}$ β_{loss} , β_{prob}, and β_{rew} are free parameters. $P_{O_{better}}$ and $P_{O_{worse}}$ are the respective probabilities of reaching the better and worse gamble outcomes (e.g. $P_{O_{better}} = P_{O1}$ when O1 has more points). $R^*_{O_{trig}}$ is the reward of the trigger outcome (the gamble outcome – O1 or O2 – with higher absolute value), 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.

The following models are compared additionally in Supplementary Fig. X:

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,

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

We used standard utility functions, $v(x) = x^{\alpha_{gain}}$ when $x \geq 0, v(x) = -(x^{\alpha_{loss}})$ when $x < 0$, and. We use the log odds linear probability weighting function, $w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}$. β, α_{gain}, α_{loss}, δ, and γ are free parameters. The effects of altering these parameters are displayed in Supplementary Fig. X. Note that to simplify the general test of relationships to outcome representation, we fit a modified version of this model which only uses a single

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 participants first select a number of samples to take, S, which we assume is drawn from an ordered probit distribution, $OrderedProbit(S \mid n, c)$. n sets the center of the distribution and is a free parameter. 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 $q(O_i)$, which is defined below. Given S samples, the subject computes an estimate of the value difference between the gamble option and safe option:

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

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}. R_{o_i} is the number of points paired with the outcome that was drawn on sample i.

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}:

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

where we took the maximum number of samples, 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. No online filters were applied during collection. The task was divided into multiple MEG sessions, with each session lasting less then 10 minutes. Participants were asked to remain still during the scanning session. Participants were able to take a rest between sessions, however they were required to remain in place in the scanner and encouraged not to move. At the start of each scanning session participant's head positions were registered.

MEG analysis

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

Pre-processing

Preprocessing was performed 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 sensors (using standard artefact rejection in OSL with default parameters - mean 8 +/- 6.01 sensors per participant), independent component analysis (ICA) was applied to denoise the data. We applied fastica, part of the AFRICA ICA procedure within the OSL software package. ICA was run with default parameters, which sets the maximum number of components that can be removed due to kurtosis to 10. In addition, several components were removed due to correlation with recorded EOG channel eye movement data. The mean and standard deviation number of components removed per run is reported in the table below:

	Mean (per run)	Stand. Dev (per run)
Total ICs removed	12.428	2.5429
Removed for Kurtosis	9.214	1.3966
Removed for Correlation with EOG	2.7045	1.3966

Table 1: mean number of ICA components removed, for either kurotosis or correlation with EOG.

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. Our decision to train classifiers from 0 ms to 500 ms post image onset was motivated by three factors. First, our analysis of classifier cross-validation accuracy revealed that we only had significant decoding (testing on the same time-point as was trained on) for train time-points from 0 ms to 560 ms post image onset (See new Fig. 3 plot below). Second, prior evidence has shown that relevant reactivation events occur for classifiers trained on a time-point less than 500 ms post image onset (Kurth-Nelson et al., 2015). Third, using an a priori hypothesis for which representations would be reactive allowed us to increase power for our key tests of relationships between behavioral and neural reactivation events.

In order to decode outcome representations, while minimizing correlations between decoded reactivations, we followed an approach recommended in (Liu, Dolan, et al., 2021) of training models to discriminate one state against a mixture of other states and null data. Thus, 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 10-ms binned timepoint, τ, following outcome stimulus onset. This generated three classifiers, one for each outcome, for each of 50 timepoints, corresponding to each 10-ms bin between 10 ms and 500 ms following outcome stimulus presentation in the localizer task. Given the task response times in addition to prior hypothesis about when relevant reactivation events occur (Kurth-Nelson et al., 2015); 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 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 $RP_{O_x}^{p,t,\tau,\tau'}$ to represent the reactivation probability output by the classifier, trained to activate for stimulus O_X (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 reactivation 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 reliance on 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, $\Delta_{RP_O}^{s,t,\tau,\tau'} = RP_{O_1}^{s,t,\tau,\tau'} - RP_{O_2}^{s,t,\tau,\tau'}$.

For each participant, s, train timepoint, and test timepoint, 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:

$$\Delta_{RP_O}^{s,t,\tau,\tau'} \sim \beta_0 + \beta_{prob(neural)}^{s,\tau,\tau'}\left[P_{O_1}^{s,t} - P_{O_2}^{s,t}\right] + \beta_{rew(neural)}^{s,\tau,\tau'}[|R_{O_1}^{s,t}| - |R_{O_2}^{s,t}|]$$

This provides an estimate of $\beta_{prob(neural)}^{s,\tau,\tau'}$, and $\beta_{rew(neural)}^{s,\tau,\tau'}$, for each participant, s, train timepoint, τ, and test timepoint, τ'. $\beta_{prob(neural)}^{s,\tau,\tau'}$ measures the extent to which, a tendency to reactivate O1 over O2 is driven by the probability of O1 relative to O2 (and vice-versa). Conversely, $\beta_{rew(neural)}^{s,\tau,\tau'}$ measures the extent to which a tendency to reactivate O1 over O2 is driven by the relative absolute reward of O1 compared to O2.

We next sought to determine whether these differences in reactivation tendencies related to behavioral reliance on reward versus probability information in choice. To examine this, in a second level, we related $\beta_{prob(neural)}^{s,\tau,\tau'}$ and $\beta_{rew(neural)}^{s,\tau,\tau'}$ to fitted parameters from 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 reliance on probability information, indexed by $\beta_{prob(behavior)}^{s}$ would be related to preferential reactivation of more probable gamble outcomes, as indexed by $\beta_{prob(neural)}^{s,\tau,\tau'}$, and that behavioral reliance on reward information, indexed by $\beta_{rew(behavior)}^{s}$ would be related to preferential reactivation of outcomes with higher absolute reward, as indexed by $\beta_{rew(neural)}^{s,\tau,\tau'}$.

We thus performed two between participant regressions: one relating $\beta_{prob(behavior)}^{s}$ to $\beta_{prob(neural)}^{s,\tau,\tau'}$ (Fig. 4) and one relating $\beta_{rew(behavior)}^{s}$ to $\beta_{rew(neural)}^{s,\tau,\tau'}$ (Fig. 5). In order to mitigate the impact of potential outliers, following previous work (Eldar et al., 2018), all between-subject behavioral-neural 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 et al., 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 max-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.

Reactivation 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 reactivation of the safe outcome (Supplementary Fig. x). As a measure of safe outcome reactivation, we computed, for each participant, s, train timepoint, τ, and test timepoint, τ', the mean reactivation probability across trials, $RP^{s,\tau,\tau'}_{O_S}$. 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 reactivation to tendency to reinstate outcomes based on their probability (Fig. 7), we repeated the previous between participant regression involving $\beta^{s,\tau,\tau'}_{prob(neural)}$, 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).

Priming perceptual discrimination task

Participants

We recruited 100 participants (mean (std) age: 27.6 (8.1), 35 female) on Prolific to perform the task online in their browser. Data from 3 participants was lost due to errors in recording. Using an equivalent exclusion criterion as used in the MEG study, an additional 5 participants were excluded due to selection of the same action on more than 80% of trials. Finally, an additional 4 participants were removed due to failure to make responses to perceptual detection trials, leaving 88 participants for analysis. For completing the study, which took approximately 65 minutes, participants were paid 9.34 GBP, with a performance dependent bonus between 0 and 3 GBP. This online study was approved by UCL ethics (ID: 16639/001).

Task

After completing instructions and passing a quiz on their contents, participants completed the BIS questionnaire followed by a probability learning task which was identical to that used in the MEG task, however only had three rather than four blocks. They then completed the risky decision making task, consisting of 288 trials. Two thirds of trials were identical to the decision trials in the MEG task, however were run slightly faster: with inter-stimulus intervals of 1 second and inter-trial intervals also of 1 second. Additionally, participants were only allowed to make a choice following observing the probability stimulus for 1 second.

On one third of trials, instead of being allowed to make a choice, the probability stimulus disappeared, and participants were shown the three outcome stimuli, one of which contained an arrow stimulus placed over it (Fig. 6A). Participants were then required to press an arrow key indicating the direction of the arrow as quickly as they could.

Inferring tendencies for what outcomes are represented from response times

We sought to estimate the extent to which individual participants represented outcome stimuli based on either their probability or absolute rewards. If participants tended to represent outcome stimuli based on their probability, they would make faster responses when higher probability outcome stimuli were the probed stimulus compared to when lower probability stimuli were the probe stimulus. Conversely, if participants tended to represent outcome stimuli based on their absolute rewards, they would make faster responses when higher absolute reward outcome stimuli were the probed stimulus compared to when lower absolute-reward stimuli were the probe stimulus.

We thus sought to estimate the effect of relative outcome probability and relative outcome absolute reward on log response times to the perceptual detection probe:

$$log\left(rt_{s,t}\right) \sim \beta_0 + \beta^s_{prob(priming)}\left[P^{s,t}_{O_{probed}} - P^{s,t}_{O_{non-probed}}\right] + \beta^s_{rew(priming)}[|R^{s,t}_{O_{probed}}| - |R^{s,t}_{O_{non-probed}}|]$$

Here, $rt_{s,t}$ is the response time of participant s on trial t. $P^{s,t}_{O_{probed}}$ and $P^{s,t}_{O_{non-probed}}$ are the respective probabilities of the probed and non-probed gamble stimuli on trial for participant s, trial *t.* Note that this regression was only applied to trials where one of the gamble stimuli was the probe. Negative values of $\beta^s_{prob(prime)}$ reflect faster responses for more probable probed stimuli, reflecting a tendency to represent outcomes based on their probability. $\left|R^{s,t}_{O_{probed}}\right|$ and $|R^{s,t}_{O_{non-probed}}|$ are the respective absolute rewards paired with the probed and non-probed gamble outcomes for participant *s* on trial *t.* Negative values of $\beta^s_{rew(prime)}$ reflect faster responses for probed stimuli with higher absolute reward, reflecting a tendency to represent outcomes based on their absolute reward.

To determine if tendencies to represent outcomes based on probability or absolute reward were related to heuristic reliance of reward and probability information in choice, we fit the additive heuristic model to participants behavior and measured spearman correlations to test for a relationship between $\beta^s_{prob(prime)}$ and $\beta^s_{prob(behavior)}$, and between $\beta^s_{rew(prime)}$ and $\beta^s_{rew(behavior)}$.

References

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, *21*(4), 503. https://doi.org/10.2307/1907921

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., Naish, K. R., Reed, D. D., & 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

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., Heinze, H.-J., & 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

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417. https://doi.org/10.1037/H0053870

Einhorn, H. J., & Hogarth, R. M. (1986). Decision Making Under Ambiguity. *The Journal of Business*, *59*(4), S225–S250. http://www.jstor.org/stable/2352758

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*(4), 643–669. https://doi.org/10.2307/1884324

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*. http://www.nber.org/papers/w23254

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *ELife*, *6*, 1–20. https://doi.org/10.7554/eLife.17086

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., Luce, D., Prelec, D., Shafir, E., & Wakker, P. (1999). On the Shape of the Probability Weighting Function. *Cognitive Psychology*, *38*, 129–166. http://www.idealibrary.comon

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

Krueger, P., Callaway, F., Gul, S., Griffiths, T., & Lieder, F. (2022). Identifying Resource-Rational Heuristics for Risky Choice. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/MG7DN

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., Dolan, R. J., Higgins, C., Penagos, H., Woolrich, M., Ólafsdóttir, H. F., Barry, C., Kurth-Nelson, Z., & Behrens, T. (2021). Temporally delayed linear modelling (Tdlm) measures replay in both animals and humans. *ELife*, *10*, 1–35. https://doi.org/10.7554/eLife.66917
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
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. *Proc. of the 40th Annual Conference of Cognitive Science Society*, 2394–2399. http://arxiv.org/abs/1801.09848
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive Strategy Selection in Decision Making. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 14, Issue 3, pp. 534–552). https://doi.org/10.1037/0278-7393.14.3.534
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
Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.
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
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
Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7
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
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.

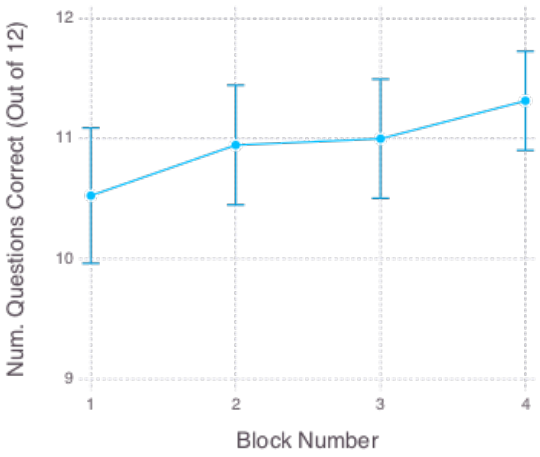
Code Availability

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

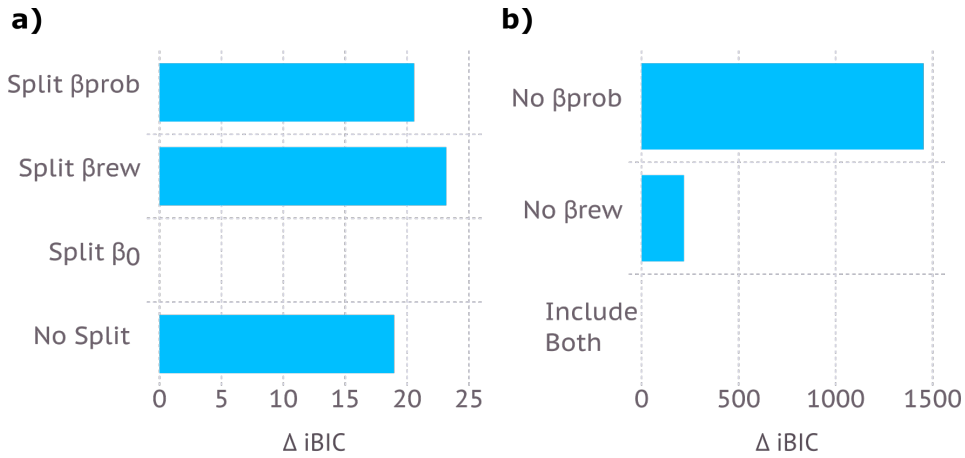
Conflict of Interest

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. 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 (iBIC. Plots show iBIC relative to best fitting model (Additive Heuristic, Fig. 2c).*

Justification for use of Additive Heuristic Model to Parameterize Heuristic Use of Reward and Probability Information

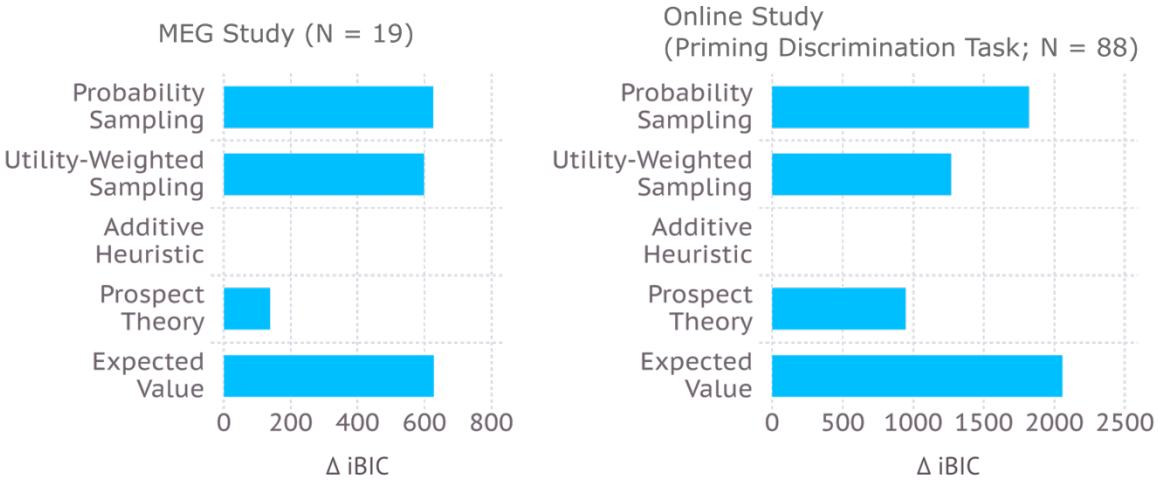
The central goal of our neural analysis was to assess whether heuristic, idiosyncratic, reliance on probability and reward information in choice reflected differing strategies for which outcome' representations were reactivated during choice. This analysis required a participant-level parameterization of idiosyncratic, heuristic, use of reward and probability information in choice, as demonstrated in their behavior. Importantly, such treatment of reward and probability information is not a property of a single model, but rather a source of choice variance that can be captured by a variety of models. For example, whereas prospect theory allows for idiosyncratic treatment of probability information through individual variation in a probability weighting function, the additive heuristic model allows this through individual variation in a behavioral probability weight.

We conducted two analyses to determine which model provided the most useful parameterization for the purpose of guiding our neural analysis. First, we determined which model provided the best explanation of participants choices, using standard approaches to model comparison. Second, a model parameterization is only useful for indexing a choice strategy if its parameters can be identified in the task. Thus, in a second analysis, we compared the identifiability of the key parameters for each model.

Model comparison. We compared the ability of several models to explain participants choices. For the purposes of parameterizing idiosyncratic use of reward and probability information, the key models of interest were the additive heuristic model and prospect theory (Kahneman & Tversky, 1979) models, which both provide parameters for heuristic use of either type of information. Additionally, we also considered two recently proposed models based on sampling outcomes according to some probability distribution. Although these

models do not explicitly have parameterization of heuristic use of reward and probability information, they do make claims about how outcome reactivations might relate to choices, and on this basis are potentially of interest for guiding MEG analysis. Finally, to compare any model to optimal weighting of reward and probability and reward information, we also compared each model to models which decided based on expected value.

Supplementary Figure 3 shows the goodness of fit of each model to participant's choices, as measured by computing the integrated Bayesian Information Criterion (iBIC) over the entire group of subjects for each model. For this analysis, all models were fit 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). This revealed the additive heuristic model provided the best fit to participants behavior ($\Delta iBIC$ Additive Heuristic Model vs Prospect Theory = 142; Supplementary Fig. 3, left). In addition to the choice data from the MEG decision study, we additionally fit models to data from the additional priming study that was performed online. This again revealed the Additive Heuristic model provided a better fit to the data than a Prospect theory model ($\Delta iBIC$ Additive Heuristic Model vs Prospect Theory = 946 Supplementary Fig. 3, right)



Supplementary Figure 3: *Comparison of model iBIC scores. Each bar gives iBIC (integrated Bayesian Information Criterion) relative to the best fitting model. Left: Results of model-comparison to behavioral data from MEG study. Right: Results of model-comparison from online priming/perceptual discrimination task data (Fig. 6). For both tasks, the best-fitting model was the additive heuristic model followed by a prospect theory model.*

In addition to computing iBIC, which provide an overall measure of model fit to all participants, we also sought to obtain unbiased individual-level participant fits, such that we could perform a hypothesis test for whether, at the group level, the additive heuristic model provided a better fit than prospect theory. Thus, for each model, for each participant, we also computed unbiased per-participant log likelihoods, fitting each model in a flat manner, without a prior. Note that because the heuristic model and prospect theory share the same number of parameters, these per-subject log-likelihoods could be directly compared. Comparison of these unbiased log-likelihoods revealed a significant advantage for Additive Heuristic model compared to Prospect Theory both for the MEG as well as the online study: (MEG study: mean +/- sem log likelihood difference = 8.40 +/- 3.97, $t(18) = 3.98$, $p = .048$, two-sided paired sample t-test. Online Study: mean +/- sem log likelihood difference = 7.77 +/- 1.164, $t(87) = 3.98$, $p < 1e\text{-}9$).

Parameter Identifiability. In addition to comparing the ability of each model to explain participant's choices, we sought to examine the identifiability of key parameters governing heuristic use of probability and reward information for the two best performing models: the Additive Heuristic Model and Prospect Theory. To do so we first fit the two models to participants behavior using an EM approach, so as to identify group-level distributions of each parameter (parameterized as normal distributions with means and variances). Then, for each model, we generated 100 new datasets (with 19 participants each), repeatedly sampling parameters for each participant from the group level distributions. For each new dataset, we re-fit both models, again using the EM approach, in order to obtain parameter estimates for each participant. We then computed Pearson correlation coefficients comparing the true parameter values (used for each participant's the simulated data) with the recovered values. Supplementary Table 1, Pearson coefficients for prospect theory parameters, shows the key prospect theory parameters governing heuristic use of reward and probability information (γ and α, see Supplementary Fig. x for visualization of their effects on subjective probability and utility functions) had poor reliability (r = .64 and .43 for γ and α respectively). In contrast, Supplementary Table 2, Pearson coefficients for the Additive Heuristic model show that the key parameters governing heuristic use of reward and probability information (β_{prob} and β_{reward} respectively) had substantially better identifiability (r = .97 and .95 respectively).

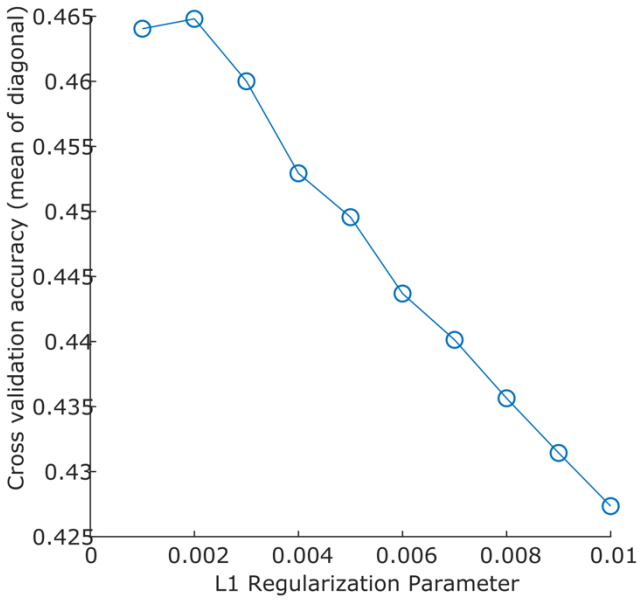
Recovered Parameters					
True Parameters		β	γ	δ	α
	β	***.54***	*.26*	*.04*	*.00*
	γ	*.36*	***.64***	*.01*	*-.04*
	δ	*-.05*	*-.01*	***.62***	*.51*
	α	*.63*	*.37*	*.27*	***.43***

Supplementary Table 1: Recoverability of prospect theory parameters. Each cell shows Pearson correlation coefficient comparing simulated (true) parameters to best fit parameters across simulated participants.

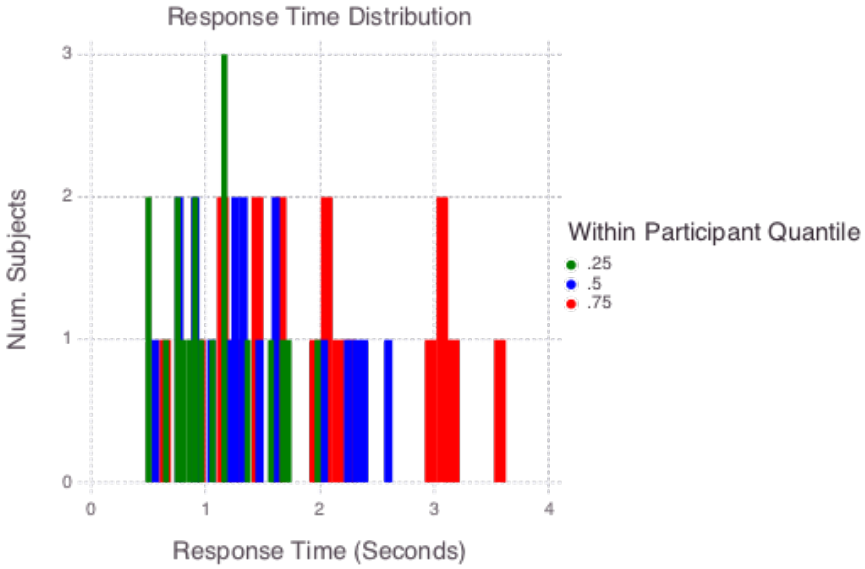
Recovered Parameters					
True Parameters		β_{gain}	β_{loss}	β_{prob}	β_{reward}
	β_{gain}	**.93**	.03	.00	.00
	β_{loss}	.03	**.91**	-.02	-.02
	β_{prob}	.02	-.02	**.97**	-.05
	β_{reward}	-0.01	-.03	.06	**.95**

Supplementary Table 2: Recoverability of additive heuristic parameters. Each cell shows Pearson correlation coefficient comparing simulated (true) parameters to best fit parameters across simulated participants.

Taken together, a better performance in explaining participants choices, and a substantially better parameter identifiability, support the use of the additive heuristic model in parameterizing participant's use of reward and probability information in choice.



Supplementary Figure 4. Mean cross validation accuracy, across train/test-time points from 0-500ms (diagonal of Fig. 3c), across participants. We selected the L1 regularization parameter that maximized this score.



Supplementary Fig. 5. 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.

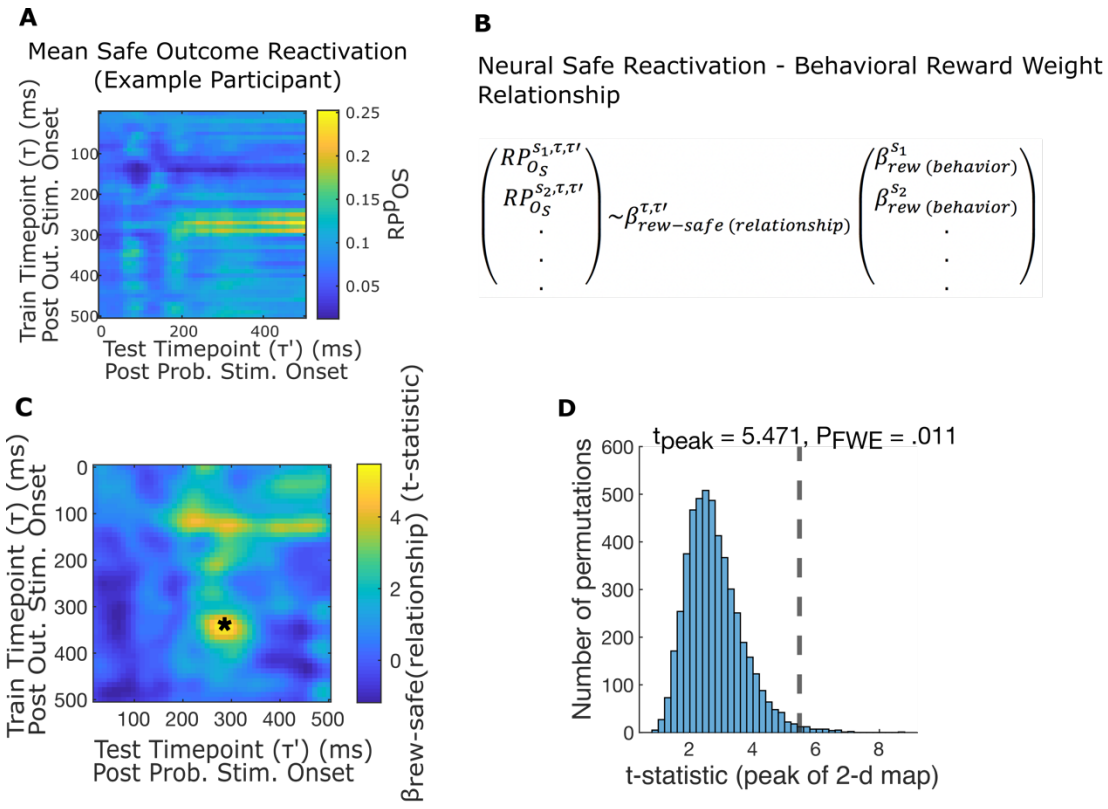
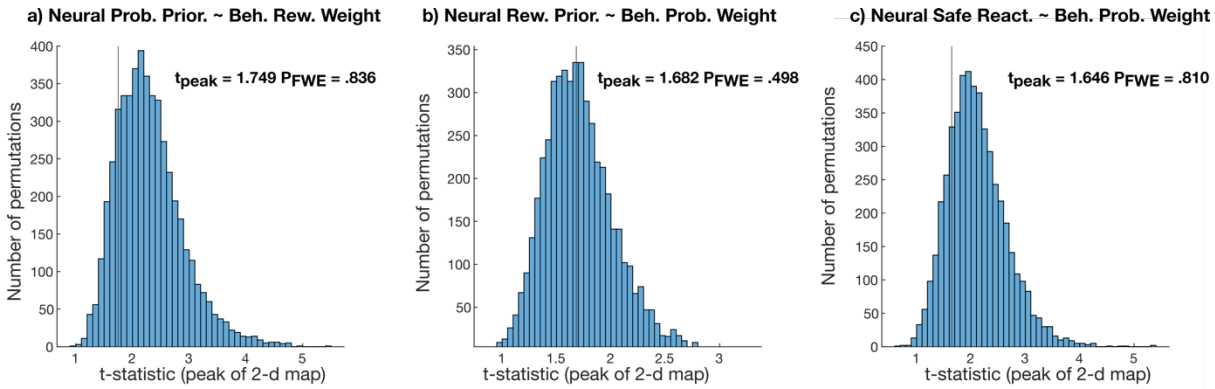


Fig. 6. Behavioral sensitivity to reward information relates to greater reactivation of safe outcome representation. *In order to measure a tendency to reactivate a safe outcome representation, we computed the mean reactivation probability of the safe outcome representation,* $RP_{O_S}^{s,\tau,\tau'}$ *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-d) 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.*



Supplementary Fig. 7. Relationships between MEG outcome reactivation 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.

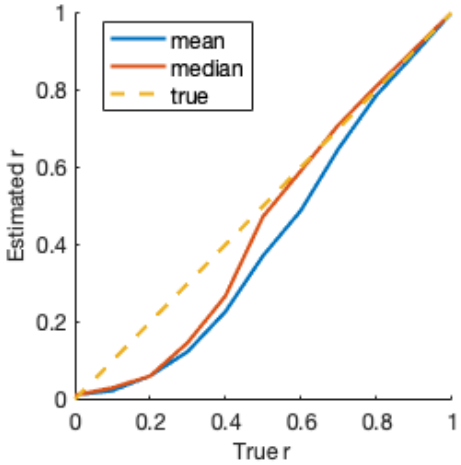
Estimating Behavioral-Neural Correlations

A central challenge with estimating effect sizes of relationships (e.g. r) between behavioral and neural measures (e.g. Neural Probability Prioritization versus Behavioral Probability Weight) is selecting a train and test time-point at which to measure the relationship. Whereas the significance of the relationship in general can be assessed by selecting the max t-statistic over the 2-d map of train and test time-points and comparing this to a null distribution of max-t statistics (generated from shuffled assignments of behavioral parameters to participants), this approach cannot be used to identify an unbiased effect size of the relationship. Notably, the effect-size at the maximum train and test time-point is biased due to the selection process. One solution here is to use a set of held-out participants in order to select a time-point (by selecting the train and test time-points which maximize the relationship), and use this held out set to evaluate the effect size at that time-point. However, using only a portion of the data to identify a time-point risks misidentifying the train and test time-points at which the relationship occurs, and necessarily uses less data to evaluate the relationship.

Consequently, we used a variant of this approach, with the aim of making maximal use of our available data. For each participant, *j*, we selected a time-point using all participants except *j*, taking the time-point which maximized the relationship. We then took the neural measure from participant *j* at that train and test time-point. Repeating this, treating each participant as the held-out participant, provided for each participant an unbiased held-out estimate of the neural effect. We then evaluated the Pearson correlation coefficient between these between-participant held-out, unbiased, neural measurements and the between-participant behavioral measurements.

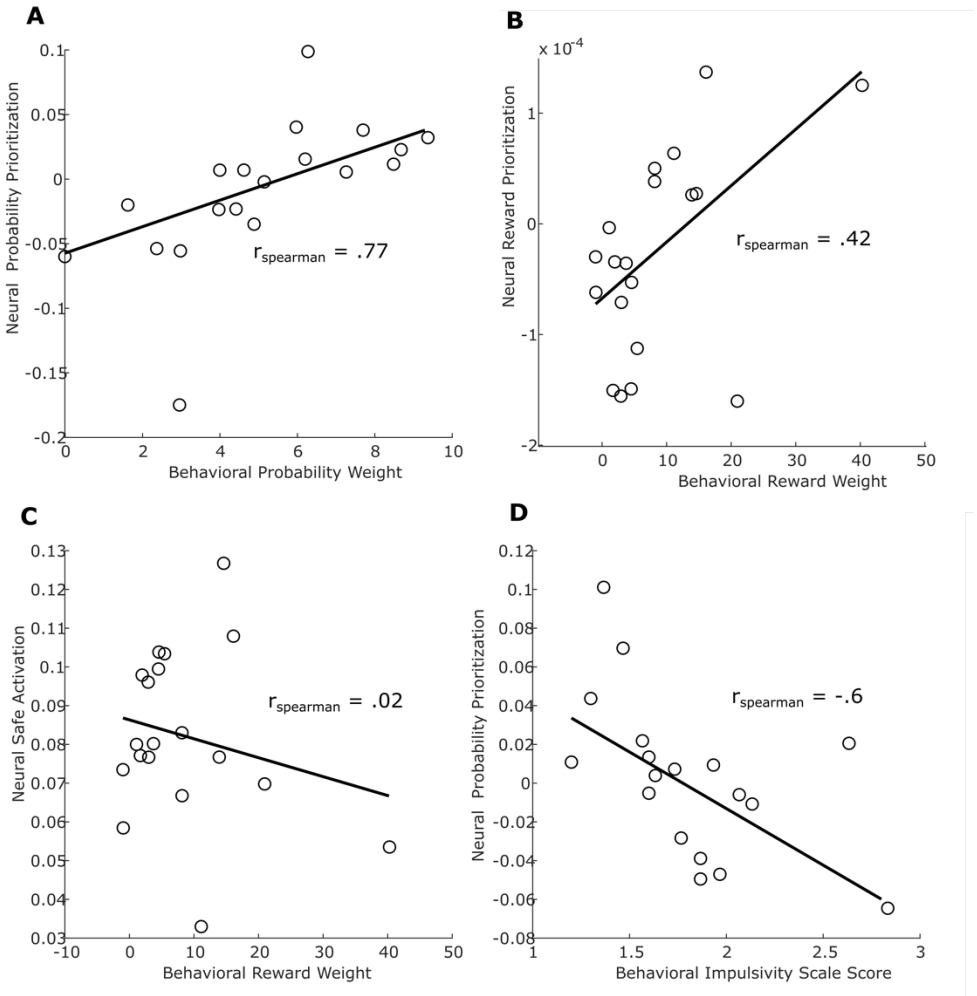
We implemented a simulation to ensure this approach does not produce biased correlation values. For each iteration of the simulation, we generated 19 (one for each participant) behavioral and neural measurements with some true correlation (by sampling from a multi-variate normal distribution). These neural measurements were then embedded into a matrix with one column for each time-point (Number-of-Time-Points=10 X 19 participants), where the other columns corresponding to alternative time-points comprised neural measurements generated to be uncorrelated with the behavioral measurements. We then estimated the correlation using the above described procedure of estimating a time-point separately for each held-out participant (using the non-held out participants) and evaluating the neural measurement for that held-out participant at the selected time-point. Supplementary Fig. x

shows the findings, for each true correlation coefficient (true r), based upon repeating this process 1000 times and taking either the mean or median estimated correlation value. As apparent from the plot, the estimated correlations are not inflated, and in fact are biased toward lower values, due the procedure frequently failing to select the appropriate time-point at which to assess participant's measurements.

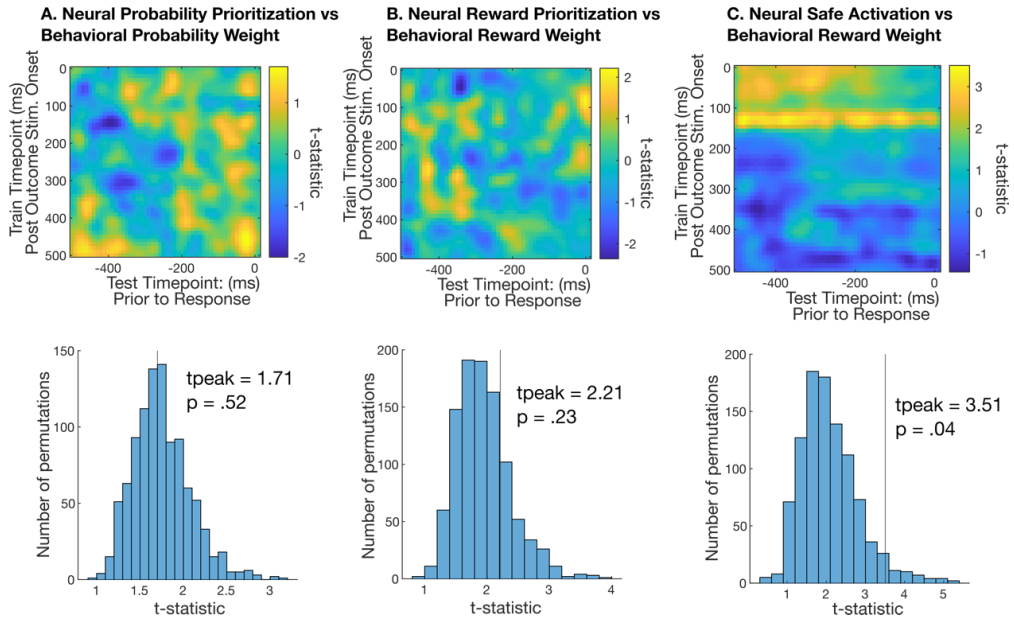


Supplementary Figure 8: Results of simulating procedure for estimating correlations by repeatedly using all participants but one to select a time-point.

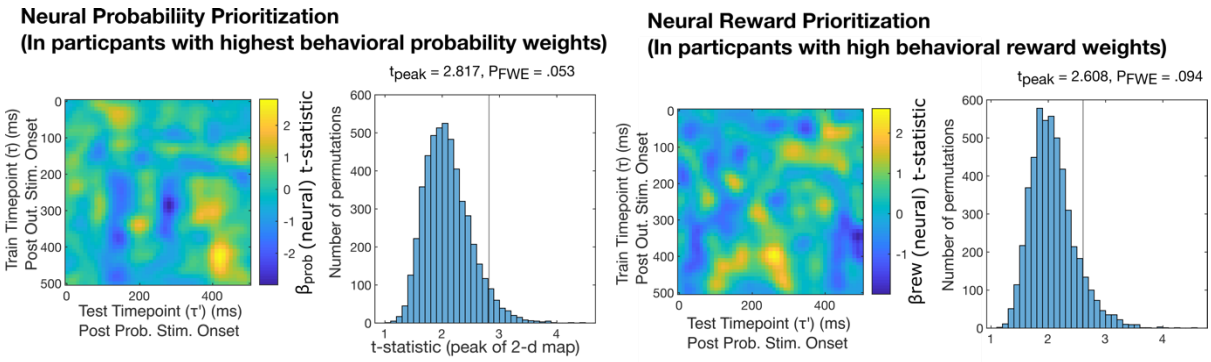
Supplementary Figure 9 shows the results of applying this procedure to our four key behavioral-neural relationships presented in the main text. We found that this procedure generated strong correlation values for all the relationships, except for the relationship between Behavioral Reward Weight and Neural Safe Activation. We think this is because this relationship, in the main-text, is strongly influenced by a single participant who has a high Behavioral Reward Weight, and also large Neural Safe Activation at a particular train and test-time-point. However, when this participant is unable to contribute to time-point selection, their Safe Activation is taken at a different activation at which it is not high.



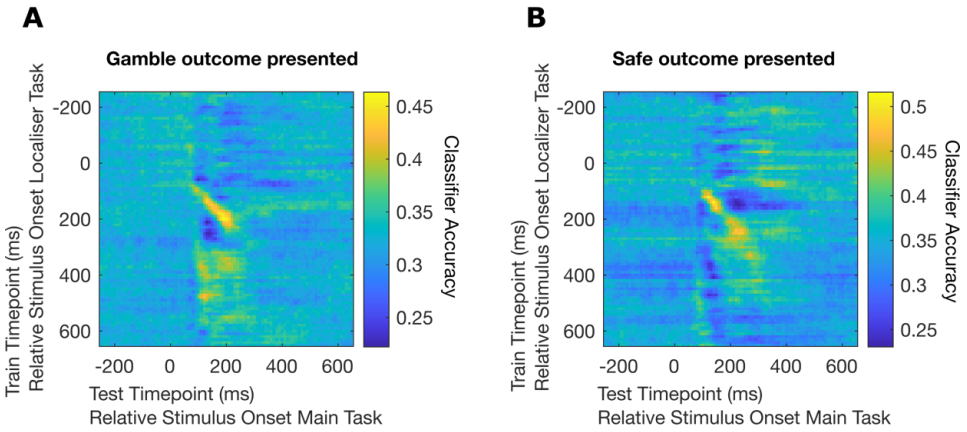
Supplementary Fig. 9: Correlations derived from unbiased estimation process. A. Relationship between Neural Probability Prioritization and Behavioral Probability weight. B. Relationship between Neural Reward Prioritization and Behavioral Reward Weight. C. Relationship between Neural Safe Activation and Behavioral Reward Weight. D. Relationship between Neural Probability Prioritization and Behavioral Impulsivity Scale (BIS) Score.



Supplementary Figure 11. *Analysis of key effects locked to response time. When locked to the time of response, we only observed a relationship between Neural Safe Activation and Behavioral Reward Weight. This suggests that key reactivation events occurred in a manner more locked to presentation of the Probability stimulus, rather than to the response. A) Relationship between Behavioral Probability Weight and Neural Probability Prioritization. B) Relationship between Behavioral Reward Weight and Neural Reward Prioritization. C) Relationship between Behavioral Reward Weight and Neural Safe Activation.*



Supplementary Fig. 12 *Examining reactivation in participants with highest Behavioral Probability Weights (A) and highest Behavioral Reward Weights (B). A) Neural Probability Prioritization in participants with highest 33 percentile Behavioral Probability Weights. B) Neural Reward Prioritization in participants with highest 33 percentile Behavioral Reward Weights.*



Supplementary Figure 13. *Classification accuracy when training on localizer task and testing on outcomes presented in the decision making task. A) Accuracy (out of three outcome stimuli) when the true outcome stimulus is a gamble outcome. B) Accuracy (out of three outcome stimuli) when the true outcome stimulus is the safe outcome.*