HHS Public Access
Author manuscript
Nat Hum Behav. Author manuscript; available in PMC 2020 March 09.

Published in final edited form as:
Nat Hum Behav. 2019 November ; 3(11): 1215–1224. doi:10.1038/s41562-019-0714-3.

Flexible combination of reward information during choice under uncertainty

Shiva Farashahi[1], **Christopher H. Donahue**[2,3], **Benjamin Y. Hayden**[4], **Daeyeol Lee**[3,5], **Alireza Soltani**[1,*]

[1]Department of Psychological and Brain Sciences, Dartmouth College, NH 03755, USA

[2]The Gladstone Institutes, San Francisco, CA 94158, USA

[3]Department of Neuroscience, Yale School of Medicine, New Haven, CT 06510, USA

[4]Department of Neuroscience and Center for Magnetic Resonance Imaging, University of Minnesota, Minneapolis, MN 55455, USA

[5]The Zanvyl Krieger Mind/Brain Institute, Department of Neuroscience, Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

Abstract

A fundamental but rarely contested assumption in economics and neuroeconomics is that decision-makers compute subjective values of risky options by multiplying functions of reward probability and magnitude. In contrast, an additive strategy for valuation allows flexible combination of reward information required in uncertain or changing environments. We hypothesized that the level of uncertainty in the reward environment should determine the strategy used for valuation and choice. To test this hypothesis, we examined choice between risky options in humans and monkeys across three tasks with different levels of uncertainty. We found that whereas humans and monkeys adopted a multiplicative strategy under risk when probabilities are known, both species spontaneously adopted an additive strategy under uncertainty when probabilities must be learned. Additionally, the level of volatility influenced relative weighting of certain and uncertain reward information and this was reflected in the encoding of reward magnitude by neurons in the dorsolateral prefrontal cortex.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

[*]Corresponding author: Alireza Soltani (soltani@dartmouth.edu; ORCID is linked to account on Manuscript Tracking System).

Author contributions
A.S. conceived the project. C.H.D., B.Y.H. and D.L. designed the experiments in monkeys. S.F. and A.S. designed human experiments. S.F. and A.S. performed model simulations and analyzed the data. C.H.D. and S.F. conducted the experiments. C.H.D., S.F., D.L., B.Y.H., and A.S. analyzed and interpreted the experimental data. D.L., B.Y.H., and A.S wrote the manuscript and all other authors contributed to revising the manuscript.

Competing interests
The authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

Most models of decision making assume that while evaluating risky options, we combine information about reward probability and stakes multiplicatively [1-3]. This approach, however, suffers from a major limitation: it cannot readily accommodate flexible weighting of reward information unless utility and probability weighting functions change dynamically, both of which are assumed to be fixed. An additive strategy for value construction (e.g., a linear combination of reward probability and magnitude) can produce a behavior similar to that of a multiplicative strategy when proper weighting is adopted [4], but more importantly, allows greater flexibility in differential weighting of reward information. Such flexible weighting is necessary when reward outcomes are probabilistic and corresponding probabilities must be estimated, resulting in different levels of uncertainty that needs to be considered for optimal combination of information [5,6]. Additionally, statistics of reward outcomes (e.g., mean reward probabilities) can change over time, giving rise to unexpected uncertainty (volatility) that requires further adjustments in learning and decision making [7,8].

One could argue, however, that if representations of reward attributes follow an exponential form, a multiplicative model becomes an additive one (Supplementary Note 1). Although this scenario seems plausible, the combination of reward information for valuation cannot be considered separately from the subsequent decision-making processes [4]. This is because an additive but not a multiplicative strategy implies decision making based on direct comparisons of reward attributes. Therefore, the fundamental difference between additive and multiplicative strategies is whether different reward attributes of each option are fused before the onset of decision-making processes or not. This distinction is important because fusion of reward attributes hinders further adjustments of the weight of each attribute on valuation and/or choice.

Considering the importance of flexibility in value-based choice, we hypothesized that the level of uncertainty in the reward environment should strongly influence the strategy used for valuation and choice. More specifically, we assumed that progressing from choice under risk when reward probabilities are known, to choice under uncertainty when probabilities must be learned, decision-makers should shift from a multiplicative to an additive strategy. In addition, we hypothesized that the levels of uncertainty associated with different pieces of reward information should further affect how they are weighted and combined for computing subjective value or making decisions.

Results

To test our hypotheses, we analyzed choice behavior and neural data from monkeys and humans, two species that have been extensively used to study decision making, performing three comparable tasks that involve only risk or different levels of uncertainty caused by changes in reward probabilities over time. During the gambling task, human participants [9] and monkeys [10] selected between pairs of gambles with independently assigned values of reward probability and magnitude (Fig. 1a,c). In this task, information about reward probabilities and magnitudes was respectively provided explicitly by the length/area and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

color of rectangular bars that represented the two options and therefore, the gambling task involved risk and not uncertainty. A previous analysis of gambling task in monkeys confirmed that all monkeys considered both lengths/areas and colors of gambles for making decisions [11,12]. During both the mixed learning (mL) and probabilistic reversal learning (PRL) tasks, human participants and monkeys [13,14] chose between two colored options that provided different amounts of reward with probabilities that had to be estimated using reward feedback. Reward magnitudes, however, were signaled by the number of dots around each given choice option for monkeys and displayed numbers for humans (see Methods and Fig. 1b,d). The mL task consisted of a stable and volatile environment; reward probabilities associated with different options were fixed in the stable environment but underwent reversals in the volatile environment. As a result, the stable environment of the mL task involved expected uncertainty associated with fixed reward probabilities that the animal was required to learn using feedback, whereas the volatile environment involved both expected and unexpected uncertainty (volatility). Finally, in the PRL task, reward probabilities underwent reversals across different blocks of trials in order to induce different levels of volatility. More specifically, the amount of volatility associated with reward probability was manipulated by changing the length of blocks in which reward probabilities were fixed (i.e., how often probabilities switched).

Identifying adopted strategies under risk and uncertainty.

To determine valuation strategies adopted by individual human participants or by each monkey during individual sessions of the experiment, we fit choice behavior using various models. These models assume different functions for how reward probability and magnitude are assessed subjectively and rely on either an additive or multiplicative strategy for combination of reward information. More specifically, we compared four variations of additive and multiplicative models in which linear and non-linear transformations of reward probability (via probability weighting function), reward magnitude (through utility function), or their combinations were included. Additionally, we also considered hybrid models that include both additive and multiplicative components. We used the Bayesian model selection (BMS)[15] method and Akaike information criterion (AIC) to identify the model that captures choice data the best (see Methods).

We first ensured that our fitting procedure can correctly identify the specific strategy adopted by individual participants and can accurately estimate relevant parameters. To that end, we generated choice data using a hybrid model over a wide range of model parameters in the gambling and three environments of the PRL tasks, and fit the simulated data using the same hybrid model (see Methods for details). We found that the hybrid model can successfully retrieve the two main parameters used for generating data (Supplementary Fig. 1): relative weight of the multiplicative component (β_{mult}); and relative weight of reward magnitude to that of reward probability (β_m/β_p), which we refer to as magnitude-to-probability weighting. We also fit the same data using the linear additive and multiplicative models and used BMS across all three models to examine model identification as a function of the parameters used to generate the data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Overall, our fitting method was able to identify the hybrid model as the most likely model followed by the additive and multiplicative when β_{mult} is close to 0 and 1, respectively (Supplementary Fig. 2). We found small errors in identification of the more dominant model for data generated for the mL task. In the PRL task, however, this error was larger and depended on β_m/β_p but mostly occurred when β_{mult} was around 0.5. For very small β_m/β_p values, the model identification was biased toward a multiplicative one but identification bias shifted toward an additive model as β_m/β_p became closer to 1 ($\log(\beta_m/\beta_p) = 0$). Considering that β_m/β_p estimated from our participants were very small ($\log(\beta_m/\beta_p) < -1$) in the PRL task (see below), these results suggest that error in model identification is small. Overall, fitting simulated data illustrate that the correct model, parameters of the hybrid model, and the dominant component of a hybrid model can be accurately retrieved using our fitting procedure.

Flexible adoption of valuation strategies in humans and monkeys.

We next used all four variations of additive, multiplicative, and hybrid models to fit individual participants' choice data in each experiment. We found that during the gambling task (choice under risk), multiplicative models and hybrid models with a dominant multiplicative component were the most likely models adopted by both monkeys and humans (Fig. 2a,d; Supplementary Fig. 3a,f; Table 1). In contrast, during the mL task (choice under uncertainty), additive models and hybrid models with a dominant additive component were the most likely models adopted by both monkeys and humans (Fig. 2b,c,e,f; Supplementary Fig. 3b,c,g,h; Table 1). Interestingly, this was true for both the stable and volatile environments indicating that an additive strategy was adopted when reward probabilities must be learned.

We found consistent results for the PRL task that also required learning of reward probabilities. Specifically, fitting choice behavior in the PRL task revealed that additive models and hybrid models with a dominant additive component were the most likely models adopted by both humans and monkeys (Fig. 3; Supplementary Fig. 3d,e,i,j; Table 1). Together, these results across three tasks illustrate that both monkeys and humans adopt a predominately multiplicative strategy under risk, whereas both switch to a predominately additive strategy under uncertainty.

To ensure that all human participants included in our data analyses actually learned reward probabilities associated with the two options during the mL and PRL tasks, we removed participants who overall did not choose the option with the higher probability of reward more than chance level (0.5; see Methods). This resulted in exclusion of 4 and 12 participants in the mL and PRL tasks, respectively. To confirm that our exclusion criteria did not bias our results in terms of the adopted strategy, we also fit the choice data from the excluded participants in the PRL task (the mL task had too few excluded participants). We did not find any credible evidence that the excluded participants adopted a strategy qualitatively different from the one used by the remaining participants (Supplementary Fig. 4). Instead, the most parsimonious explanation for our data was that the excluded participants simply failed to learn reward probabilities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Flexible combination of reward information under uncertainty.

Together, results based on the fit of choice behavior suggest that there is a major and heretofore undetected effect of expected uncertainty on the strategies that participants use to combine information across reward dimensions. Why might an additive approach be favored under uncertainty? We hypothesized that the additive strategy may allow more flexibility and may therefore be favored in uncertain environments in which decision-makers must learn reward attributes and associated uncertainty in order to adjust the weight of each attribute with a different level of uncertainty on the overall value or choice. If true, this also predicts that weighting of a given piece of information should depend on its level of uncertainty and thus, the relative weighting of reward information should change according to volatility of the environment.

To test this prediction, we compared the effect of volatility on how different pieces of reward information were combined using the estimated model parameters based on the simplest hybrid model in the two environments of the mL as well as PRL tasks. In this model, the additive component was a linear function of reward magnitude and probability, and the multiplicative component was equal to expected value (EV). We focused on the simplest hybrid model so we could directly compare behavioral and neural adjustments in monkeys and interpret our results more clearly (see the last section of Results). We found that the relative weighting of reward information differed in the two environments of the mL task in both species (Fig. 4a,c). More specifically, both monkeys and human participants exhibited significantly larger magnitude-to-probability weighting in the volatile compared with the stable environment (one-sided Wilcoxon signed-rank test; monkeys: median±IQR: 0.53±1.34, $P < 0.001$, $d = 0.54$, $N = 316$, 95% CI = [0.49 0.76]; humans: median±IQR: 0.47±1.54, $P < 0.001$, $d = 0.45$, $N = 46$, 95% CI = [0.14 0.89]). Similarly, magnitude-to-probability weighting was larger in the more volatile compared with the less volatile environment of the PRL task (Fig. 4b,d; one-sided Wilcoxon signed-rank test; monkeys: median±IQR: 0.16±0.57, $P < 0.001$, $d = 0.22$, $N = 118$, 95% CI = [0.08 0.26]; humans: median±IQR: 0.20±1.81, $P = 0.04$, $d = 0.29$, $N = 38$, 95% CI = [−0.01 0.27]). These results confirm our prediction and illustrate that the relative weighting of the certain (i.e., reward magnitude) to that of the uncertain information (i.e., reward probability) increased as the reward environment became more volatile increasing uncertainty in reward probability.

If changes in volatility cause adjustments in the relative weighting of reward information, then one would predict that there would be larger changes between the stable and volatile environments of the mL task compared with the less and more volatile environments of the PRL task. This is because the range of uncertainty in reward probabilities was larger in the mL task than in the PRL task. Consistent with this prediction, we found that the differences between magnitude-to-probability weighting in the volatile and stable environments of the mL task to be larger than the differences between magnitude-to-probability weighting in the more and less volatile environments of the PRL task; this effect, however, was only significant in monkeys (one-sided Wilcoxon rank-sum test; monkeys: $P < 0.001$, $d = 0.33$, $N = 432$, 95% CI = [0.07 0.55]; humans: $P = 0.1$, $d = 0.06$, $N = 82$, 95% CI = [−0.2 0.68]). Interestingly, we did not find any consistent evidence for effects of volatility on the extent to which an additive strategy was adopted using the likelihood of sessions in monkeys (or

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

human participants) with additive and hybrid strategies, or using estimated β_{EV} values (Figs. 2, 3; see Table 2 for detailed statistics). Together, these results suggest that volatility associated with reward probability can strongly influence how this information is weighted relative to reward magnitude.

Adjustments of learning to uncertainty.

It has been previously shown that volatility of the environment influences the learning rates [16]. Therefore, we also compared the estimated learning rates between the two environments of the mL and PRL tasks. We found that in the mL task, the learning rates were larger in the volatile compared with the stable environment (Fig. 5; c.f., Figure 2 of [14]). In contrast, we did not find any credible evidence for an increase in the learning rates between the less and more volatile environments of the PRL task for monkeys or humans (Supplementary Fig. 5). Less consistent effects of volatility on the learning rates compared with effects of volatility on the relative weighting of reward information suggests that changes in weighting of reward information might be a more fundamental adjustment to volatility in the reward environment.

We should note that the absence of volatility effects on the learning rates should not be considered as the lack of evidence for adjustments in learning processes in the PRL task. As we have previously shown [8], learning of the better and worse options (in terms of reward probability) follows different dynamics in the less and more volatile environments of the PRL task (Supplementary Fig. 6a). Nevertheless, we performed additional analyses of choice behavior in the PRL task to directly (without using model fitting) show that volatility influences both the speed of learning and the relative weighting of reward information (Supplementary Fig. 6c; Supplementary Note 2).

Flexible representations of reward information in the prefrontal cortex.

Finally, we looked for neural correlates of these behavioral adjustments in the activity of the dorsolateral prefrontal cortex (dlPFC) neurons recorded during the PRL task that consist of two comparable environments with different levels of volatility [13]. We first used two separate multiple linear regression models for the two environments in order to characterize neural response to various events/signals in the current and previous trials, and any changes in these responses due to volatility (see Methods). We found a significant difference in a few regression coefficients between the two environments across all neurons. This includes the relative positions of target colors (POS_{RG}), the previous chosen color ($C_{RG}(t\text{-}1)$), the interaction of the positions of target colors and previous chosen color ($POS_{RG} \times C_{RG}(t\text{-}1)$), and the difference in and sum of reward magnitudes of the two options presented on each trial ($m_r - m_l$, and $m_r + m_l$) (Fig. 6a-d, Supplementary Fig. 7a-f; see Eq. 11 in Methods).

The regression coefficient for the difference in reward magnitude quantifies how strongly this variable is encoded. Considering the relevance of this encoding for an additive integration or direct comparison of reward attributes, we next examined the relationship between adjustments in the dlPFC activity and behavior in response to changes in volatility of the environment. Among “magnitude-difference selective” neurons, we found a significant positive correlation between behavioral adjustments and changes in encoding of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the difference in reward magnitudes for the two options (Kendall correlation ($N = 45$): $r = 0.27$, $P = 0.018$, 95% CI = [0.045 0.43]; Fig. 6f). Specifically, a stronger dlPFC encoding of the difference in magnitudes accompanied larger behavioral weighting of reward magnitude relative to reward probability (magnitude-to-probability weighting) in the more volatile environment. In contrast, we did not find credible evidence for correlation between behavioral adjustments and changes in encoding of any other variables that were significantly represented in neural activity including the sum in reward magnitudes of the two options (Kendall correlation ($N = 45$); $m_r + m_l$: $r = -0.052$, $P = 0.5$, 95% CI = [−0.21 0.10]; POS_{RG}: $r = -0.082$, $P = 0.4$, 95% CI = [−0.29 0.11]; C_{RG}(t-1): $r = 0.063$, $P = 0.5$, 95% CI = [−0.14 0.26]; $POS_{RG} \times C_{RG}$(t-1): $r = 0.003$, $P = 1$, 95% CI = [−0.18 0.20]; Fig. 6e, Supplementary Fig. 7g-i). Together, analyses of neural data suggest a direct link between observed behavioral adjustments and adjustments in the representation of the most relevant variable (i.e. the difference in reward magnitudes) in the dlPFC neurons.

Discussion

Using three tasks with different levels of uncertainty in monkeys and humans, we examined how these two species adjust to uncertainty in the reward environment. Our results demonstrate that under uncertainty, that is, when reward probabilities have to be estimated from reward feedback, both humans and monkeys spontaneously adopt an additive strategy for valuation (i.e., a linear combination of reward probability and magnitude or functions of them). In contrast, both species adopt a multiplicative strategy (i.e., multiplying functions of reward probability and magnitude) under risk when reward probabilities are known. The additive strategy allows humans and monkeys to dynamically adjust their weighting of reward magnitude relative to reward probability based on the environmental volatility associated with reward probability. These behavioral adjustments in turn are accompanied by corresponding adjustments in the strength of reward-magnitude encoding in the dlPFC, suggesting that prefrontal neurons could flexibly adjust their representations of task-relevant information according to the level of uncertainty in the environment [14].

A fundamental difference between the multiplicative and additive strategies is that different reward attributes have to be fused for each option before the onset of decision-making processes in the former but not necessary the latter. This is because an additive strategy for the construction of subjective value (followed by choice) is equivalent to decision making based on the weighted sum of the differences in each dimension; that is, choice can be made by direct comparisons of reward attributes in each dimension separately [17]. The difference between the two strategies has important implications for the flexibility of choice behavior because fusion of reward attributes results in an integrated value that hinders further independent adjustments of the weight of each attribute on valuation and choice.

Traditional approaches to behavioral economics and neuroeconomics hold that laboratory measures of economic attitudes (especially with regards to risk and time) measure stable and universal preferences and strategies. However, recent empirical evidence supports the idea that, in humans and other animals, economic preferences are constructed on the fly and vary substantially based on ostensibly small contextual factors [18]. For example, it has been shown that humans adaptively adjust their choice behavior according to statistics of attended

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

variables, time to receive the reward, and current resources [19-21]. Similarly, animal studies of decision making have demonstrated time-dependent, task-dependent, and sequence-dependent choice preferences [11,22,23]. Therefore, the present findings are consistent with the broader evidence that human and animal decision-makers are not hard-wired to follow fixed strategies assumed by normative models, and instead, are endowed with flexibility needed for learning and choice under uncertainty.

Finally, previous studies have shown that the anterior cingulate cortex (ACC) carries signals related to volatility [16] and is crucial for leaning from reward feedback under uncertainty [24]. Here, we find that the dlPFC neurons change their encoding of reward magnitude according to volatility of the environment, suggesting that volatility information may be routed from the ACC to the dlPFC to support flexible behavior under uncertainty. However, future manipulation studies are required to test this prediction in order to better elucidate circuit-level mechanisms of adaptive learning [25].

Methods

Ethics statement.

All experimental procedures in monkeys were approved by the Institutional Animal Care and Use Committee (IACUC) at Yale University, the University Committee on Animal Resources at the University of Rochester, or the Institutional Animal Care and Use Committee at the University of Minnesota. All experimental procedures in humans were approved by the Dartmouth College Institutional Review Board, and informed consent was obtained from all participants before participating in the experiment.

Animal preparation.

For the gambling task, three male rhesus monkeys (B, C and J) were used. All three monkeys were habituated to laboratory conditions and then trained to perform oculomotor tasks for liquid reward. Eye position was sampled at 1,000 Hz by an infrared eye-monitoring camera system (SR Research). Stimuli were controlled by a computer running Matlab (Mathworks Inc.) with Psychtoolbox [26] and Eyelink Toolbox [27]. For the probabilistic reversal learning task (PRL), two male rhesus monkeys (O and U) were used. Monkey O had been previously trained on a manual joystick task but had not been used for electrophysiological recordings before this experiment. Monkey U had not been used for any prior experiments. Both animals were socially housed throughout these experiments. For the mixed learning (mL) task, two male rhesus monkeys (U and X) were used. Monkey U had been trained on the probabilistic reversal learning (PRL) task, while Monkey X had not been used for any prior experiments. Eye movements were monitored at a sampling rate of 225 Hz with an infrared eye tracker (ET49, Thomas Recording, Germany). Stimuli were controlled using Orion or Picto custom code written in C++ (https://medicine.yale.edu/lab/dlee/technology).

Neurophysiological recording.

For the PRL task, activity of individual neurons in the dorsolateral prefrontal cortex (dlPFC) was recorded extracellularly (left hemisphere in both monkeys) using a 16-channel multi-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

electrode recording system (Thomas Recording, Germany) and a multichannel acquisition processor (Plexon, TX). On the basis of magnetic resonance images, the recording chamber was centered over the principal sulcus and located anterior to the genu of the arcuate sulcus (monkey O, 4 mm; monkey U, 10 mm). All neurons selected for analysis were located anterior to the frontal eye field, which was defined by eye movements evoked by electrical stimulation in monkey O (current <50 μA). The recording chamber in monkey U was located sufficiently anterior to the frontal eye field, so stimulation was not performed in this animal. Each neuron in the data set was recorded for a minimum of 320 trials (77 and 149 neurons in monkeys O and U, respectively), and on average for 518.8 trials (s.d. = 147.2 trials). We did not preselect neurons on the basis of activity, and all neurons that could be sufficiently isolated for the minimum number of trials were included in the analyses.

Human participants.

For the gambling task in humans, 64 participants (38 females; ages 18–22 years) were recruited from the Dartmouth College student population. No participant was excluded from data analyses for the gambling task. For the mL and PRL tasks in humans, 50 participants (35 females; ages 18–22 years) were recruited from the Dartmouth College student population. Because both the mL and PRL task involved learning reward probabilities associated with the two color targets, we used a criterion to remove participants whose performance (in terms of selecting the option with the higher probability of reward) was not significantly better than chance (0.5). More specifically, we used a performance threshold of 0.5513 equal to 0.5 plus 2 times s.e.m., based on the average of 380 trials after excluding the first 10 trials of each environment. This resulted in the exclusion of data from 4 of 50 participants in the mL task and 12 of 50 participants in PRL task, respectively. No participants had a history of neurological or psychiatric illness.

Participants in all the experiments were compensated with a combination of money and “T-points,” which are extra credit points for classes within the Department of Psychological and Brain Sciences at Dartmouth College. The base rate for compensation was $10/h or 1 t-point/h. Participants were then additionally rewarded based on their performance by up to $10/h.

Gambling task in monkeys.

Three male monkeys performed 70,700 (monkey B), 24,700 (monkey C), and 12,872 trials (monkey J) of a gambling task for a total of 146 sessions and 108,272 trials. On each trial of this task, they selected one of two options (Fig. 1a). Options offered either a gamble or a safe (100% probability) bet for liquid (water or dilute cherry juice, depending on the animal’s preference) reward. Gamble offers were defined by two parameters, reward size and probability. Each gamble rectangle was divided into two portions: one red and the other either blue or green. The size of the green or blue portions signified the probability of winning a medium (0.165 ml) or large reward (0.24 ml), respectively. Probabilities were drawn from a uniform distribution between 0 and 100%, with 1% precision and excluding upper limit. The rest of the bar was colored red; the size of the red portion indicated the probability of no reward. A safe option existed in 11.11% of trials which was entirely gray and carried a 100% probability of a small reward (0.125 ml).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

On each trial, one offer appeared on the left side of the screen and the other appeared on the right. The side of the first and second offer (left and right) was randomized by trial. Following presentation of both offers individually, both offers appeared simultaneously and the animal indicated its choice by shifting gaze to its preferred offer and maintaining fixation on it. Following a successful fixation, the gamble was immediately resolved and reward delivered [10].

Gambling task in humans.

Each participant performed a gambling task which he/she selected between a pair of offers on every trial and were provided with reward feedback (Fig. 1c). Gambles were presented as rectangular bars divided into one or two portions. A portion's color indicated the reward magnitude of that outcome, and its size signaled its probability. The task consists of either choice between a safe option and a gamble that yields either a reward larger than that of the safe option or no reward with complementary probabilities, or choice between two gambles. Participants evaluated and selected between a total of 63 unique gamble pairs, each of which was shown four times in a random order (total of 252 trials).

Before the beginning of the task, participants completed a training session in which they selected between two safe options. These training sessions were used to familiarize participants with the associations between four different colors (purple, magenta, green, and gray) and their corresponding reward values. Reward values were always 0, 1, 2, and 4 points and no reward (0 points) was always assigned to the gray color. The color-reward assignment remained consistent for each participant throughout both the training session and its corresponding task. The color-reward assignments, however, were randomized between participants [9].

PRL task in monkeys.

Monkey O and U completed 45 and 73 sessions (a total of 118 sessions and 66,148 trials) of the PRL task, respectively, in which, they had to choose between a red and a green circle on each trial (Fig. 1b). A set of yellow tokens was also presented around each target to indicate the magnitude of potential reward on a given target. On each trial, one of the target colors was associated with a high reward probability (80%) whereas the other was associated with the complimentary low reward probability (20%). These reward probabilities were fixed within a block of trials and alternated across blocks of 20 (more volatile) or 80 (less volatile) trials to induce different levels of volatility. That is, the block length L was used to manipulate volatility of the environment. If animals' choice on a given trials was rewarded, they were given the amount of apple juice associated with the magnitude of the chosen target. Each token corresponded to one drop of juice (0.1 ml). The reward magnitudes associated with each target color were drawn from the following ten possible pairs: {(1,1), (1,2), (1,4), (1,8), (2,1), (2,4), (4,1), (4,2), (4,4), (8,1)}. Each magnitude pair was counter-balanced across target locations so that reward magnitude did not provide any information about the location of reward. We did not find any systematic differences in either animal's behavior, therefore we combined the data from both monkeys. More details about the task and behaviors of the animal have been reported previously [13].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mL task in monkeys.

Monkey U and X completed 182 and 134 sessions (a total of 316 sessions and 166,912 trials) of the mL task, respectively (Fig. 1b). This task is very similar in construct to the PRL task in monkeys except in one of the two conditions (stable environment) the reward probabilities associated with a pair of two targets did not change over time. Animals had to choose between a pair of two physically distinct color targets while a set of yellow tokens was also presented around each target, indicating the magnitude of potential reward on a given target. One of the two targets in each pair was associated with a high reward probability (80%) and the other was associated with the complimentary low reward probability (20%). Reward probabilities associated with different targets had to be learned using reward feedback in two conditions: stable environment in which reward probabilities were fixed, and volatile environment in which reward probabilities underwent reversals similarly to the PRL task but with the block length sampled randomly from 20 and 40. The physical characteristics of the targets indicated the two conditions of the tasks. Red and green target were used in the volatile environment whereas pairs of orange and cyan targets or pairs of white diamond and white square targets were used in the stable environment. The animals received the magnitude of juice associated with the chosen target, with each token equal to one drop of juice (0.1 ml). The magnitudes associated with each target were drawn from the following ten possible pairs: {(1,1), (1,2), (1,4), (1,8), (2,1), (2,4), (4,1), (4,2), (4,4), (8,1)}. Each magnitude pair was counter-balanced across target locations. More details about the task have been reported previously [14].

mL and PRL tasks in humans.

Each participant completed two sessions of the experiment, corresponding to the mL and PRL tasks, in which she/he was asked to choose between blue and red or cyan and magenta squares, respectively (Fig. 1d). The magnitude of potential reward (reward points) on a given target was presented as yellow numbers inside each target. Participants were told to select between the two targets on the basis of both presented reward magnitude on each trial and the experienced outcomes associated with each color in the preceding trials in order to maximize the total number of reward points.

The first session of the experiment (the mL task) started with 200 trials in which the probability of either red or blue target being rewarded was fixed at 80% or 20% (stable environment). This was followed by a super-block of 200 trials in which reward probabilities associated with the two targets switched between 80% and 20% every 20 or 40 trials (volatile environment). The second session of the experiment (the PRL task) started with either a super-block of 160 trials in which reward probabilities for the two targets switched every 20 trials (more volatile environment) followed by a super-block of 240 trials in which reward probabilities for the two targets switched every 80 trials (less volatile environment), or vice versa. The order of the less and more volatile environments were counter balanced across participants. Throughout the experiment, reward magnitudes were selected from the following ten possible combinations: {(1,1), (1,2), (1,4), (1,8), (2,1), (2,4), (4,1), (4,2), (4,4), (8,1)} similar to the mL and PRL tasks in monkeys. The target color associated with the higher probability of reward during the initial block of each session of the experiment was randomly assigned and counter-balanced across participants.

Author Manuscript

Analysis of behavioral data.

For both the mL and PRL tasks, we first fit choice data using a reinforcement learning (RL) model to estimate reward probabilities learned by each participant over time. More specifically, we tested an RL model with two learning rates for rewarded and unrewarded trials (α_{rew} and α_{unr}). In this model, the two options (colored targets or distinct shapes) are assigned with complementary probabilities: say $\hat{p}_R$ and $\hat{p}_G = 1 - \hat{p}_R$ for the red and green targets, respectively. We made this assumption because actual reward probabilities were complementary in all our experiments and models based on complementary estimates provided better fits than those based on independent estimates [8]. If the red target is selected, the estimated probability for the red target is updated as follows:

$$\hat{p}_R(t+1) = \hat{p}_R(t) + (\delta_{r(t),1}\alpha_{rew} + \delta_{r(t),0}\alpha_{unr})(r(t) - \hat{p}_R(t)) \qquad \text{(Eq. 1)}$$

where t represents the trial number, $\hat{p}_R(t)$ is the estimated reward probability of the red target on trial t, $r(t)$ is the trial outcome (1 for rewarded, 0 for unrewarded), and α_{rew} and α_{unr} are the learning rates for rewarded and unrewarded trials, respectively, and $\delta_{r(t),X}$ is the Kronecker delta function ($\delta_{r(t),X}$, if $r(t) = X$, and 0 otherwise). On trials when the green target is selected, the update rule is equal to:

$$\hat{p}_R(t+1) = \hat{p}_R(t) + (\delta_{r(t),1}\alpha_{rew} + \delta_{r(t),0}\alpha_{unr})(r(t) - \hat{p}_R(t)) \qquad \text{(Eq. 2)}$$

because of the assumption about the complementary nature of reward probabilities.

To systematically examine how each participant combined reward magnitude and estimated (in the mL and PRL tasks) or given reward probability (in the gambling task) into a subjective value, we compared several variations of models in which probabilities and magnitudes were combined additively or multiplicatively. We also considered hybrid models that combine both additive and multiplicative models.

In the additive models, the subjective value of each gamble is computed as follows:

$$SV_L = \alpha_m u(m_L) + \alpha_p w(p_L) \qquad \text{(Eq. 3)}$$

where SV_L is the subjective value of left gamble, m_L is the magnitude of left gamble, and p_L is the provided or estimated (Eq. 1) reward probability of the left gamble, $u(m)$ is the utility function, $w(p)$ is probability weighting function (see below), and α_m and α_p are the weights assigned to the magnitude and probability, respectively.

In the multiplicative models, the subjective value of each gamble is computed as follows:

$$SV_L = \beta(u(m_L) * w(p_L)) \qquad \text{(Eq. 4)}$$

where β is the inverse temperature.

In the hybrid models, the subjective value of each gamble is computed as follows:

$$SV_L = (\alpha_m u(m_L) + \alpha_p w(p_L)) + \alpha_{mult}\left(u(m_L) * w(p_L)\right) \quad \text{(Eq. 5)}$$

where α_m and α_p are the weights assigned to reward magnitude and probability, respectively, and α_{mult} is the weight of the multiplicative component on the subjective value. We normalize these weights to define a set of relative weights (β_{mult}, β_m, and β_p) as follows:

$$\begin{cases} \beta_{mult} = \dfrac{\alpha_{mult}}{\alpha_m + \alpha_p + \alpha_{mult}} \\ \beta_m = \dfrac{\alpha_m}{(\alpha_m + \alpha_p)} \\ \beta_p = \dfrac{\alpha_p}{(\alpha_m + \alpha_p)} \end{cases} \quad \text{(Eq. 6)}$$

where β_{mult} is the relative weight assigned to the multiplicative component, and β_m and β_p measures the relative weight of reward magnitude and probability in the additive component, respectively. Using these definitions, the subjective value in the hybrid model can be written as:

$$SV_L = \beta \times \left((1 - \beta_{mult})(\beta_m u(m_L) + \beta_p w(p_L)) + \beta_{mult}(u(m_L) * w(p_L))\right) \quad \text{(Eq. 7)}$$

where $\beta = \alpha_m + \alpha_p + \alpha_{mult}$. The model with $\beta_{mult} = 0$ is purely additive and the model with $\beta_{mult} = 1$ is purely multiplicative.

The estimated subjective values are then used to compute the probability of selecting left and right based on a logistic function:

$$logit\ p(Left) = \beta_0 + (SV_L - SV_R) + \beta_{stay} D_{pc} POS_{RG} \quad \text{(Eq. 8)}$$

where $p(Left)$ denotes the probability of choosing the left gamble. The first and third terms only were used to fit choice behavior in the volatile environments to capture the bias in choosing the options on left or right (β_0) and the tendency to repeat the previous chosen target color (β_{stay}), respectively. These terms were confounded with reward values and thus, were not used when probabilities were known as in the gambling task or fluctuated very little as in the stable environment of the mL task. Finally, D_{pc} is a dummy variable ($D_{pc} =$ −1, 1 if the previous choice was green or red, respectively), and $POS_{RG}(t)$ is the relative position of the red and green targets (1 if red is on the right, and −1 otherwise).

We examined four variations of the additive, multiplicative, and hybrid models (EV, EV +PW, EU, and SU) in which the actual or nonlinear transformations of probabilities and magnitudes were combined additively or multiplicatively. In the expected value (EV) models, linear functions of reward probabilities and magnitudes were used to estimate the subjective value of each gamble ($u(m) = m$, $w(p) = p$). In the expected utility (EU) model we considered a nonlinear function of reward magnitude to determine the subjective utility of a given reward outcome:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$u(m) = m^{\rho_G} \quad \text{(Eq. 9)}$$

where ρ_G is the exponents of the power law function and determines risk aversion. However, the probability weighting was linear in this model. In the EV+PW model, we considered a linear function of magnitude and a nonlinear probability weighting function (PW). The PW was computed using the 1-parameter Prelec function as follows:

$$w(p) = e^{-(-log(p))^{\gamma}} \quad \text{(Eq. 10)}$$

where $w(p)$ is the PW and γ is a parameter that determines the amount and direction of distortion in the probability weighting function. Finally, in the SU model, we used both nonlinear utility and nonlinear probability weighting functions to estimate the subjective value of each gamble. This procedure was similar for the mL, PRL, and gambling tasks.

All models were fit to experimental data by minimizing the negative log likelihood of the predicted choice probability given different model parameters using the *fminsearch* function in MATLAB (Mathworks Inc.). To avoid over-fitting and to deal with different numbers of parameters, we applied variational Bayes model selection (BMS) approach to identify the most likely models that could account for our data. We calculated likelihood of each model using the estimated Dirichlet density from which models are sampled to generate participant-specific data [15]. The procedure was repeated 50 times using 80% of the data for a given monkey or human participant in a given task in order to calculate the mean and standard deviation of model likelihood in capturing the data. We confirmed our results by computing the Akaike information criterion (AIC) that penalized the use of additional parameters in a given model. The smaller value for this measure indicates a better fit of choice behavior. Finally, we found similar results using Bayesian information criterion for all experimental data and cross-validation for monkey data for which this method could be applied (data not shown).

To avoid local minima, the fitting procedure was repeated 20 times for data from each monkey and human participant. For more complex models (EV+PW, EU, and SU), we used the estimated parameters of the simplest model (EV) as the initial values for searching the parameters. We adopted this method to ensure that more complex models achieve negative log likelihood not bigger than the best corresponding simplest model, which could happen due to converging to local minima with a large number of parameters.

Validation of the fitting procedure.

To investigate whether our fitting procedure can be used to distinguish between alternative models and accurately estimate model parameters, we simulated choice data using a hybrid model of value construction (Eq. 7 with linear utility and probability weighting functions) in the gambling and the PRL tasks and fit this data using an additive, multiplicative, and hybrid models. The simulated data in the PRL task was generated using an RL model with two learning rates. We constrained β_m and β_{mult} to be in the range [0, 0.5] and [0, 1], respectively, but kept β equal to 5 for all simulations. We simulated 10 sets of choice data, each with 4000 trials in the gambling task and three environments of the PRL task: stable

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

environment with block length of 200, less volatile environment with block length of 80, and more volatile environment with block length of 20. To ensure proper learning, we set the learning rates to 0.4, 0.2 and 0.1 for rewarded and 0.1, 0.05 and 0.025 for unrewarded trials in the more volatile, less volatile, and stable environments, respectively. For simplicity, we used linear utility and probability weighting functions. We then fit the simulated data to estimate model parameters and compute the BMS likelihood. The BMS likelihood and the estimated model parameters were computed by averaging over all fits.

Analysis of neural data.

A linear regression model was used to investigate how individual neurons encode various types of information in the PRL task. We included the terms that were shown to have a neural representation in the dorsolateral prefrontal cortex [5]. To analyze how the way features are encoded by single neurons is influenced by volatility, we compared the fit of two simple regression models to activity in the less and more volatile blocks using the following equation:

$$\begin{aligned} y(t) = {} & \beta_0 + \beta_1 C_{LR}(t) + \beta_2 C_{LR}(t-1) + \beta_3 R(t-1) + \beta_4 POS_{RG}(t) + \beta_5 \\ & (m_r(t) + m_l(t)) + \\ & \beta_6 (m_r(t) - m_l(t)) + \beta_7 C_{RG}(t) + \beta_8 C_{RG}(t-1) + \beta_9 R(t-1) \times POS_{RG}(t) + \\ & \beta_{10} C_{RG}(t-1) \times \\ & \quad POS_{RG}(t) + \beta_{11} C_{RG}(t-1) \times R(t-1) + \beta_{12} PRL(t) + \beta_{13} HVL(t) \end{aligned} \tag{Eq. 11}$$

where $y(t)$ is the firing rate of a neuron for a given epoch on trial t, $C_{LR}(t)$ is the location of the chosen target on trial t, $R(t)$ is the outcome on trial t, $POS_{RG}(t)$ is the position of the red and green target on trial t, $(m_r(t) - m_l(t))$ and $(m_r(t) + m_l(t))$ is the sum and the difference in reward magnitude of left and right targets on trial t, $C_{RG}(t)$ is the color of the chosen target on trial t. The $PRL(t)$ term stands for the location associated with the high reward probability target ($PRL(\mathrm{t}) = C_{LR}(t-1) \times R(t-1)$), and the $HVL(t)$ term indicates the location of color associated with the high reward probability target ($HVL(t) = C_{RG}(t-1) \times R(t-1) \times POS_{RG}(t)$).

To compare the regression coefficients across the two volatility conditions (less and more volatile environment), we randomly removed a subset of trials in each pair of reward magnitudes so that the proportion of trials in which the animal chose the high-reward probability target was equated for the two conditions. While reducing the difference between these proportions we ensured that the lowest number of trials are removed for each condition in each session. We repeated this procedure 50 times for each session (removing different sets of trials in each repetition) and averaged the regression coefficients across the repetitions. The “magnitude-difference” selective neurons are defined as neurons that were selective to the difference in magnitudes considering both sessions together. Finally, we did not correct for multiple comparisons to identify the bins at which a given regressor was significantly different from 0 because of the overlap between spikes in the neighboring bins (due to sliding window). Nevertheless, to assign two successive bins as significant, we required those bins to have significant values and the fractions of neurons with significant regressor to be larger than 0.15. The latter was done to avoid false positives due to small number of neurons. The statistical comparisons were performed using two-sided Wilcoxon

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

signed-rank test. Finally, for correlation analysis, we only considered spikes between 750-1250 ms after target onset when magnitude information was presented on the screen [13].

Relative modulation due to volatility.

To quantify the modulations due to volatility, we computed different quantities for behavioral and neural estimates. Specifically, we defined a relative neural modulation index using estimated standardized regression coefficients as below:

$$\text{Rel. neural mod.} = sign(\beta_{i(mvol)} + \beta_{i(lvol)}) * (\beta_{i(mvol)} - \beta_{i(lvol)}) \qquad \text{(Eq. 12)}$$

where $i = \{1, \ldots 13\}$ is the regressor index, and $\beta_{i(lvol)}$, and $\beta_{i(mvol)}$ are the estimated regression coefficient for the less and more volatile environments, respectively. Similarly, we defined a relative behavioral modulation index using the behavioral estimate of the ratio of weights for reward probability and magnitude as below:

$$\text{Rel. behavioral mod.} = sign(\frac{\beta_{m(mvol)}}{\beta_{p(mvol)}} + \frac{\beta_{m(lvol)}}{\beta_{p(lvol)}}) * (\frac{\beta_{m(mvol)}}{\beta_{p(mvol)}} - \frac{\beta_{m(lvol)}}{\beta_{p(lvol)}}) \qquad \text{(Eq. 13)}$$

where $\frac{\beta_{m(lvol)}}{\beta_{p(lvol)}}$ and $\frac{\beta_{m(mvol)}}{\beta_{p(mvol)}}$ are the ratio of estimated weights for reward magnitude and probability (magnitude-to-probability weighting) in the less and more volatile environments, respectively, based on the fit of behavioral data using the simplest hybrid model.

Data analysis.

Data collection and analysis were not performed blind to the conditions of the experiments. Unless otherwise mentioned, data distribution was assumed to be non-normal but this was not formally tested. The statistical comparisons were performed using Wilcoxon signed-rank test in order to test the hypothesis of zero median for one sample or the difference between paired samples. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous similar publications [7, 16]. We used an alpha level of .05 for all statistical tests. The reported effect sizes are Cohen's d values. All behavioral analyses, model fitting, and simulations were done using MATLAB 2018a (MathWorks Inc., Natick, MA).

Code availability

Custom computer codes that support the findings of this study are available from the corresponding author upon request.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Emily Chu, Sada Nichols-Worley, and Lauren Tran for collecting human data, and Caleb Strait and Marc Mancarella for collecting monkey data in the gambling task. This work is supported by the National Science Foundation (CAREER Award BCS1253576 to B.Y.H. and EPSCoR Award 1632738 to A.S.), and the National Institutes of Health (grant R01 DA038615 to B.Y.H., grants R01 DA029330 and R01 MH108629 to D.L., and grant R01 DA047870 to A.S.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Bernoulli D Expositions of a new theory of the measurement of risk. Econometrica 22, 23–36 (1954).
2. Edwards W The theory of decision making. Psychol. Bull 51, 380 (1954). [PubMed: 13177802]
3. Kahneman D & Tversky A On the Psychology of Prediction. Psych Rev 80, 237–251 (1973).
4. Stewart N Information integration in risky choice: Identification and stability. Front. Psychol 2, 301 (2011). [PubMed: 22110455]
5. Ernst MO & Banks MS Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 429 (2002). [PubMed: 11807554]
6. Hunt LT, Dolan RJ & Behrens TE Hierarchical competitions subserving multi-attribute choice. Nat. Neurosci 17, 1613–1622 (2014). [PubMed: 25306549]
7. Farashahi S, Rowe K, Aslami Z, Lee D & Soltani A Feature-based learning improves adaptability without compromising precision. Nat. Commun 8, 1768 (2017). [PubMed: 29170381]
8. Farashahi S et al. Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. Neuron 94, 401–414 (2017). [PubMed: 28426971]
9. Spitmaan M, Chu E & Soltani A Salience-driven value construction for adaptive choice under risk. J. Neurosci 39, 5195–5209 (2019). [PubMed: 31023835]
10. Strait CE, Blanchard TC & Hayden BY Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. Neuron 82, 1357–1366 (2014). [PubMed: 24881835]
11. Farashahi S, Azab H, Hayden B & Soltani A On the flexibility of basic risk attitudes in monkeys. J. Neurosci 38, 2260–17 (2018).
12. Hayden B, Heilbronner S & Platt M Ambiguity aversion in rhesus macaques. Front. Neurosci 4, 166 (2010). [PubMed: 20922060]
13. Donahue CH & Lee D Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. Nat. Neurosci 18, 295–301 (2015). [PubMed: 25581364]
14. Massi B, Donahue CH & Lee D Volatility Facilitates Value Updating in the Prefrontal Cortex. Neuron 99, 598–608 (2018). [PubMed: 30033151]
15. Stephan KE, Penny WD, Daunizeau J, Moran RJ & Friston KJ Bayesian model selection for group studies (vol 46, pg 1005, 2009). NeuroImage 48, 311–311 (2009).
16. Behrens TEJ, Woolrich MW, Walton ME & Rushworth MFS Learning the value of information in an uncertain world. Nat. Neurosci. 10, 1214–1221 (2007). [PubMed: 17676057]
17. Tversky A Intransitivity of preferences. Psychol. Rev 76, 31 (1969).
18. Lichtenstein S & Slovic P The construction of preference. (Cambridge, UK: Cambridge University Press, 2006).
19. Ariely D, Loewenstein G & Prelec D “Coherent arbitrariness”: Stable demand curves without stable preferences. Q. J. Econ 118, 73–106 (2003).
20. Frederick S, Loewenstein G & O’donoghue T Time discounting and time preference: A critical review. J. Econ. Lit 40, 351–401 (2002).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

21. Kolling N, Wittmann M & Rushworth MF Multiple neural mechanisms of decision making and their competition under changing risk pressure. Neuron 81, 1190–1202 (2014). [PubMed: 24607236]

22. Ferrari-Toniolo S, Bujold PM & Schultz W Probability distortion depends on choice sequence in rhesus monkeys. J. Neurosci 39, 2915–2929 (2019). [PubMed: 30705103]

23. Hayden BY Time discounting and time preference in animals: a critical review. Psychon. Bull. Rev. 23, 39–53 (2016). [PubMed: 26063653]

24. Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ & Rushworth MFS Optimal decision making and the anterior cingulate cortex. Nat. Neurosci 9, 940–7 (2006). [PubMed: 16783368]

25. Soltani A & Izquierdo A Adaptive learning under expected and unexpected uncertainty. Nat. Rev. Neurosci In Press (2019).

26. Brainard DH The psychophysics toolbox. Spat. Vis 10, 433–436 (1997). [PubMed: 9176952]

27. Cornelissen FW, Peters EM & Palmer J The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. Behav. Res. Methods Instrum. Comput 34, 613–617 (2002). [PubMed: 12564564]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

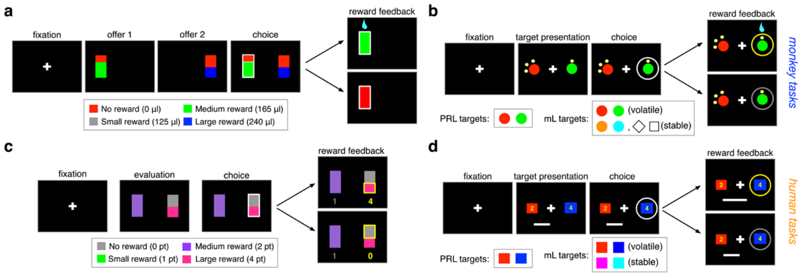

Figure 1. Experimental paradigms.

(**a, c**) Timeline of the gambling task in monkeys (a) and humans (c). On each trial, participants were presented with two options, each offering a gamble. Gambles were represented by a rectangle consisting of one or two portions with colors indicating different amounts of reward as indicated in the inset. The area of the colored portion indicates the probability that choosing that offer would yield the corresponding reward. Reward feedback was provided at the end of each trial indicating the gamble outcome followed by reward juice or points in monkeys and humans, respectively. (**b, d**) Timeline and reward schedules of the mixed learning and probabilistic reversal learning tasks in monkeys (b) and humans (d). On each trial, participants selected between two targets (colored circles or squares as shown in the insets) and subsequently received reward feedback (reward or no reward) on the chosen target. The reward was assigned probabilistically to one of the two targets, whereas the target with a larger probability of reward changed after a certain number of trials in volatile environments of the mL and PRL tasks. Reward magnitudes expected from each target were signaled by the number of dots around each target for monkeys and displayed numbers for human participants.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

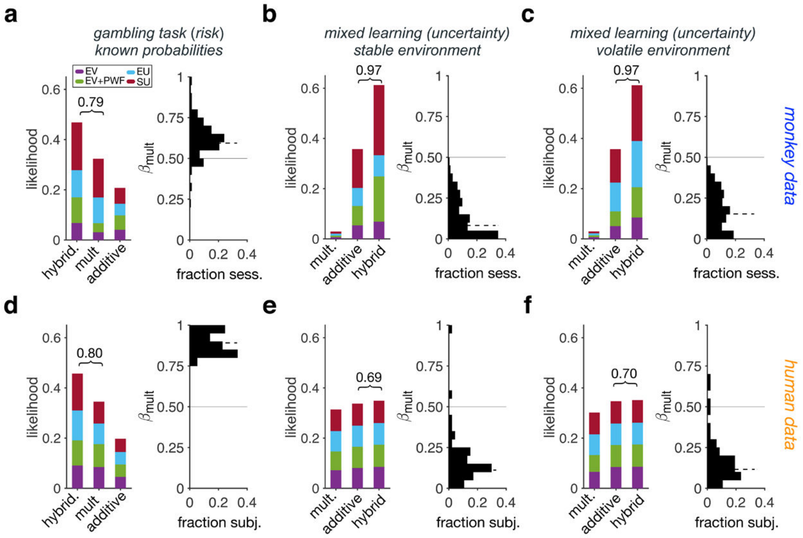

Figure 2. Different strategies for combination of reward information under risk and uncertainty.

(**a**) Left panel: Likelihood of different strategies adopted by monkeys during the gambling task (choice under risk) using the Bayesian model selection ($N = 146$). Different colors indicate different models: expected value (EV), EV with probability weighting (EV+PW), expected utility (EU), and subjective utility (SU), used for the estimation of subjective value. The values above the bracket shows the sum likelihood of the more prevalent strategy (additive or multiplicative) and the hybrid models with larger weighting of that strategy. Right panel: Distribution of the estimated values of β_{mult} using the hybrid model. The solid and dashed lines show 0.5 and median, respectively. (**b-c**) Same as in panel a but for monkeys in the stable (b) and volatile (c) environments of the mL task ($N = 316$). (**d-f**) The same as in panels **a-c** but for human participants (gambling: $N = 64$, mixed learning: $N = 46$). Under risk, multiplicative models can explain choice behavior better for both monkeys and humans, whereas additive models provide better fits to choice under uncertainty.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

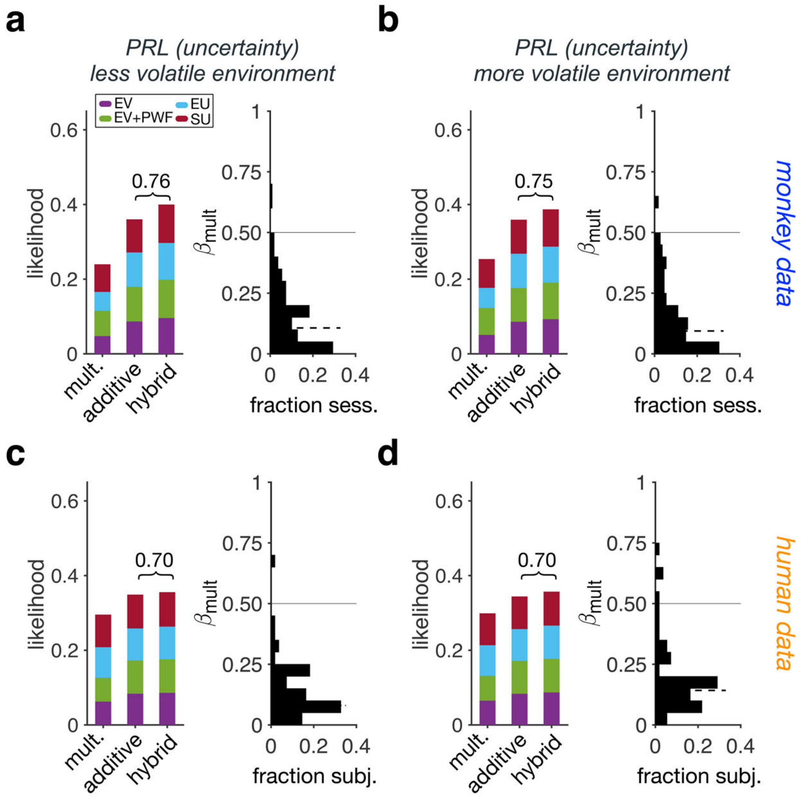

Figure 3. Additive models explain choice under uncertainty.

(**a-b**) Left panel: Likelihood of different strategies adopted by monkeys during the more (a) and less volatile (b) environments of the PRL task (choice under uncertainty) using the Bayesian model selection ($N = 118$). Different colors indicate different models: expected value (EV), EV with probability weighting (EV+PW), expected utility (EU), and subjective utility (SU), used for the estimation of subjective value. The values above the bracket shows the sum likelihood of the more prevalent strategy (additive or multiplicative) and the hybrid models with larger weighting of that strategy. Right panel: Distribution of the estimated values of β_{mult} using the hybrid model. The solid and dashed lines show 0.5 and median, respectively. (**c-d**) The same as in panels **a-b** but for human participants ($N = 38$).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

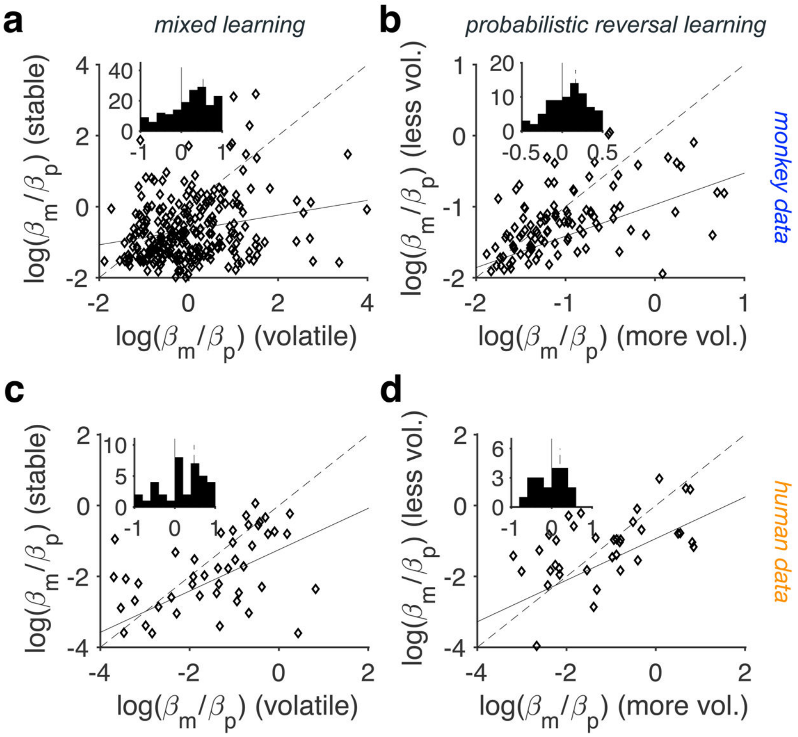

Figure 4. Adjustment of choice behavior to volatility of the environment.

(**a-b**) Plotted is the log ratio of estimated relative weights for reward magnitude and probability (magnitude-to-probability weighting) in the stable vs. volatile environment of the mL task (a), and the less vs. more volatile environment of the PRL task (b) in monkeys (mixed learning: $N = 316$, PRL: $N = 118$). The insets show the histogram of the difference in the log ratio of magnitude-to-probability weighting between the volatile and stable environments of the mL task (a), and the more and less volatile environments of the PRL task (b) (mixed learning: $N = 46$, PRL: $N = 38$). (**c-d**) Same as in panels **a-b** but for human participants during the mL (c) and PRL tasks (d).

Author Manuscript

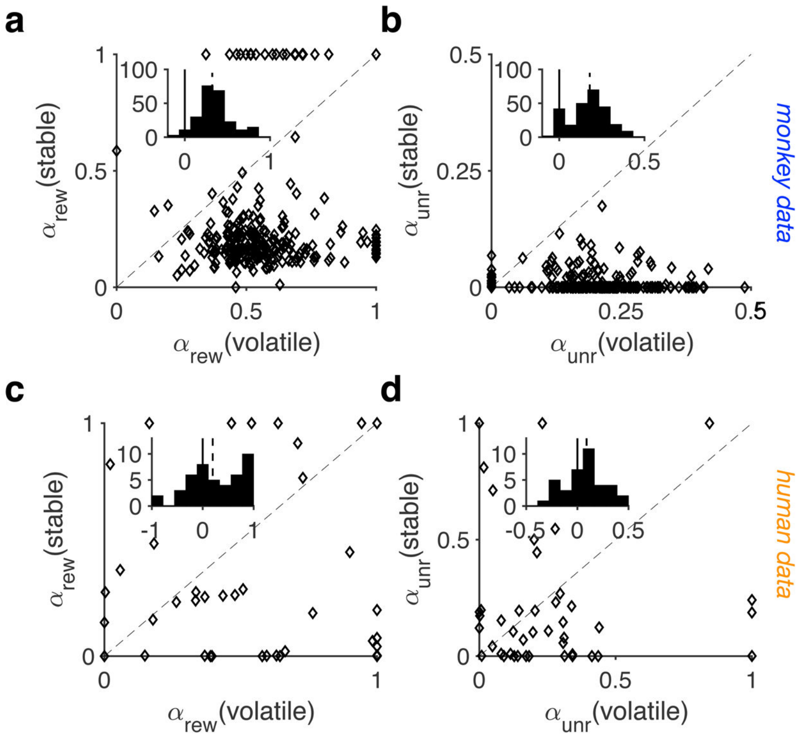

Figure 5. Behavioral adjustments in response to changes in volatility of the environment in the mL task.

(**a-b**) Plotted are the estimated learning rates (for the rewarded (α_{rew}) and unrewarded (α_{unr}) trials) in the stable vs. volatile environments of the mL task in monkeys. Insets show the histograms of the difference in the estimated learning rates between the stable and volatile environments. The solid and dashed lines show 0 and median, respectively. These results are based on session-by-session fit of the data to a simple additive model. There was a significant change in the both learning rates between the two environments (two-sided Wilcoxon signed-rank test; α_{rew} median±IQR: 0.26±0.26, $P < 0.001$, $d = 1.22$, $N = 316$, 95% CI = [0.23 0.30]; α_{unr} median±IQR: 0.18±0.13, $P < 0.001$, $d = 0.83$, $N = 316$, 95% CI = [0.16 0.20]). (**c-d**) The same as in panels **a-b** but for human data (two-sided Wilcoxon signed-rank test; α_{rew} median±IQR: 0.17±0.7, $P = 0.02$, $d = 0.51$, $N = 46$, 95% CI = [0.11 0.42]; α_{unr} median±IQR: 0.09±0.33, $P = 0.04$, $d = 0.16$, $N = 46$, 95% CI = [−0.05 0.18]).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

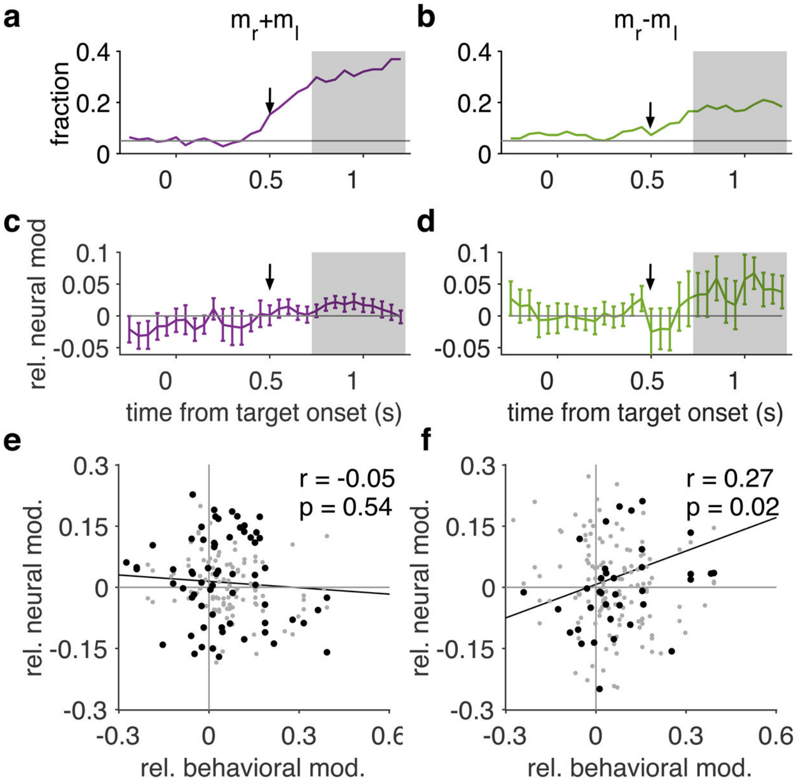

Figure 6. Neural signature of behavioral adjustments to volatility in the dlPFC.

(**a-b**) Plotted is the percentage of neurons that significantly encode the sum (a) and the difference (b) in reward magnitudes of the two options presented on each trial (*N* = 118). The arrows show the time at which magnitude cues were presented on the screen. (**c-d**) Plotted is the median of the relative neural modulation related to volatility for neurons that significantly encode the sum (c) and the difference (d) in reward magnitudes of the two options. Error bars show s.e.m. Gray background shows the period between 0.75 s and 1.25 s after target onset. These results were obtained with a sliding window of 500 ms. (**e-f**) Plotted is the change in encoding of the sum (e) and the difference (f) in reward magnitude (relative neural modulation due to volatility) in the dlPFC neurons vs. relative behavioral modulation. Black dots indicate magnitude-sum (e) and magnitude-difference (f) selective neurons.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

The relative weighting of the multiplicative strategy across different experiments.

Reported are median±IQR values of the estimated β_{mult} (relative weight assigned to the multiplicative component of the hybrid model), *p*-values, effect sizes (Cohen's d), and 95% confidence intervals for comparison of the estimated β_{mult} with 0.5 in different environments of the three tasks, and the size of dataset, separately for monkey and human data. The *p*-values are calculated using two-sided Wilcoxon signed-rank test. The multiplicative component was dominant under risk whereas the additive component was dominant under uncertainty.

	gambling task (risk)	**mixed learning task (uncertainty)**		**probabilistic reversal learning task (uncertainty)**	
	known probability	**stable**	**volatile**	**less volatile**	**more volatile**
monkey	0.59±0.12	0.08±0.20	0.15±0.18	0.11±0.19	0.09±0.20
	$P = 5.3\times10^{-18}$	$P = 1\times10^{-45}$	$P = 1\times10^{-45}$	$P = 1.7\times10^{-19}$	$P = 1.6\times10^{-19}$
	$d = 0.86$	$d = 3.37$	$d = 3.01$	$d = 2.72$	$d = 2.62$
	$N = 146$	$N = 316$	$N = 316$	$N = 118$	$N = 118$
	CI = [−0.42 −0.39]	CI = [0.38 0.41]	CI = [0.33 0.36]	CI = [0.34 0.39]	CI = [0.34 0.39]
human	0.89±0.12	0.08±0.08	0.09±0.11	0.06±0.10	0.11±0.10
	$P = 5.1\times10^{-11}$	$P = 9.8\times10^{-8}$	$P = 1.1\times10^{-6}$	$P = 3.8\times10^{-9}$	$P = 6.8\times10^{-9}$
	$d = 6.8$	$d = 2.78$	$d = 2.06$	$d = 3.43$	$d = 2.25$
	$N = 64$	$N = 46$	$N = 46$	$N = 38$	$N = 38$
	CI = [−0.11 −0.08]	CI = [0.34 0.40]	CI = [0.33 0.39]	CI = [0.37 0.42]	CI = [0.32 0.39]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Comparison of the relative weighting of the multiplicative strategy between the two levels of volatility of the mL and PRL tasks.

Reported are median±IQR values of the difference in estimated β_{mult} for the volatile and stable environments of the mL task and more volatile and less volatile environments of the PRL task, *p*-values, effect sizes (Cohen's d), and 95% confidence intervals for the tests of this comparison, and the size of dataset. The *p*-values are calculated using two-sided Wilcoxon signed-rank test. There was no consistent effect of volatility on the extent to which the additive vs. multiplicative strategy was adopted.

	mixed learning task (volatile vs. stable)	**probabilistic reversal learning task (more vol. vs. less vol.)**
monkey	0.03±0.184 $P = 4\times10^{-6}$ $d = 0.30$ $N = 316$ CI = [0.02 0.06]	−0.01±0.18 $P = 0.9$ $d = 0.01$ $N = 118$ CI = [−0.03 0.04]
human	0.01±0.21 $P = 0.8$ $d = 0.07$ $N = 46$ CI = [−0.05 0.06]	0.05±0.17 $P = 0.006$ $d = 0.24$ $N = 38$ CI = [0.01 0.09]