NEUROSCIENCE

Dynamic prospect theory: Two core decision theories coexist in the gambling behavior of monkeys and humans

Agnieszka Tymula[1]*, Xueting Wang[2], Yuri Imaizumi[3], Takashi Kawai[4], Jun Kunimatsu[4,5,6], Masayuki Matsumoto[4,5,6], Hiroshi Yamada[4,5,6]*

Research in the multidisciplinary field of neuroeconomics has mainly been driven by two influential theories regarding human economic choice: prospect theory, which describes decision-making under risk, and reinforcement learning theory, which describes learning for decision-making. We hypothesized that these two distinct theories guide decision-making in a comprehensive manner. Here, we propose and test a decision-making theory under uncertainty that combines these highly influential theories. Collecting many gambling decisions from laboratory monkeys allowed for reliable testing of our model and revealed a systematic violation of prospect theory's assumption that probability weighting is static. Using the same experimental paradigm in humans, substantial similarities between these species were uncovered by various econometric analyses of our dynamic prospect theory model, which incorporates decision-by-decision learning dynamics of prediction errors into static prospect theory. Our model provides a unified theoretical framework for exploring a neurobiological model of economic choice in human and nonhuman primates.

INTRODUCTION

Neuroeconomics (*1*, *2*) has made notable progress toward understanding how the brain makes economic choices. Two mathematical models, prospect theory (*3*) and reinforcement learning theory (*4*), have been frequently used to construct a biologically viable framework explaining economic choice. These two models were in principle developed to describe decision-making behavior in different situations. Prospect theory focuses on risk preferences, whereas reinforcement learning theory focuses on learning from outcomes of previous decisions. Both models describe similar neural activity related to risky decision-making with and without learning; however, it remains unclear how these two theories come together to explain economic choices [see Shen *et al.* (*5*) for an exception].

Prospect theory (*3*) was conceived to capture risky decision-making through static preference parameters. In the present study, we focused on two elements of prospect theory: the curvature of a utility function that captures the desirability of rewards and an inverse S-shaped probability weighting function that captures the underweighting of small and overweighting of large probabilities (our study focuses on the gain domain, and thus, loss aversion is beyond its scope). By multiplying the utilities of rewards by the weighted probabilities with which these rewards are received, decision makers calculate the subjective expected utilities for risky decisions. In humans, utility measured in laboratory experiments is typically concave in the domain of gains, whereas probability weighting is typically inverse S-shaped at aggregate levels (*3*, *6*, *7*) [see previous studies (*8*–*11*) for potential alternate views]. In laboratory monkeys—a close relative to humans—concave or convex utility functions and various shapes of probability weighting functions have been found (*12*–*16*), leading to inconsistent and sometimes controversial results. Inverse S-shaped probability weighting is not a general feature. For example, when choosers learn probabilities through the experience of rewards, the pattern of probability weighting is reversed in both humans (*17*) and monkeys (*15*). Similarly, in humans, feedback about the chosen lottery outcome reduces the willingness to take gambles with a small probability of winning (*18*). Overall, the existing literature suggests that the probability weighting assumed in prospect theory might be heterogeneous among individuals and that an individual's probability weighting might be adjusted on the basis of probabilistic events experienced in an environment during gambling. However, it remains unclear whether the experience of gambling outcomes causes systematic changes in the prospect theory's utility or probability weighting functions.

Reinforcement learning theory (*4*), on the other hand, describes how decision makers learn the values of available options through the experience of rewards under the assumption that they learn the values of all alternatives. These values are dynamically updated by calculating the reward prediction error (RPE) (*4*), which is the difference between the value of the obtained reward and its predicted value signaled in the brain (*19*, *20*). This mathematical algorithm yields learning dynamics for optimal choice behavior under uncertainty and has been extensively examined to describe decision-making behavior and related neural activity in humans, monkeys, and rodents (*21*–*23*). However, whether the RPE algorithm updates the utility of rewards or the probability weighting function with which the reward is received has not yet been explored. It is

[1]School of Economics, University of Sydney, Sydney, NSW 2006, Australia. [2]School of Economics, Finance and Marketing, College of Business and Law, RMIT University, Melbourne, VIC 2476, Australia. [3]Medical Sciences, University of Tsukuba, 1-1-1 Tenno-dai, Tsukuba, Ibaraki 305-8577, Japan. [4]Division of Biomedical Science, Institute of Medicine, University of Tsukuba, 1-1-1 Tenno-dai, Tsukuba, Ibaraki 305-8577, Japan. [5]Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tenno-dai, Tsukuba, Ibaraki 305-8577, Japan. [6]Transborder Medical Research Center, University of Tsukuba, 1-1-1 Tenno-dai, Tsukuba, Ibaraki 305-8577, Japan.
*Corresponding author. Email: h-yamada@md.tsukuba.ac.jp (H.Y.); agnieszka.tymula@sydney.edu.au (A.T.)

challenging to elucidate how these two influential decision theories are combined into a single model to explain risky decision making.

The present study combines the approaches of the static prospect theory and dynamic reinforcement learning theory into a combined model, including key features from both theories: utility function, probability weighting function, and RPE. In our experiment, participants (humans and monkeys) repeatedly made choices between lotteries that systematically varied in reward magnitudes, probabilities, and RPE levels. After each decision, the lottery was resolved, and participants received feedback that yielded RPE, allowing us to precisely test whether RPEs affect the parameters of the utility and probability weighting functions. In our procedure, gambles were constructed from a completely orthogonal payoff matrix of probabilities and magnitudes of rewards, and we collected decisions from laboratory monkeys over hundreds of trial blocks spanning several months. Using a closely matched task, we collected data from humans. These data allowed us to reliably estimate and compare the parameters of the combined model to examine its descriptive ability in monkeys and humans.

Previous studies have implied that findings from monkey research expand our understanding of human behavior by enabling researchers to understand the evolutionary roots of choice, as well as to conduct research that is not easily feasible with human participants. For example, we have previously described utility and probability weighting functions in monkeys using the same dataset in the present study (*24*). Leveraging the use of the same experimental paradigm for both species enabled us to compare monkey and human behavior through the lens of two major decision theories and to contribute to the long-standing dispute on the extent to which monkeys are a good model for human decision-making. To date, such comparisons have mostly been performed across studies using substantially different paradigms for each species [see (*25*) for an exception]. A fair comparison between monkeys and humans helps settle matters under debate in the literature (e.g., whether monkeys and humans have similar or different risk preferences), because our analyses and experimental paradigm were specifically designed for comparisons between these two closely related species. Using various analytic approaches to test our combined model, we consistently found that the probability weighting functions between species were dynamically adjusted decision-by-decision based on RPE information. Thus, we provide a novel and unified theoretical framework for the neuroeconomic study of decision-making under risk.

RESULTS

Relation to our previous studies

Our previous publications (*24*, *26*) reported estimates of the expected value model and static prospect theory models for risky choices of the monkeys using the same datasets in this study. We have never reported human experimental results or a dynamic prospect theory model. Thus, the contribution of the present study is the analysis of the effect of RPE on prospect theory parameters and interspecies comparison of risky choice behavior.

Experimental setting for monkeys

To construct and test a combined model of prospect theory and reinforcement learning theory, we trained monkeys to perform a lottery choice task (*26*) (Fig. 1A) similar to previous experiments performed with human participants in economics (*27*) and reinforcement learning (*28*). The task involved choosing between two options, each of which offered a liquid reward with a certain probability. The monkeys fixated on a central gray target, and two options were subsequently presented visually as pie charts displayed on the left and right sides of the screen. The number of green pie segments indicated the magnitude of the liquid reward in 0.1-ml increments (0.1 to 1.0 ml), and the number of blue pie segments indicated the probability of receiving the reward in 0.1 increments (0.1 to 1, where 1 indicates a 100% chance). After the pie charts disappeared, the gray target in the center reappeared for 0.5 s. Thereafter, the monkeys chose between the left and right targets by fixating on one side. Following the choice, the monkeys either received or did not receive the amount of liquid reward associated with the chosen option according to the option′s corresponding probability. In each choice trial, two of the 100 possible combinations of reward probability and magnitude (Fig. 1B) were randomly allocated to the left- and right-side target options. We used all data collected after each monkey learned to associate the probability and magnitude with the pie chart stimuli. The dataset includes 44,883 decisions made by monkey SUN (884 blocks over 242 days) and 19,292 decisions made by monkey FU (571 blocks over 127 days). We collected numerous decisions from each animal to simultaneously evaluate the free parameters of the two models. The well-trained monkeys showed behavior consistent with utility maximization, similar to a typical human participant, selecting, on average, options with a higher expected value (Fig. 1C).

Experimental setting for humans

Each experimental session began with written instructions explaining the task and the payment. The instructions explained that green and blue pie charts represent reward magnitude and probability, respectively. The same gamble display as in the monkey experiment was used. The reward magnitudes ranged from \$1 to \$5 (in \$1 increments), and the reward probabilities ranged from 0.1 to 1 (see Materials and Methods). We decided to use money rather than liquid rewards for human participants because it is known to be an attractive reward, and previous research suggests that risk attitudes are highly correlated across monetary, water, and food rewards (*29*, *30*). In contrast, liquid rewards are the most effective for homecaged monkeys under controlled water access (*31*, *32*). In each trial, participants decided between two gambles based on the reward probability and magnitude cued by the visual pie charts. To achieve as accurate a comparison as possible between humans and monkeys, we collected and analyzed 18,000 decisions from 72 human participants, with 250 decisions made by each participant. After each decision, the participants received feedback on whether the chosen lottery resulted in a win or not. To maintain incentive compatibility, after the participants made all decisions, they received payments for 20 outcomes of the randomly selected lotteries that they have chosen. We aimed to maintain as much consistency as possible between the two species′ experiments using an experimental design with the following features (see Materials and Methods for more details): (i) an orthogonal payoff matrix of probabilities and magnitudes of rewards involving various reward magnitudes, probabilities, and RPE levels; (ii) feedback on reward outcomes provided to the participants after every decision; and (iii) visual pie chart stimuli to convey information regarding reward probability and magnitude. These shared features between

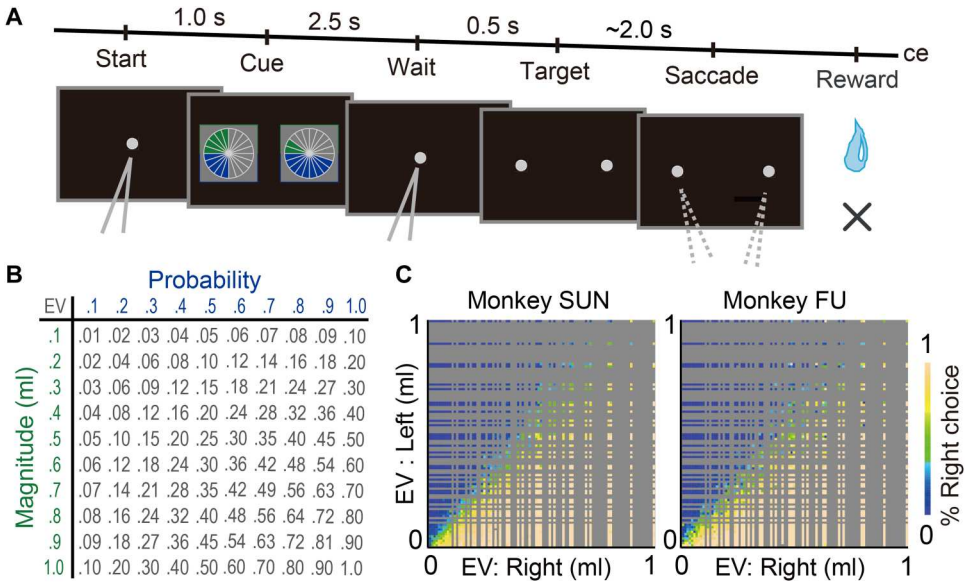

Fig. 1. Lottery choice task and monkey choice behavior. (**A**) A sequence of events in choice trials. Two pie charts representing the available options were presented to the monkeys on the left and right sides of the screen. Monkeys chose either target by fixating on the side where it appeared. (**B**) Payoff matrix. Each magnitude was fully crossed with each probability, resulting in a pool of 100 lotteries from which two were randomly allocated to the left- and right-side target options on each trial. Expected values (EVs) are calculated in milliliters. (**C**) The frequency with which the target on the right side was selected for the expected values of the left and right target options. (A) and (C) are published in (*26*).

the two experiments and the large number of trials at the aggregate level allowed for fair comparisons of the experimental results between the species.

We note that aggregated data analysis of human behavior is a standard procedure in the economic literature (*33*). In addition, the differences between the two experiments were as follows: (i) Monkeys decided on liquid rewards, whereas humans decided on monetary rewards; (ii) we collected numerous decisions from two monkeys and a relatively small number of individual decisions from many human participants (although still a large number of decisions for a human study); (iii) monkeys learned the magnitudes and probabilities associated with the visual pie chart lottery representation through experience before data collection, whereas human participants learned the meaning of the visual pie charts by reading the instructions; and (iv) reward delivery was not preceded by a tone for human participants.

Estimation of static prospect theory parameters

First, we estimated the utility and probability weighting functions using the standard parameterizations used in the literature. For the utility function, except for the expected value model, we used the power utility function $U(m) = m^{\alpha}$, where m indicates the magnitude of the reward, $\alpha > 1$ indicates convex utility (risk-seeking), $\alpha < 1$ indicates concave utility (risk-aversion), and $\alpha = 1$ indicates linear utility (risk neutrality). For the probability weighting function, we used one- and two-parameter functions that are commonly used in the literature. Overall, we estimated six models of the utility of receiving a reward magnitude m with probability p, $V(m, p)$ (see Materials and Methods).

To determine which model best described participants′ behavior, we used the Bayesian information criterion (BIC), which measures the goodness of fit, with a penalty (see Materials and Methods for more details). Among the six models, P2 and GE clearly had the lowest BIC and outperformed EV, EU, TK, and P1 (Fig. 2A, top and middle; see Materials and Methods for the model names), particularly in monkeys, with a reasonable range of estimated parameters for each model (Table 1). In humans, EV showed the worst performance, and EU largely improved the fit performance, although P2 and GE were the best and second-best models, respectively (Fig. 2A, bottom; see BIC values in Table 1). We also checked the degree of stochasticity (parameter β in Table 1) and psychometric choice functions (fig. S1) for our participants. Thus, the prospect theory model, which contains two free parameters in the probability weighting function, is the best for both monkeys and humans. We note that the fit performance was largely improved by the two-parameter probability weighting functions in monkeys, whereas adding the curvature to the utility function improved the fit performance in humans. We further examined these differences in the robustness check sections.

The estimates of the fitted parameters showed that the curvature of the utility function in the best-fit model was predominantly concave for both monkeys and humans (Fig. 2B; GE, green; P2, orange), but more concave for humans than for monkeys (see also the alpha values in P2 and GE in Table 1). Notably, the probability weighting functions diverged significantly from the traditionally assumed overweighting of small and underweighting of large probabilities in both species (Fig. 2C, green and orange, respectively). For example, the two-parameter probability weighting functions in monkeys were clearly concave (top and middle rows: GE, green; P2, orange). In humans, the probability weighting functions of P2 and GE were both predominantly concave. However, these two models yielded slightly different shapes (Fig. 2C, bottom). In particular, pronounced (P2, orange) or moderate (GE, green) overweighting was observed for almost all probability values, with

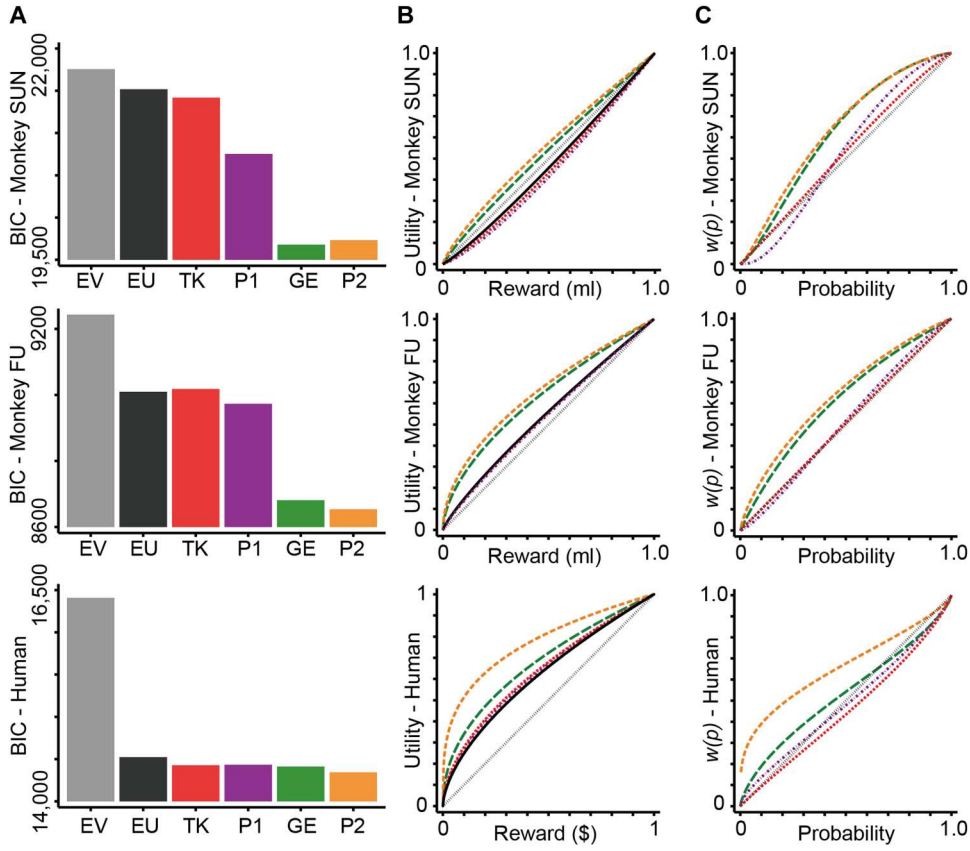

Fig. 2. Model estimates for monkeys and humans. (**A**) BIC values of the standard economic models: EV, EU, TK, P1, GE, and P2. See Materials and Methods for details; monkey SUN (top), monkey FU (middle), and human participants (bottom). The respective color for each model is presented. (**B**) Illustration of the estimated utility functions in all models. For illustrative purposes and to capture the same range of utility values on the *y* axis, the *x* axis is limited to 0 and 1 for humans to allow for an easier comparison with monkeys' utility functions. See fig. S4 for utility over 0 to 5 range. Gray, black, red, purple, green, and orange dashed lines indicate EV, EU, TK, P1, GE, and P2 models, respectively. (**C**) Illustration of the estimated probability weighting functions in all models.

minimal underweighting of very high probabilities in the GE model (please compare the delta and gamma values in Table 1 for interspecies differences). Thus, both monkeys and humans exhibited concave (or predominantly concave) utility functions and overweighted essentially all probabilities.

We further examined this result without parameterization using logistic regression analysis, where we compared the lottery choice frequency among different probability ranges: low ($P \leq 0.3$), middle ($0.4 \leq P \leq 0.6$), and high ($P \geq 0.7$). Both monkeys and humans chose lotteries with probabilities below 0.3 and above 0.7 less often than lotteries with middle-range probabilities, as indicated by the significant negative regression coefficients for $P \leq 0.3$ and $P \geq 0.7$ (Table 2). This choice pattern is consistent with the probability weighting functions shown in Fig. 2C. If the participants exhibited the traditionally assumed inverse S-shaped probability weighting function, they would have chosen lotteries with low probabilities more often and lotteries with high probabilities less often than lotteries with middle-range probabilities (after controlling for the linear impact of probability). The smaller effect of $P \leq 0.3$ in humans compared to monkeys was similar to that in the parametric analyses. Thus, consistent results were obtained in the parametric and nonparametric analyses of probability weightings.

Why was the utility function of monkey SUN estimated as convex in the EU, TK, and P1? The estimates from the two-parameter probability weighting models (Fig. 2, B and C; GE, green; P2, orange) were remarkably different from those from the one-parameter models, particularly in monkeys (Fig. 2, B and C; EU, black; TK, red; P1, purple). The EU, TK, and P1 models yielded more convex (monkey SUN) and less concave (monkey FU and humans) utility functions. One remarkable difference is that probability weighting is not assumed in EU. Moreover, by design, the TK and P1 probability weighting functions cannot achieve a concave shape; thus, because of this mathematical constraint, when using these models, the estimated probability weighting was either absent in monkey FU (Fig. 2C, middle, red and purple) or very slight in monkey SUN and humans (Fig. 2C, top and bottom, red and purple; Table 1 for the gamma parameters in TK and P1). Thus, as the EU, TK, and P1 models cannot capture the increased risk-taking stemming from concave probability weighting, they compensate by estimating more convex/less concave utility functions (i.e., more risk tolerant).

In summary, for the static prospect theory model, our orthogonal data matrix estimated using flexible models of probability weighting led to the conclusion that our participants, both monkeys and humans, distorted the probability in a similar manner, but differently from what is usually assumed in the

Table 1. Fit comparison of the six standard economic models. The estimated parameters of the six models are presented. SEs are shown in parentheses. For the human participants, SEs were clustered at the participant level. *P < 0.05, **P < 0.01, and ***P < 0.001 indicate significant differences from 0. The estimates with † were not significantly different from 1 at P = 0.05. Bold indicates the smallest BIC values.

	Monkey SUN					
	EV	**EU**	**TK**	**P1**	**GE**	**P2**
Alpha		1.1314*** (0.0088)	1.2047*** (0.0121)	1.2770*** (0.0113)	0.8787*** (0.0156)	0.7989*** (0.0173)
Delta					2.4187*** (0.0549)	0.5655*** (0.0120)
Gamma			1.1077*** (0.0111)	1.4789*** (0.0224)	1.3137*** (0.0220)	1.4299*** (0.0187)
Beta	0.0590*** (0.0006)	0.0561*** (0.0006)	0.0551*** (0.0007)	0.0572*** (0.0007)	0.0647*** (0.0008)	0.0647*** (0.0007)
N	44,883	44,883	44,883	44,883	44,883	44,883
BIC	21,754	21,516	21,417	20,752	**19,678**	19,730
	Monkey FU					
	EV	**EU**	**TK**	**P1**	**GE**	**P2**
Alpha		0.8323*** (0.0098)	0.8413*** (0.0123)	0.8618*** (0.0111)	0.5776*** (0.0186)	0.5208*** (0.0208)
Delta					1.9634*** (0.0754)	0.5665*** (0.0221)
Gamma			†1.0206*** (0.0165)	1.1428*** (0.0230)	†0.9818*** (0.0232)	1.1233*** (0.0212)
Beta	0.0590*** (0.0010)	0.0596*** (0.0010)	0.0597*** (0.0010)	0.0617*** (0.0011)	0.0636*** (0.0011)	0.0622*** (0.0012)
N	19,292	19,292	19,292	19,292	19,292	19,292
BIC	9,243	9,009	9,018	8,973	8,681	**8,653**
	Human					
	EV	**EU**	**TK**	**P1**	**GE**	**P2**
Alpha		0.6045*** (0.0394)	0.5557*** (0.0367)	0.5735*** (0.0375)	0.4678*** (0.0344)	0.2888*** (0.0405)
Delta					1.1847*** (0.1098)	0.5115*** (0.0708)
Gamma			0.8314*** (0.0266)	0.8264*** (0.0368)	0.7374*** (0.0323)	0.7298*** (0.0325)
Beta	0.5480*** (0.0426)	0.2784*** (0.0247)	0.2308*** (0.0232)	0.2358*** (0.0248)	0.1954*** (0.0168)	0.1191*** (0.0155)
N	18,000	18,000	18,000	18,000	18,000	18,000
BIC	16,407	14,522	14,425	14,432	14,411	**14,341**

theoretical literature. These distortions were best described as an overestimation of the probabilities. Furthermore, when subjective distortions in probability were accounted for, participants' estimated utility functions were more concave for both species. Thus, using the same experimental design and analysis, this is the first demonstration of a remarkable similarity between human and monkey utility and probability weighting functions estimated using standard economic models. Following the analysis of the static model, we examined whether the experience of gambling outcomes causes systematic changes in prospect theory's utility and/or probability weighting functions through the reinforcement learning model.

Reinforcement learning model simulation

We first performed simulations to illustrate how the typical reinforcement learning approach would model participant lottery valuation under our experimental conditions (see Materials and Methods). At the end of each trial t, the monkeys and humans received feedback about the lottery outcome (reward or no-reward), as instructed by a combination of the probability and magnitude of the chosen lottery. According to the reinforcement learning model, after receiving a reward or no-reward, the value function is updated through a learning algorithm $V(m, p)_{t+1} = V(m, p)_t + A\Delta_t$, where $\Delta_t = r_t - V(m, p)_t$ is the RPE [the difference between the received reward, r_t, and the predicted value of the lottery, $V(m, p)_t$]. A is

Table 2. Additional evidence for concave probability weighting functions. Results of logistic regression analysis with an indicator variable for whether the monkey selected target 1 (right) in a trial. m_i and p_i denote the reward magnitude and probability of receiving a reward at the target (1, right; 2, left), respectively. $p_1 \leq 0.3$ ($p_1 \geq 0.7$) is an indicator variable equal to one if p_1 is smaller than or equal to 0.3 (larger than or equal to 0.7). The SEs are shown in parentheses. For the human participants, SEs were clustered at the participant level. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$

	Monkey SUN	Monkey FU	Humans
$p_1 \leq 0.3$	−1.6556*** (0.0681)	−0.8861*** (0.1017)	−0.2931* (0.1172)
$p_1 \geq 0.7$	−1.0975*** (0.0666)	−0.6201*** (0.1016)	−0.5779*** (0.1187)
p_1	6.7635*** (0.1503)	7.5412*** (0.2343)	6.7081*** (0.6285)
m_1	7.6721*** (0.0865)	6.7615*** (0.1239)	0.8271*** (0.0650)
p_2	−6.2147*** (0.0776)	−7.3912*** (0.1297)	−4.7144*** (0.4701)
m_2	−7.0435*** (0.0826)	−6.6346*** (0.1231)	−1.3767*** (0.1141)
Constant	−0.0338 (0.0870)	0.3352* (0.1324)	−0.4423* (0.1979)
N	44,883	19,292	18,000

the learning rate at which the value function is updated (the learning rate A is changed from what is usually used in the literature to distinguish between the parameters of the reinforcement learning and prospect theory models). We simulated this reinforcement learning model using different learning rates for each of the 100 lotteries over 10,000 trials in a monkey experimental setting. After 10,000 trials, which was sufficient to fully learn the lottery value, the algorithm produced lottery valuations, $V(m, p)_{t=10,000}$, which converged to the expected value (Fig. 3A). There were slight deviations in the predictions from the expected value (diagonal line) due to the RPE, which yielded trial-by-trial dynamics in $V(m, p)_t$ even after substantial learning, and the extent of these fluctuations was causally related to the learning rate (Fig. 3A, comparison between panels). It was also clear that these deviations were observed in both the positive and negative components of the RPE (Fig. 3, B and C), as they were estimated from $r_t - V(m, p)_t$.

This simple simulation exercise demonstrated how the typical reinforcement learning model captured the trial-by-trial dynamics of gambling behavior in our experiments using $V(m, p)_t$. Reinforcement learning is limited in that it does not reveal whether utility curvature and/or probability weighting are influenced by the RPE. Thus, under the reinforcement learning approach alone, it remains unclear whether, after receiving a reward that is larger than predicted, the reward itself becomes more or less valuable, or whether an individual's belief regarding the probability of receiving the reward changes.

Construction of the dynamic prospect theory model

We combined the elements of the reinforcement learning theory and prospect theory into a single framework (see Materials and Methods). In doing so, we make three assumptions. First, on the basis of BIC values, we used P2 as our baseline probability weighting model, as it was one of the two models with the best fit (and best in two of the three datasets—Table 1; see also the robustness check section for the other model fits). Second, we allow the RPE to affect the parameters of both the utility and probability weighting functions to investigate whether the RPE affects risk-taking behavior through the utility, probability weighting function, or both. Third, we assumed that RPE would be derived from the expected values and received rewards as $\Delta_t = r_t - p_t m_t$. This implies that, in our experimental setting, after the reward is received, a positive RPE always occurs (except when the reward is expected with 100% probability), whereas after the no-reward, a negative RPE always occurs.

We constructed a fully combined model in two ways. In our first "standard" RPE model, we assumed that each of the parameters of the static prospect theory model depends on the RPE experienced in the last trial Δ_{t-1}. In our second "separate" RPE model, we assumed that each of the parameters of the static prospect theory model depends on two separate RPEs experienced in the last trial, $t-1$: a positive RPE after receiving a reward ($\Delta\text{positive}_{t-1}$) and a negative RPE after receiving a no-reward ($\Delta\text{negative}_{t-1}$). This is one of the common assumptions in reinforcement learning models because positive and negative RPEs differently affect the activity of dopamine neurons (*19*, *20*), and some previous studies have allowed for positive and negative RPEs to differentially affect economic choices (*34*, *35*). In the separate RPE model 2, 10 free parameters were included, and in the standard RPE model 1, 7 free parameters were included (see Materials and Methods). If the estimation of the model reveals that only the constant parameters α_0, δ_0, and γ_0 are significant, the model collapses into a traditional static prospect theory model. Conversely, if either α_i or δ_i or γ_i, which depends on the RPEs, is significantly different from zero, the RPEs are thought to affect choice through that parameter, and valuation and/or probability weighting adjusts consistently with the reinforcement learning model.

Nonparametric test of the dynamic prospect theory model

Before parameterizing dynamic prospect theory models 1 and 2, we first examined whether the use of two separate RPEs was adequate. For this purpose, we examined whether positive and negative RPEs influence risk-taking behavior in monkeys and human participants in a simpler analysis that preceded the structural estimation of our models (see Materials and Methods). In this analysis, we first determined which lottery was riskier in a trial by calculating the coefficient of variation for each lottery as the ratio of the SD of the lottery to its mean (*36*). We then regressed whether the chooser selected a riskier option in a given trial on the variables contained in the dynamic prospect theory models with standard and separate RPE (Table 3). We found that separating RPE into positive and negative RPEs was important for monkey FU and human participants (model 2 in Table 3). Monkey FU and human participants were significantly more risk-taking when the positive RPE was larger, whereas they were significantly less risk-taking when the negative RPE was larger (less negative RPE). In other words, the larger the absolute value of the RPE, the greater the risk-taking. Monkey SUN, on the other hand, took significantly more risks with a larger overall RPE (model 1 in Table 3), while the separate effect of positive or negative RPEs was significant for positive RPE only (model 2 in Table 3).

The estimates of the standard and separate RPE models suggest that the overall effect of RPE on risk-taking in monkey FU and human participants was not significant (model 1 in Table 3) and

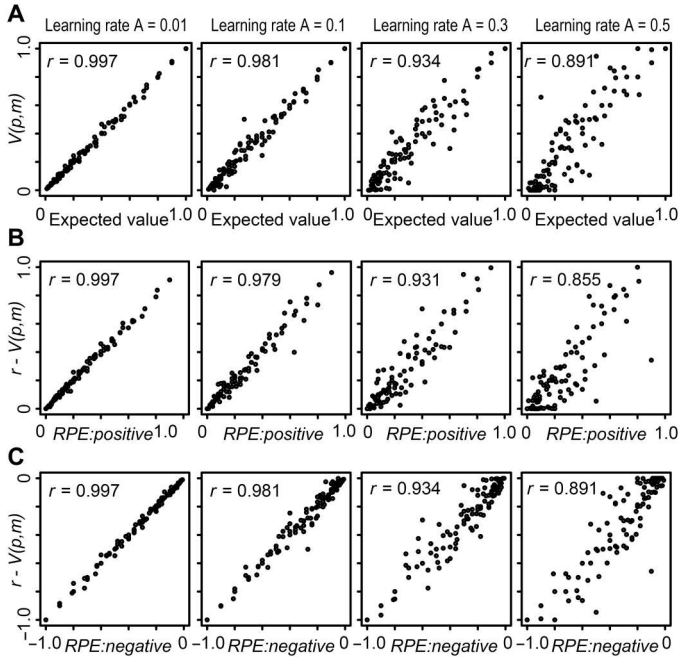

Fig. 3. Value function and its RPE estimated by using the reinforcement learning model. (**A**) Plots of the $V(p,m)_{t=10,000}$ against the expected value defined mathematically, i.e., probability time magnitude. (**B**) $r - V(p,m)_{t=10,000}$ after reward plotted against the positive component of RPE, i.e., obtained reward magnitude minus the expected values. (**C**) $r - V(p,m)_{t=10,000}$ after no-reward (hence, r is zero) plotted against the negative component of RPE, i.e., zero minus expected value. Plots were made for all stimuli, as a function of different learning rates. r is the correlation coefficient.

that the positive and negative RPEs have opposite effects (see coefficients for $\Delta\text{positive}_{t-1}$ and $\Delta\text{negative}_{t-1}$). In monkey SUN, the coefficients of both positive ($\Delta\text{positive}_{t-1}$) and negative RPE ($\Delta\text{negative}_{t-1}$) have the same positive sign (model 2), which may have contributed to the significant effect of the overall RPE. The significance of the positive RPEs was observed in all three datasets at $P < 0.05$ (model 2 in Table 3), suggesting that a positive RPE affects the internal subjective valuation of the lottery under risk. Because this simple analysis cannot determine whether the utility or probability weighting function is affected by RPEs, we next estimated our structural models to answer this question.

We note that although the following additive analyses were exploratory and not driven theoretically, we have separated the impact of the reward magnitude and probability conveyed by the RPE, as well as the effects of the chosen, received, and unchosen options (see table S1), but did not find consistent effects. Moreover, because positive and negative RPEs are moderately correlated (0.404 and 0.391 in monkey datasets and 0.315 in human dataset; table S2), we analyzed the effect of either positive or negative RPE (table S3). The results were not identical, but were almost consistent with those from the following parametric analyses.

Parametric test of the dynamic prospect theory model

We fit our dynamic prospect theory model to the data to parameterize the RPE effect through utility and probability weighting functions (Table 4). Our analysis can determine whether the RPE affects either one, both, or none of the functions. The separate RPE model 2 showed a dependence of the probability weighting function parameter delta on positive RPE in both monkeys and humans (Table 4; see delta parameters, $\Delta\text{positive}_{t-1}$ in model 2). The significant negative coefficients of delta, $\Delta\text{positive}_{t-1}$, observed in all three datasets indicate that the larger the positive RPE, the more concave the probability weighting function (Fig. 4B, compare solid orange and gray dotted curves). In contrast, no significant effect was observed in the gamma parameter (Table 4, see gamma).

In the standard RPE model 1, which does not distinguish between positive and negative RPEs, we did not find a consistent effect of RPE on the delta parameter between monkeys and humans (Table 4, see delta parameters, Δ_{t-1} in model 1). This is because the effect of negative RPE varied between the participants (Table 4, see delta parameters, $\Delta\text{negative}_{t-1}$ in model 2), and the

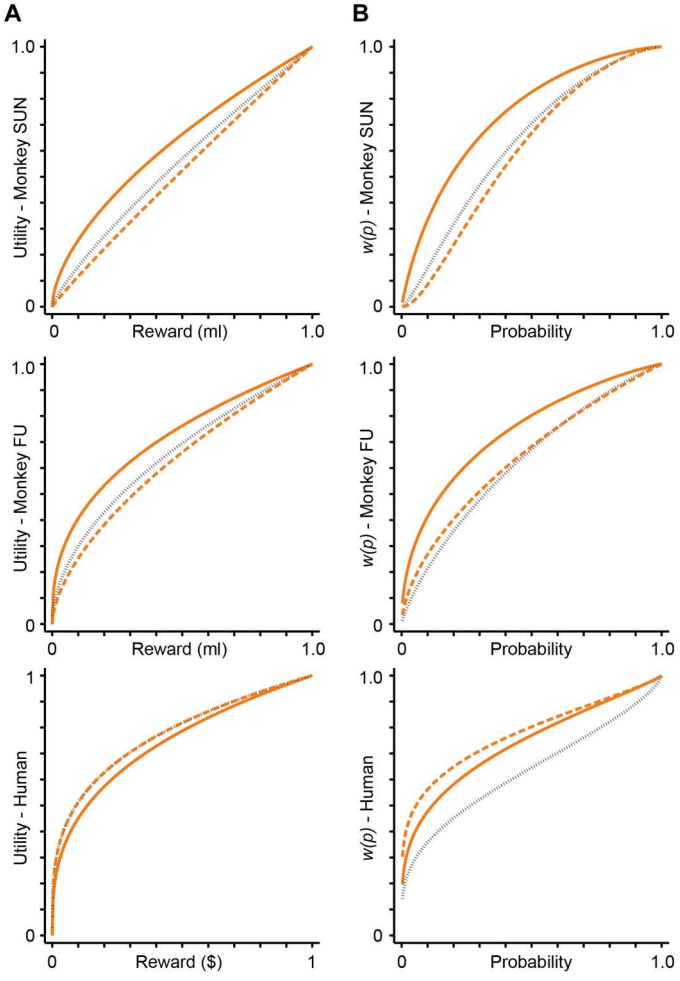

Fig. 4. The effect of the RPE through utility and probability weighting functions in the dynamic prospect theory model. Effect of RPE on the utility function (**A**) and probability weighting function (**B**) for monkey SUN (top), monkey FU (middle), and human participants (bottom). All curves were drawn using the estimates from Table 4 (model 2). Solid orange lines are drawn assuming the largest RPE (0.9 for monkeys and 4.5 for humans), dashed orange lines are drawn assuming the smallest RPE (−0.9 for monkeys and −4.5 for humans), and the dotted gray line represents RPE = 0. Note that in humans, the dotted gray and dashed orange curves overlap in (A). For illustrative purposes, we limited the *y* axis and *x* axis to 0 and 1 for humans to allow for an easier comparison with monkeys' utility functions.

Table 3. RPEs influence risk-taking behavior. Logistic regression with response variable = 1 if the riskier lottery is selected and 0 otherwise. Two alternative RPE models were used: (1) a standard RPE model and (2) a separate RPE model with separate positive and negative components. All the models included constant and lottery magnitudes and probabilities. The SEs are shown in parentheses. Fixed effects were included in the analysis of human data, and SEs were clustered at the participant level in human data. $^{+}P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

	Monkey SUN		Monkey FU		Humans	
	(1)	(2)	(1)	(2)	(1)	(2)
Δ_{t-1}	0.1312* (0.0582)		−0.1732^{+} (0.0925)		−0.0261 (0.0198)	
$\Delta\text{positive}_{t-1}$		0.2530* (0.1169)		0.3906* (0.1851)		0.1794*** (0.0411)
$\Delta\text{negative}_{t-1}$		0.0365 (0.0980)		−0.5978*** (0.1522)		−0.2140*** (0.0354)
m_{risky}	6.4970*** (0.0825)	6.4983*** (0.0825)	5.9308*** (0.1239)	5.9315*** (0.1240)	0.8447*** (0.0639)	0.8479*** (0.0641)
p_{risky}	8.2466*** (0.1131)	8.2471*** (0.1131)	8.6596*** (0.1811)	8.6746*** (0.1813)	6.8084*** (0.6092)	6.8542*** (0.6088)
m_{safe}	−9.0517*** (0.1005)	−9.0530*** (0.1005)	−8.4309*** (0.1494)	−8.4419*** (0.1496)	−1.4145*** (0.1158)	−1.4205*** (0.1156)
p_{safe}	−4.5701*** (0.0974)	−4.5719*** (0.0974)	−5.7667*** (0.1624)	−5.7760*** (0.1625)	−4.6067*** (0.4566)	−4.6258*** (0.4555)
Constant	−0.8470*** (0.0660)	−0.8695*** (0.0686)	−0.5486*** (0.1016)	−0.6488*** (0.1057)	−0.6622*** (0.1746)	−0.8343*** (0.1801)
N	43,999	43,999	18,721	18,721	17,928	17,928

overall RPE in model 1 reflects the average of these two effects. Thus, a positive RPE consistently explains changes in the probability weighting function. These findings suggest that monkeys and humans may share the common effect of a more optimistic perception of reward probability after an unexpected positive outcome, such as good fortune or winning a jackpot.

The utility function governed by the alpha parameter was not consistently affected by RPE, as shown by no significant effects of positive and negative RPEs in most cases (Table 4, see alpha in model 2; see Fig. 4A). In model 1, the curvature of the utility function was affected by the overall RPE in monkey SUN and human participants, but with the opposite effects (Table 4; see alpha in model 1). Thus, it is likely that the RPE does not consistently affect the utility curvature in the same way, in contrast to the consistent effect of a positive RPE on the probability weighting function parameter delta. Although separation of RPEs into the positive and negative component seems important for understanding how RPE influences choice, we note that the separate RPE model (Table 4, model 2) had a better fit according to BIC than the standard RPE model (Table 4, model 1) for humans but not for monkeys, likely due to the additional three free parameters.

Robustness checks

Because our separate RPE model contains a large number of free parameters (10 free parameters in total; see Materials and Methods), there is a possibility that the estimated parameters of the model may not be accurate, possibly because of the moderate correlation between the RPE variables in our dataset (table S2). We therefore tested the robustness of our findings in four ways. First, we tested whether the results of our combined model remained consistent when we used different probability weighting functions, especially GE, the other two-parameter probability weighting function. Second, we performed parameter recovery simulations (*37*) to examine the ability of our model to recover the assumed parameters of the separate RPE model. If the estimated parameters of the model were not the ones that were assumed, this would imply that the original parameters cannot be correctly recovered in this process. Third, we searched for a combination of parameters that best fit our data (i.e., model selection using BIC) to examine whether positive RPE affects the probability weighting function through the delta parameter in the best-fit model. Fourth, we performed cross-validation to avoid potential overfitting of the models to the data.

(i) Other probability weighting functions. As a first robustness check, we estimated the standard and separate RPE models using all four probability weighting functions: TK, P1, GE, and P2. For both dynamic prospect theory models with the standard and the separate RPE, P2 showed the lowest BIC for monkey FU and human participants, and GE showed the lowest BIC for monkey SUN (see table S4, models 1 and 2). Furthermore, the results were qualitatively the same when using P2 and GE probability weighting functions (fig. S2, B and D). In particular, models with either P2 or GE probability weighting specifications found that participants are more optimistic after larger positive RPE. The effect of RPEs on the utility function seemed to be similar between these two models (fig. S2, A and C), although consistent significance of these effects was not observed across participants (see model 2 for GE and P2 in table S4). As in the main analysis, we found that models with one-parameter probability weighting functions are not flexible enough to capture these effects (see TK and P1 models in table S4). Overall, the impact of RPEs on probability weighting was similar for GE and P2.

Table 4. Estimated parameters in the dynamic prospect theory model. Two alternative RPE models were used: (1) a standard RPE model and (2) an RPE model with separate positive and negative components. All models included a constant. For the human participants, SEs were clustered at the participant level. $^{+}P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

	Monkey SUN		Monkey FU		Humans	
	(1)	(2)	(1)	(2)	(1)	(2)
Alpha						
Δ_{t-1}	−0.1804** (0.0610)		−0.1136^{+} (0.0661)		0.0110* (0.0049)	
$\Delta\text{positive}_{t-1}$		−0.2430* (0.1213)		−0.1475 (0.1107)		0.0125 (0.0111)
$\Delta\text{negative}_{t-1}$		−0.1364 (0.1035)		−0.0817 (0.1078)		−0.0002 (0.0101)
Constant	0.7991*** (0.0173)	0.8101*** (0.0263)	0.5154*** (0.0210)	0.5240*** (0.0288)	0.2905*** (0.0302)	0.2896*** (0.0392)
Delta						
Δ_{t-1}	−0.2004*** (0.0430)		−0.1210 (0.0779)		0.0321** (0.0108)	
$\Delta\text{positive}_{t-1}$		−0.2882*** (0.0837)		−0.2936** (0.1122)		−0.0463* (0.0221)
$\Delta\text{negative}_{t-1}$		−0.1281^{+} (0.0735)		0.0495 (0.1111)		0.0606*** (0.0140)
Constant	0.5686*** (0.0120)	0.5858*** (0.0187)	0.5616*** (0.0223)	0.6005*** (0.0314)	0.5154*** (0.0505)	0.5674*** (0.0726)
Gamma						
Δ_{t-1}	−0.0451 (0.0641)		0.0960 (0.0697)		−0.0154 (0.0112)	
$\Delta\text{positive}_{t-1}$		0.0429 (0.1220)		0.0277 (0.1434)		0.0351 (0.0233)
$\Delta\text{negative}_{t-1}$		−0.1034 (0.1140)		0.1108 (0.1085)		−0.0212 (0.0147)
Constant	1.4311*** (0.0188)	1.4148*** (0.0279)	1.1234*** (0.0216)	1.1334*** (0.0316)	0.7270*** (0.0159)	0.7096*** (0.0356)
Beta						
Constant	0.0636*** (0.0008)	0.0636*** (0.0008)	0.0619*** (0.0012)	0.0619*** (0.0012)	0.1186*** (0.0119)	0.1200*** (0.0151)
N	43,999	43,999	18,721	18,721	17,928	17,928
Number of free parameters	7	10	7	10	7	10
BIC	19,158	19,187	8,418	8,430	14,286	14,239

(ii) Parameter and model recovery. The moderate correlation between explanatory variables in our dataset may affect the accuracy to estimate their impact on the variables in the dynamic prospect theory. The correlation coefficients between the positive and negative RPEs were 0.404, 0.391, and 0.315 in monkeys SUN, FU, and human participants, respectively (table S2). To examine whether our most complicated model, the separate RPE model (model 2 in Table 4), can recover all 10 parameters correctly, we created choice datasets using simulated parameters. Each simulation was characterized by a utility and probability weighting parameters that were randomly selected from a reasonable range of the values. Using these simulated parameters, we generated choices that a chooser characterized by such parameters would have made in our experimental task. Thereafter, we used the same model that had been used to generate the simulated choice data to estimate its parameters and examined whether the estimated model parameters were correlated with the original parameters used to create the simulation data. As a benchmark, we also simulated the choices for the EU and static prospect theory (P2) models.

This parameter recovery exercise showed high correlation coefficients between the sets of randomly generated parameters and the estimates obtained in the EU, P2, and separate RPE models (table S5). For static parameters, the correlations were above 0.98 for all models. For parameters that depended on the positive and negative RPEs, the correlations ranged from 0.76 to 0.94. Using the separate RPE model, we also confirmed that the assumed parameters fell within the estimated confidence intervals in more than 85% of the simulated datasets (table S6). Thus, we concluded that the separate RPE model can precisely recover the correct parameters, indicating that the estimated parameters are reliable.

A separate question is how often the data-generating model is recognized by model comparisons. To answer this question, we first use our simulated dataset to check how often EU or P2 chooser would be incorrectly identified as separate RPE chooser. We found that this never happens. When the generative choosers have EU preferences, BIC of the EU model fit is always lower than that of separate RPE model. Similarly, when the generative choosers have P2 preferences, P2 is always identified as a better model than separate RPE according to BIC. After checking these static prospect theory models, we checked how often the choosers with separate RPE preferences are correctly identified according to BIC. We found that of our 40 simulated choosers, none were ever mistaken for EU choosers. However, we would mistakenly classify eight as P2 according to BIC. For these simulated participants, the BIC differences between the models were minimal (fig. S3). Only one of them would be classified as P2 chooser according to Akaike′s Information Criterion (AIC) that applies lower penalty for extra parameters. Overall, we concluded that it is very unlikely that we would classify EU or P2 choosers as separate RPE choosers, but it is possible to sometimes mistake separate RPE choosers for P2 choosers due to the extra parameter penalty applied by BIC and to a lesser extent by AIC.

(iii) Model selection to search for the best model. We found that in the separate RPE model (model 2 in Table 4), some of the included explanatory variables were not significant. Hence, we attempted to find the best-fit dynamic prospect theory model with separate RPE and to determine whether in the best-fit model the parameter delta is still affected by the positive RPE. We adopted a model selection approach that finds combinations of parameters with the lowest BIC (see Materials and Methods; see also cross-validation for additional considerations).

In our best-fit model based on BIC (P2), we replicated the same significant effect of a positive RPE on delta in both monkeys and humans and found no effect on the gamma parameter (Table 5; the estimates in Tables 4 and 5 differ because different sets of parameters are included due to the model selection approach in Table 5.). In addition, the delta parameter in human participants was also affected by a negative RPE. In contrast, the utility curvature was only affected by the negative RPE in monkey FU. These results are consistent with the full model tested above (model 2 in Table 4). In addition, we searched for the best-fit model replacing P2 with the other probability weighting functions (model 3 in table S4). As with P2, the best-fit model with GE showed that probability weighting, rather than utility, was affected by the positive RPE. This close replication of the results using GE probability weighting function indicated that the experiences of positive RPE lead to a more pronounced overweighting of probabilities. Note that the model performances according to BICs were only slightly different across three dynamic prospect theory models with P2 and GE, indicating that the performances of our dynamic prospect theory models are almost equivalent.

Further examination showed that if we select the best-fit models using one-parameter probability weighting functions (model 3 with TK and P1 in table S4), utility was affected by RPEs and probability weighting was unaffected. The conclusions of models with a one-parameter probability weighting functions were different than with two-parameter probability weighting functions. However, these models also showed larger BIC values compared to the two-parameter models (see BIC values in table S4; GE in monkey SUN and P2 in monkey FU and human). These differences in the results were consistent with the differences in the static prospect theory model analysis (Fig. 2 and Table 1). We attributed them to the inability of the one-parameter probability weighting functions to capture the kind of probability distortions displayed by our participants.

(iv) Cross-validation. The differences in BIC were not particularly large between the models (Fig. 2A), and our dynamic prospect theory models have a large number of free parameters, which can lead to overfitting. To address potential overfitting problems, we conducted a four-fold cross-validation that assesses how the models perform out of sample. We compared EU, static prospect theory, dynamic prospect theory with standard RPE, dynamic prospect theory with separate RPE, and our best-fit model determined in the previous section. Because P2 probability weighting function mostly outperformed other probability weighting functions, we used it for these comparisons. For cross-validation, we used log likelihood of the model estimated for test samples (i.e., out-of-sample), instead of BICs (see Materials and Methods for this reason).

Consistent with our analysis in the preceding section, the log likelihoods were almost equal across the three dynamic prospect theory models (table S7). The best-fit model selected using the full dataset according to BIC and the dynamic prospect theory with separate RPE came as the best or the second best models (although the log likelihood values were very similar). Across all datasets, the dynamic prospect theory models have lower log likelihoods than static prospect theory and EU. Thus, cross-validation suggested that a dynamic prospect theory model would better describe gambling behavior in human and nonhuman primates.

Result summary

Interpreting these results together, we conclude that the probability weighting function is affected by RPE, especially the positive RPE as observed in both human and monkey participants. Our conclusion was based on the finding that models that incorporate the positive RPE into the prospect theory probability weighting function consistently showed a significant effect in both parametric and nonparametric analyses (Tables 3 to 5 and tables S3 and S4) and a slightly better performance compared to the static prospect theory under cross-validation. Further, we have conducted a range of robustness checks and demonstrated that despite the large number of free parameters, they can be correctly recovered using our econometric procedures (table S5). The effect of positive RPE on probability weighting survives model selection procedure (Table 5 and table S4), and the results are consistent under both frequently used two-parameter probability weighting functions (table S4). On the other hand, the effect of the negative RPE and the effect on utility function were not consistently established across monkeys and across species.

After these robustness checks, we found the effect of the positive RPE particularly remarkable and perhaps unexpected as our monkeys underwent substantial and prolonged training and knew the choice environment well, and the stimuli were intuitive and clearly explained to the human participants. Our results suggest that the RPE signal is also important in choice situations in which participants do not need to learn the reward magnitudes and probabilities. The existence of such dynamic changes in lottery valuations may help explain the decision-by-decision dynamics in risky choices.

Table 5. Estimated parameters in the selected combined model based on model selection. Statistics are presented in the best selected model based on P2, which showed the lowest BIC.

	Monkey SUN				Monkey FU				Humans			
Partial model	**Coefficient**	**SE**	***z***	***P***	**Coefficient**	**SE**	***z***	***P***	**Coefficient**	**SE**	***z***	***P***
Alpha												
$\Delta positive_{t-1}$	–	–	–	–	–	–	–	–	–	–	–	–
$\Delta negative_{t-1}$	–	–	–	–	−0.160	0.039	−4.08	<0.001	–	–	–	–
Constant	0.798	0.018	45.73	<0.001	0.501	0.021	24.23	<0.001	0.307	0.039	7.80	<0.001
Delta												
$\Delta positive_{t-1}$	−0.140	0.029	−4.79	<0.001	−0.145	0.045	−3.24	0.001	−0.047	0.013	−3.67	<0.001
$\Delta negative_{t-1}$	–	–	–	–	–	–	–	–	0.055	0.011	5.13	<0.001
Constant	0.582	0.013	45.23	<0.001	0.579	0.024	24.64	<0.001	0.585	0.077	7.64	<0.001
Gamma												
$\Delta positive_{t-1}$	–	–	–	–	–	–	–	–	–	–	–	–
$\Delta negative_{t-1}$	–	–	–	–	–	–	–	–	–	–	–	–
Constant	1.429	0.019	76.23	<0.001	1.124	0.022	52.07	<0.001	0.736	0.032	22.84	<0.001
Beta	0.064	0.001	85.34	<0.001	0.062	0.001	51.79	<0.001	0.125	0.015	8.02	<0.001
N	43,999				18,721				17,928			
BIC	19,146				8,394				14,208			

DISCUSSION

Reinforcement learning and prospect theory

We adopted a novel framework to test whether RPE influences the parameters of the utility and probability weighting functions in a stable reward environment in which the associations between stimuli and rewards are known. As learning usually occurs even in stable environments with lower learning rates, a priori, we hypothesized that the reinforcement learning algorithm drives a decision maker′s valuation of gambles through changes in, either or both, the utility and probability weighting functions. Our empirical testing indicated that the probability weighting function dynamically adjusts on a trial-by-trial basis based on outcome experiences. Despite extensive training of monkeys in the visual lottery, they remained alert to the RPE and updated their subjective valuations of the gambles through experience. Consistently, human participants were sensitive to RPE, although they made decisions based on known probabilities and magnitudes of rewards, presented using visual pie charts. After a positive RPE, both monkeys and humans significantly overweighted all probabilities (Fig. 4B, solid orange line) more than after no RPE (Fig. 4B, gray line), suggesting that both species share a common valuation: greater optimism regarding the probability of winning after experiencing a larger positive RPE. After a positive RPE, dopamine is released from the synaptic terminal (*19*, *20*) to guide learning and form predictions (*24*) for the next decision. This mechanism may drive animals to maximize utility in a broader decision-making context when the environment is less stable by learning predictions in the target brain regions (*38*–*40*).

A substantial number of studies in behavioral and experimental economics have found that people′s willingness to take risks changes after unexpected changes in the environment, such as after natural disasters or war [see introduction in (*41*) for a summary of recent evidence] or even after inconsequential exposure to reward distributions (*42*). Wealth effects, framing, and background risk have been suggested as potential factors that explain these effects. In line with this view, our findings suggest that sensitivity to RPEs, an integral part of the valuation and prediction system, could at least partially explain some of the findings in the literature on decision-making in unstable environments, because learning never stops for probabilistic outcomes even in a stable environment. Even in our task, where the reward outcomes are not informative about future rewards, as soon as feedback occurs, animals and humans use the RPE to guide their next decision.

Probability weighting across species

One of the main challenges in studying probability distortions in humans is the limited number of choices that an experimenter can feasibly ask a human participant to make. This undermines the ability to obtain reliable estimates of the probability weighting and utility curvature at the individual level. To dissociate these two functions during parameter estimation, experimental tasks often include choices with large-reward low-probability lotteries (*6*), such as a 1% chance of winning USD 500. Rewards of this scale, which may correspond to more than 1000 ml of water in monkeys, are never included in monkey experiments. Therefore, it remains unclear how monkeys would behave when faced with low probabilities of very large gains. This sampling procedure may affect the estimated probability distortions because the lower and upper limits largely affect parameter estimation. However, such an experimental design in human studies introduces a major limitation: The negative correlation between reward magnitudes and probabilities can affect estimates, which poses a challenge for reliable parameter estimation. To overcome this obstacle, we used a dataset that includes a large number of choices and a payoff matrix that fully orthogonalizes reward magnitudes and

probabilities. These features allowed us to reliably estimate and compare many free parameters of the dynamic prospect theory model in monkeys and humans at the aggregate level.

In the economic literature, most studies conducted with human participants have found inverse S-shaped probability weighting functions at the aggregate level, with a large amount of heterogeneity at the individual level (*7*–*11*, *43*, *44*). The estimated probability weighting functions were also found to vary with changes in mood (*11*), age, and sex (*9*). A handful of recent studies have investigated static distortions in probability weighting of captive macaques, with inconsistent results (*13*–*16*, *45*, *46*). The probability weighting function was found to be inverse S-shaped (*14*, *15*), S-shaped (*13*, *16*), and concave (*15*) when estimated on the basis of economic choice tasks and either concave or convex when estimated based on nonchoice tasks (*47*). Most of these existing studies have not provided a mechanistic explanation for why heterogeneities across species may exist, and many reasons, such as those related to methodology, could explain the inconsistencies in findings across the animal and human literature. To reconcile these discrepancies, we conducted comparable experiments with monkeys and human participants using the same experimental design principles and identical data analysis procedures. We found that probability weighting was similar for the two species, generally overweighting all levels of probability. This is the first demonstration of a remarkable similarity between human and monkey probability weighting functions estimated using standard economic models.

Our analysis clearly shows that functional assumptions in the empirical analysis of probability weighting have crucial implications. We found that the different functional forms yielded remarkably different probability weighting functions (Fig. 2C). For example, we estimated both of the more flexible two-parameter functions to be consistently predominantly concave, whereas the one-parameter probability weighting functions (TK and P1) were essentially linear, suggesting no or minimal probability distortions. Thus, the selection of the appropriate functional form is critical and affects the results, particularly if the probability distortions are not in the traditionally assumed S shape [e.g., see simulations of the utility shapes that can be captured by different functions in (*48*)]. Our findings should caution researchers to either use the most flexible probability weighting functions or estimate a range of the functions suggested in the literature.

Last, we showed that the assumption in prospect theory that the probability weighting function is static would not be sufficient to accurately describe the behavior of macaques and humans (Fig. 4B and Table 5). Our combined model indicated that both monkeys and humans overweighed all probabilities significantly more when the positive RPE was larger compared to no RPE. The observation that probability weighting is not static was supported by earlier findings in the monkey literature, where the same animals exhibited different probability distortions under different experimental conditions (*14*, *15*), one of which showed that the reward sequence structure affects probability distortion (*15*). Probability weighting has also been shown to be unstable in studies involving human participants (*18*). This evidence supports our finding that changes in probability weighting can be modeled by allowing probability weighting to be dynamically adjusted according to the RPE, a critical algorithmic feature embedded in the nervous system (*19*, *20*, *23*, *49*).

Utility across species

Most previous studies found that monkeys have convex (*13*, *14*) or concave (*12*, *45*, *46*, *50*) utility over rewards in the gain domain. The monkeys and humans in our study exhibited the same concave utility shape (Fig. 2B), with the utility function of humans more concave than the utility function of the two monkeys. A possible explanation for the convexity of the utility function in earlier monkey studies is that these studies did not account for the possibility of optimistic probability weighting. When we analyzed our data under the assumption of no probability distortions or using an inflexible one-parameter probability weighting function, our monkeys were estimated to have convex utility or much less concave utility than when the probability distortions were accounted for (Fig. 2, B and C, and Table 1). The preference for taking risk is captured through a more convex utility function when optimistic distortions of probabilities are not accounted for. Probability distortions should be estimated to reliably identify the shape of the utility function.

Several aspects distinguish experiments involving monkeys from typical human studies. Monkeys make many choices for which they obtain immediate rewards, which are usually very small (less than 0.3 ml of a liquid), whereas humans usually make fewer choices for larger rewards. Both these features may drive monkeys to take more risks than humans under typical experimental conditions. Consistent with this idea, Genest *et al.* (*51*) found that when large rewards were used in a single trial (>0.8 ml), monkeys′ utility function was concave. A similar phenomenon was observed in our previous study (*12*). In other experimental conditions, similar to natural foraging, the utility function of monkeys was also concave (*45*). Furthermore, existing literature has established the effect of satiety on utility functions and reward valuations (*12*, *52*–*54*), but it remains unclear whether an individual′s level of satiety or financial wealth affects subjective probabilities. These aspects are critical for the comparison of human and macaque economic decision-making.

Comparisons between humans and monkeys

Comparing humans and monkeys directly, Hayden and Platt (*25*) found that human decision makers can also display risk-seeking and convex utility functions for juice and monetary rewards under conditions resembling those in monkey experiments. In line with this study, we found that when the experimental design and analyses were the same between species, the utilities were of the same shape, although, in our case, they were concave. Although we chose different rewards for monkeys and humans to make them more desirable and motivating (i.e., liquid reward for thirsty monkeys and money for humans), the results should be comparable because of the high correlations in utility curvature estimated across monetary, liquid, and food rewards in humans (*30*).

One significant difference between species was the improvement in the model fit among the sequentially developed models (Fig. 2A). In humans, the EU largely improved the explanatory ability, while the two-parameter probability weighting functions improved it in monkeys. Moreover, the effect of negative RPEs may also differ between these two species, but this difference was inconsistent (Fig. 4B), possibly because the sensitivity to negative events varied between individuals (*34*, *35*). Together, these two studies reinforce the view that methodology across studies must be aligned for cross-species comparisons. This calls for further studies in which different species are studied under the same paradigm.

Neural mechanisms of economic choice

How are RPE, utility, and probability weighting integrated in the brain? Each component of these signals was processed in the reward circuit. For example, midbrain dopamine neurons send RPE signals (*19*, *20*) to cortical and subcortical regions to learn the values of items and options (*21*–*23*). Neural signals in target areas of dopamine neuron's projections reflect utility or probability weighting in human and nonhuman primates (*30*, *43*, *44*, *55*, *56*). A recent finding suggests that dopamine neurons signal prediction errors from the predicted utility of rewards (*57*) and may influence dynamic and static signals related to economic decisions in target brain regions such as the striatum and cortices. These brain regions may learn values (*24*, *26*, *58*) with different learning rates that underlie the neural basis of reinforcement learning (*39*, *49*, *59*, *60*). These previous findings, as well as our present finding of RPE effects on probability distortions, link symbol- and experience-based decision processes (*61*), which are studied across different species using different approaches. However, both may be driven by a common neural mechanism in the reward decision circuitry. The neural signal reflecting utility and probability weighting that incorporates RPE dynamics might be widely present within this circuitry and can be explored using the dynamic prospect theory model presented here.

MATERIALS AND METHODS

Monkey study details

Subjects and experimental procedure

Two rhesus monkeys (*Macaca mulatta*, SUN, 7.1 kg, male; *Macaca fuscata*, FU, 6.7 kg, female) performed the tasks. All experimental procedures were approved by the Animal Care and Use Committee of the University of Tsukuba (protocol no. H30.336) and were performed in compliance with the U.S. Public Health Service's *Guide for the Care and Use of Laboratory Animals*. Each animal was implanted with a head-restrained prosthesis. Eye movements were measured using a video camera system at 120 Hz. Visual stimuli were generated using a liquid crystal display at 60 Hz placed 38 cm from the monkey's face when seated. The subjects performed lottery tasks 5 days a week. They practiced lottery tasks for 10 months, after which they became proficient in choosing lottery options.

Lottery tasks

The animals learned to perform two visually cued lottery tasks: a single-lottery task and a lottery choice task. The single-lottery task was aimed at providing monkeys with experiences of associations between visual cues and lottery outcomes. The lottery choice task aimed to provide monkeys with experience in choosing a lottery from two visual cues and then experiencing the chosen outcomes.

Single-lottery task. At the beginning of each trial, the monkeys had 2 s to align their gaze to within 3° of a 1°-diameter gray central fixation target. After fixating for 1 s, an 8° pie chart providing information regarding the probability and magnitude of rewards was presented for 2.5 s at the same location as that of the central fixation target. The magnitude and probability are indicated by the number of green and blue pie chart segments, respectively. The pie chart was then removed, and 0.2 s later, 1- and 0.1-kHz tones of 0.15-s duration were used to inform the monkeys whether they would receive the reward or no-reward, respectively. The high tone preceded the reward delivery by 0.2 s. The low tone indicates that no-reward is delivered. The animals received a liquid reward, as indicated by the number of green pie chart segments, with the probability indicated by the number of blue pie chart segments. An intertrial interval of 4 to 6 s followed each trial.

Lottery choice task. At the beginning of each trial, the monkeys had 2 s to align their gaze to within 3° of a 1°-diameter gray central fixation target. After fixation for 1 s, two peripheral 8° pie charts providing information regarding the probability and magnitude of rewards for each of the two target options were presented for 2.5 s at 8° to the left and right sides of the central fixation location. The gray 1° choice targets appeared at the same locations. After a 0.5-s delay, the fixation target disappeared, cueing saccade initiation. The monkeys were allowed 2 s to make their choice by shifting their gaze to either target within 3° of the choice target. Respective 1- and 0.1-kHz tones sounded for 0.15 s to denote reward and no-reward outcomes. The animals received a liquid reward, as indicated by the number of green pie chart segments for the chosen target, with the probability indicated by the number of blue pie chart segments. An intertrial interval of 4 to 6 s followed each trial.

Payoff, block structure, and data collection

Green and blue pie charts indicated reward magnitudes from 0.1 to 1.0 ml, in 0.1-ml increments, and reward probabilities from 0.1 to 1.0, in 0.1 increments, respectively. In total, 100 pie chart combinations were used in this study. In the single-lottery task, each pie chart was presented once in random order, allowing monkeys to experience all 100 lotteries within a certain period. For the choice task, two pie charts were randomly allocated to the left and right options for each trial. Approximately 30 to 60 trial blocks of the choice task were sometimes interleaved with 100 to 120 trial blocks of the single-lottery task.

Calibration of the reward supply system

A precise amount of liquid reward was delivered to the monkeys using a solenoid valve. An 18-gauge tube (of 0.9-mm inner diameter) was attached to the tip of the delivery tube to reduce variation across trials. The amount of reward in each payoff condition was calibrated by measuring the weight of water with 0.002-g precision (2 μl) on a single trial basis. This calibration method was the same as that previously used (*58*).

Human study details

Subjects

Seventy-two participants [34 female, 36 male; mean age, 21.65 years; SD (age), 3.82 years] were recruited from the student research volunteer database at the University of Sydney (*62*). All procedures were approved by the Human Ethics Committee of the University of Sydney (protocol no. 2020/558), and informed consent was obtained from all participants.

Choice task

To arrive at a practical number of decision scenarios for the human study, we decreased the number of reward magnitudes and probabilities relative to the monkey task. We created 25 unique lotteries by crossing five reward magnitudes ($1, $2, $3, $4, and $5) with five probability levels (0.1 to 0.9 in 0.2 increments). We then created all possible binary decision scenarios from these 25 lotteries and removed those that involved a first-order stochastically dominated option. We added 25 more decision scenarios by pairing $3 received with certainty with each of the 25 lotteries. This resulted in a total of 125 unique decision scenarios. Participants performed two blocks

of these scenarios, for a total of 250 decisions per individual. The scenarios were presented in an order that was individually randomized for each participant within each block, and participants could take up to a 2-min break between the blocks. There was no limit on response times, and the participants made decisions at their own pace by clicking on the preferred option. The lottery reward magnitudes and probabilities were presented in an identical manner as in the monkey study via green and blue pie charts, and the participants received written instructions on the meaning of these pie charts before they started the task. After every decision, the participants received feedback on whether their chosen lottery had won or not. In the first case, they saw the amount won in red at the center of the screen. In the second case, they would see a gray X at the center of the screen. To ensure that the decisions were consequential, each participant was paid for 20 randomly selected decisions using PayPal. The expected payment for a risk-neutral chooser in this task was $39. The sample sizes required to detect effect sizes on the aggregated level (number of total decision trials) were defined on the basis of the previous monkey studies (*14*, *15*).

Statistical analysis

Stata was used for statistical analyses. All statistical tests were two tailed. The structural models were estimated for each monkey separately. For human data, because of a smaller number of choices, we use representative agent approach, i.e., estimate one model for all 70 participants with SEs clustered at the participant level. A limitation of this approach is that we did not investigate individual heterogeneity in behavior within human species.

Mathematical models

We used standard economic models to estimate the utility function and probability weighting function in Stata. We used a standard reinforcement learning model to estimate the RPE using R software.

Economic models

We estimated the parameters of the utility and probability weighting functions within a random utility framework. Specifically, a lottery $L(m, p)$ denotes a gamble that pays m (magnitude of the offered reward in milliliters for monkeys and in Australian dollars for humans) with probability p, and 0 otherwise. We assumed a popular constant relative risk attitude (also known as power) utility function, $U(m) = m^{\alpha}$, and considered four previously proposed probability weighting functions: TK (*3*): $w(p) = \frac{p^{\gamma}}{[p^{\gamma}+(1-p)^{\gamma}]^{1/\gamma}}$; P1 (*63*): $w(p) = \exp[-(-\log p)^{\gamma}]$; P2 (*63*): $w(p) = \exp[-\delta(-\log p)^{\gamma}]$; and GE (*64*): $w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}$. We assumed that subjective probabilities and utilities are integrated multiplicatively per the standard economic theory, yielding the expected subjective utility function $V(m, p) = w(p)m^{\alpha}$.

The following six models were used:

(i) EV: expected value $V(m, p) = pm$

(ii) EU: expected utility $V(m, p) = pm^{\alpha}$

(iii) TK: prospect theory with $w(p)$ as in Kahneman and Tversky (*3*)

$$V(m,p) = \frac{p^{\gamma}}{[p^{\gamma}+(1-p)^{\gamma}]^{1/\gamma}}\, m^{\alpha}$$

(iv) P1: prospect theory with $w(p)$ with one parameter as in Prelec (*63*)

$$V(m,p) = \exp[-(-\log p)^{\gamma}]m^{\alpha}$$

(v) P2: prospect theory with $w(p)$ with two parameters as in Prelec (*63*)

$$V(m,p) = \exp[-\delta(-\log p)^{\gamma}]m^{\alpha}$$

(vi) GE: prospect theory with $w(p)$ as in Goldstein and Einhorn (*64*)

$$V(m,p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma}+(1-p)^{\gamma}}\, m^{\alpha}$$

where α, δ, and γ are free parameters, and m and p are the magnitude and probability cued by the lottery, respectively. These six models were nested and sequentially developed from simple EV to construct more complex prospect theory models.

The probability of participants choosing the lottery on the right side (L_R) instead of the lottery on the left side (L_L) was estimated using a logistic choice function, $P(L_R) = 1/(1 + e^{-Z})$, where $Z = \frac{V(L_R)-V(L_L)}{\beta}$ and the free parameter β captures the degree of stochasticity observed in choice. We fit the data by maximizing the log likelihood. The multiple free parameters in each model were estimated using the maximum likelihood procedure (*65*), resulting in one estimate with a confidence interval corresponding to each parameter in each model.

In each fitted model, we applied the Wald test to determine whether α, β, δ, and γ were significantly different from zero and from one. The number of stars in the results tables indicates the level of significance of the estimate being different from zero: $^{+}P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

We chose the best structural model to describe participant behavior using BIC (*66*): $BIC(\text{model}) = -2L + k\ ln(n)$, where L is the maximum log likelihood of the model, k is the number of free parameters, and n is the sample size of the model. We only used the BIC (except for in the "Robustness checks" section) because the dataset was large, meaning that the additional parameter penalty would be too small if we used the AIC: $AIC(\text{model}) = -2L + 2k$, where L is the maximum log likelihood of the model and k is the number of free parameters.

Reinforcement learning model

We used a standard temporal difference (TD) learning model (*4*) to estimate reward prediction for lottery stimuli. Let $V(m, p)_t$ represent the reward prediction from the chosen lottery option in trial t. Let r_t be the reward amount received in trial t in milliliters for monkeys and in Australian dollars for humans if the trial is rewarded and 0 if the trial is not rewarded. In trial t, upon receiving the actual reward (or not), the participants updated their reward prediction $V(m,p)_{t+1}$ for future trials using the TD model, as follows

$$V(m,p)_{t+1} = V(m,p)_t + A\Delta_t$$

where A is the learning rate ($0 \le A \le 1$) and Δ_t is the RPE. The RPE was defined as $\Delta_t = r_t - V(m, p)_t$.

Assuming that $V(m, p)_0 = 0$, we simulated the TD algorithm for 10,000 experimental trials for each lottery using different values of the learning rate A. In Fig. 3, we plot $V(m, p)_{10,000}$, which is the lottery valuation at which the algorithm arrived, against its expected

value. Ultimately, the algorithm values the lotteries close to their expected values, particularly at low learning rates. The RPE estimated using the TD algorithm is close to the obtained reward minus the expected value. $V(m, p)_t$ was updated for the chosen target during the choice task and for the lottery during the single-lottery task.

Dynamic prospect theory models

We calculated RPE as the difference between the reward received and the expected value of the chosen lottery. Therefore, by design, the RPE was positive ($\Delta\text{positive}_{t-1}$) in our experiment after receiving the reward, except (p = 1.0), and negative ($\Delta\text{negative}_{t-1}$) after no-reward was provided. Our dynamic prospect theory model includes dynamic parameters indexed by trial t

$$V(p, m)_t = \exp[-\delta_t(-\log p)^{\gamma_t}]\ m^{\alpha_t}$$

In this P2-based combined model, we allowed the dynamic parameters α_t, δ_t, and γ_t to depend on the RPE variables as follows

$$\alpha_t = \alpha_0 + \sum_i \alpha_i z_i$$
$$\delta_t = \delta_0 + \sum_i \delta_i z_i$$
$$\gamma_t = \gamma_0 + \sum_i \gamma_i z_i$$

where z_i is a set of RPE variables that capture the experience in the last trial, $t-1$.

In the standard RPE model, we assumed that only the difference between the expected value of the outcome and the received outcome affects each parameter. The standard RPE model included seven free parameters: α_0, δ_0, γ_0, $\sum_i \alpha_i z_i = \alpha_1 \Delta_{t-1}$, $\sum_i \delta_i z_i = \delta_1 \Delta_{t-1}$, $\sum_i \gamma_i z_i = \gamma_1 \Delta_{t-1}$, and β, which measures the choice stochasticity. In the separate RPE model, z_i was $\Delta\text{positive}_{t-1}$ and $\Delta\text{negative}_{t-1}$. The separate RPE model included 10 free parameters: α_0, δ_0, γ_0, $\sum_i \alpha_i z_i = \alpha_1 \Delta\text{positive}_{t-1} + \alpha_2 \Delta\text{negative}_{t-1}$, $\sum_i \delta_i z_i = \delta_1 \Delta\text{positive}_{t-1} + \delta_2 \Delta\text{negative}_{t-1}$, $\sum_i \gamma_i z_i = \gamma_1 \Delta\text{positive}_{t-1} + \gamma_2 \Delta\text{negative}_{t-1}$, and β, which measures the choice stochasticity. Whether α_0, δ_0, γ_0, α_i, δ_i, and γ_i were significantly different from zero was determined using the Wald test. The number of stars in the results tables indicates the level of significance of each result: $^{+}P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$. To select a model with the best-fitting selection of explanatory variables, we used the separate RPE model and eliminated the least significant variable, one at a time, and continued until the BIC stopped improving.

Parameter recovery

We generated 50 synthetic choosers endowed with a randomly selected set of parameters for utility and probability weighting functions. For each simulated chooser, we generated 5000 decisions. The decision scenarios were randomly selected from a set of all possible decision scenarios (10,000). The sets of parameters were randomly selected from the following uniform distributions U(lowest, highest): alpha ~ U(0.1, 1.9), delta ~ U(0.1, 3), gamma ~ U(0.1, 3), and the six parameters that depend on the positive and negative RPEs ~ U(−1,1). For each chooser in each decision scenario, we calculated the utility of each option with noise, with a mean of zero and an SD of 0.03. We then assumed that the choosers always chose the option with higher utility. We repeated this procedure for three models: EU, static prospect theory with P2, and dynamic prospect theory separate RPE model with P2. We used these simulated choice datasets to estimate parameters of the generative model. Then, we estimated the correlation between the estimated parameters and the original parameters used to generate the data. We reported the parameter recovery correlations for 40 simulated choosers because the static prospect theory P2 model yielded unreasonably low estimates for five sets of assumed parameters and did not converge for the other five sets of parameters.

Cross-validation

As in the previous analysis, we analyzed our three datasets—monkey SUN, monkey FU, and human participants—separately. We randomly divided each dataset into four subsamples using split-sample command in Stata. Data from one block in monkeys and from the same participant in humans were not separated. Then, we used three subsamples (~75% of the data) as a training sample and one subsample (~25%) as a holdout sample (i.e., test data). We used the training sample to estimate EU, static prospect theory, dynamic prospect theory with standard RPE, dynamic prospect theory with separate RPE, and the best-fit model selected according to BIC using whole data. SEs were clustered on the participant level in the human data. Then, we used the estimated model parameters to calculate the expected subjective utilities of lotteries and predict choices in the holdout sample. We calculated log likelihood for each model in the holdout sample as a measure of model performance. We simply compared the log likelihoods instead of BICs to evaluate the effect of overfitting in our samples (*67*). If the overfitting of the model occurs, log likelihood of the models estimated in the holdout samples becomes worse. This is also because we did not estimate the model parameters in the holdout samples, and no free parameters existed when evaluating the model performances in the holdout samples. We repeated the procedure four times, each time using a different subsample as a holdout sample and calculating the log likelihoods. Last, we summed up the log likelihoods obtained in each holdout subsample (table S7 presents the results).

Nonparametric analyses

To examine the robustness of the finding that the probability weighting functions are concave, we examined how reward probability affects participants' willingness to take gambles using logistic regression analysis (Table 2). The likelihood of choosing target 1 (right) was analyzed using logistic regression.

$$P\text{chooses}_1 = 1/(1 + e^{-Z})$$

where Z combines a constant as well as variables that capture information about the reward probability and magnitude of targets 1 (right) and 2 (left) in a trial as follows

$$Z = b_0 + b_1 m_1 + b_2 m_2 + b_3 p_3 + b_4 p_4 + b_5 p_1 \le 0.3 + b_6 p_1 \ge 0.7$$

where b_0 is the intercept. m_1 and m_2 denote the reward magnitudes for targets 1 (right) and 2 (left), respectively. p_1 and p_2 denote the probability of receiving a reward for targets 1 (right) and 2 (left), respectively, and capture the linear effect of the probabilities on the choice. To identify nonlinearities in probability weighting, we differentiated between three probability levels for target 1 (right): $p_1 \le 0.3$, p_1 = 0.4 to 0.6, and $p_1 \ge 0.7$. We prepared two indicator variables: The first equals one when the reward probability for target 1 is small ($p_1 \le 0.3$), and zero otherwise, and the second equals one

when the reward probability for target 1 is large ($p_1 \geq 0.7$), and zero otherwise. Through a comparison to an omitted category in which the reward probability was between 0.4 and 0.6, our regression allowed us to check whether the probability of choosing target 1 was larger or smaller than what of a linear model would predict when the reward probability was high or low. If probabilities contribute to the decision linearly, coefficients b_5 and b_6 should be insignificant. If b_5 and b_6 were smaller than 0, with a statistical significance level of $P < 0.05$, the participant's choice was consistent with the concave probability weighting function. If b_5 is larger and b_6 is smaller than 0, with a statistical significance level of $P < 0.05$, this would be consistent with an inverse S-shaped probability weighting function.

To examine whether RPEs with a nonparametric analysis affected the participants' choices (Table 3), we used a logistic regression to examine whether the RPEs in the previous trials affected the monkey's willingness to take risks. For this analysis, we determined whether target 1 or 2 was riskier in each trial by calculating and comparing the coefficient of variation for each lottery as the ratio of the SD of the lottery to its mean (*36*). We subsequently regressed the explanatory variables on whether the chooser selected a riskier option as follows

$$P\text{chooses}_{\text{risky}} = 1/(1 + e^{-Z})$$
$$Z = b_0 + b_1 m_{\text{risky}} + b_2 m_{\text{safe}} + b_3 p_{\text{risky}} + b_4 p_{\text{safe}} + b_5 \Delta_{t-1}$$

where b_0 is the intercept; m_{risky} and m_{safe} denote the reward magnitudes for the riskier and safer options in a trial, respectively; and p_{risky} and p_{safe} denote the probability of receiving a reward in the riskier and safer options in a trial, respectively. Δ_{t-1} is the RPE of the previous trial. If the coefficient b_5 was different from 0 with a statistical significance level of $P < 0.05$, we would conclude that RPE significantly affects risk-taking behavior.

We also modeled two separate RPEs as follows

$$Z = b_0 + b_1 m_{\text{risky}} + b_2 m_{\text{safe}} + b_3 p_{\text{risky}} + b_4 p_{\text{safe}} + b_5 \Delta\text{positive}_{t-1} + b_6 \Delta\text{negative}_{t-1}$$

where $\Delta\text{positive}_{t-1}$ is the positive RPE after receiving the reward, and $\Delta\text{negative}_{t-1}$ is the negative RPE after receiving no-reward. If the coefficients b_5 and b_6 are different from 0, with a statistical significance level of $P < 0.05$, we would conclude that the positive and negative RPEs significantly affect risk-taking behavior.

Supplementary Materials

This PDF file includes:
Figs. S1 to S4
Tables S1 to S7
Legends for data S1 to S9

Other Supplementary Material for this manuscript includes the following:
Data S1 to S9

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. P. W. Glimcher, C. F. Camerer, E. Fehr, R. A. Poldrack, *Neuroeconomics: Decision Making and the Brain*, P. W. Glimcher, C. F. Camerer, E. Fehr, R. A. Poldrack, Eds. (Elsevier Press, 2008).
2. C. Camerer, G. Loewenstein, G. Prelec, Neuroeconomics: How neuroscience can inform economics. *J. Econ. Lit.* **43**, 9–64 (2005).
3. D. Kahneman, A. Tversky, Prospect theory: An analysis of decision under risk. *Econometrica* **47**, 263–292 (1979).
4. R. S. Sutton, A. G. Barto, *Reinforcement Learning* (The MIT Press, 1998).
5. Y. Shen, M. J. Tobia, T. Sommer, K. Obermayer, Risk-sensitive reinforcement learning. *Neural Comput.* **26**, 1298–1328 (2014).
6. G. Wu, R. Gonzalez, Curvature of the probability weighting function. *Manage. Sci.* **42**, 1676–1690 (1996).
7. A. Bruhin, H. Fehr-Duda, T. Epper, Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica* **78**, 1375–1412 (2010).
8. M. Abdellaoui, Parameter-free elicitation of utility and probability weighting functions. *Manage. Sci.* **46**, 1497–1512 (2000).
9. W. Harbaugh, K. Krause, L. Vesterlund, Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Exp. Econ.* **5**, 53–84 (2002).
10. G. W. Harrison, E. E. Rutstrom, Expected utility theory and prospect theory: One wedding and a decent funeral. *Exp. Econ.* **12**, 133–158 (2009).
11. H. Fehr-Duda, T. Epper, A. Bruhin, R. Schubert, Risk and rationality: The effects of mood and decision rules on probability weighting. *J. Econ. Behav. Organ.* **78**, 14–24 (2011).
12. H. Yamada, A. Tymula, K. Louie, P. W. Glimcher, Thirst-dependent risk preferences in monkeys identify a primitive form of wealth. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15788–15793 (2013).
13. W. R. Stauffer, A. Lak, P. Bossaerts, W. Schultz, Economic choices reveal probability distortion in macaque monkeys. *J. Neurosci.* **35**, 3146–3154 (2015).
14. S. Farashahi, H. Azab, B. Hayden, A. Soltani, On the flexibility of basic risk attitudes in monkeys. *J. Neurosci.* **38**, 4383–4398 (2018).
15. S. Ferrari-Toniolo, P. M. Bujold, W. Schultz, Probability distortion depends on choice sequence in rhesus monkeys. *J. Neurosci.* **39**, 2915–2929 (2019).
16. A. Nioche, N. P. Rougier, M. Deffains, S. Bourgeois-Gironde, S. Ballesta, T. Boraud, The adaptive value of probability distortion and risk-seeking in macaques' decision-making. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20190668 (2021).
17. R. Hertwig, I. Erev, The description-experience gap in risky choice. *Trends Cogn. Sci.* **13**, 517–523 (2009).
18. R. K. Jessup, A. J. Bishara, J. R. Busemeyer, Feedback produces divergence from prospect theory in descriptive choice. *Psychol. Sci.* **19**, 1015–1022 (2008).
19. P. R. Montague, P. Dayan, T. J. Sejnowski, A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
20. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
21. S. J. Gershman, N. D. Daw, Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annu. Rev. Psychol.* **68**, 101–128 (2017).
22. B. B. Averbeck, V. D. Costa, Motivational neural circuits underlying reinforcement learning. *Nat. Neurosci.* **20**, 505–512 (2017).
23. A. S. Lowet, Q. Zheng, S. Matias, J. Drugowitsch, N. Uchida, Distributional reinforcement learning in the brain. *Trends Neurosci.* **43**, 980–997 (2020).
24. Y. Imaizumi, A. Tymula, Y. Tsubo, M. Matsumoto, H. Yamada, A neuronal prospect theory model in the brain reward circuitry. *Nat. Commun.* **13**, 5855 (2022).
25. B. Y. Hayden, M. L. Platt, Gambling for gatorade: Risk-sensitive decision making for fluid rewards in humans. *Anim. Cogn.* **12**, 201–207 (2009).
26. H. Yamada, Y. Imaizumi, M. Matsumoto, Neural population dynamics underlying expected value computation. *J. Neurosci.* **41**, 1684–1698 (2021).
27. J. D. Hey, C. Orme, Investigating generalizations of expected utility theory using experimental data. *Econometrica* **62**, 1291–1326 (1994).
28. Y. Wang, N. Ma, X. He, N. Li, Z. Wei, L. Yang, R. Zha, L. Han, X. Li, D. Zhang, Y. Liu, X. Zhang, Neural substrates of updating the prediction through prediction error during decision making. *Neuroimage* **157**, 1–12 (2017).
29. S. L. Cheung, A. Tymula, X. Wang, Present bias for monetary and dietary rewards. *Exp. Econ.* **25**, 1202–1233 (2022).
30. D. J. Levy, P. W. Glimcher, Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain. *J. Neurosci.* **31**, 14693–14707 (2011).
31. H. Yamada, K. Louie, P. W. Glimcher, Controlled water intake: A method for objectively evaluating thirst and hydration state in monkeys by the measurement of blood osmolality. *J. Neurosci. Methods* **191**, 83–89 (2010).
32. T. Minamimoto, H. Yamada, Y. Hori, T. Suhara, Hydration level is an internal variable for computing motivation to obtain water rewards in monkeys. *Exp. Brain Res.* **218**, 609–618 (2012).

33. H.-M. von Gaudecker, A. van Soest, E. Wengstrom, Heterogeneity in risky choice behavior in a broad population. *Am. Econ. Rev.* **101**, 664–694 (2011).
34. H. Seo, X. Cai, C. H. Donahue, D. Lee, Neural correlates of strategic reasoning during competitive games. *Science* **346**, 340–343 (2014).
35. M. Matsumoto, K. Matsumoto, H. Abe, K. Tanaka, Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* **10**, 647–656 (2007).
36. E. U. Weber, S. Shafir, A. R. Blais, Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychol. Rev.* **111**, 430–445 (2004).
37. S. B. Broomell, S. Bhatia, Parameter recovery for decision modeling using choice data. *Decision* **1**, 252–274 (2014).
38. J. O'Doherty, P. Dayan, J. Schultz, R. Deichmann, K. Friston, R. J. Dolan, Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
39. H. Yamada, H. Inokawa, N. Matsumoto, Y. Ueda, K. Enomoto, M. Kimura, Coding of the long-term value of multiple future rewards in the primate striatum. *J. Neurophysiol.* **109**, 1140–1151 (2013).
40. K. Yamanaka, Y. Hori, T. Minamimoto, H. Yamada, N. Matsumoto, K. Enomoto, T. Aosaki, A. M. Graybiel, M. Kimura, Roles of centromedian parafascicular nuclei of thalamus and cholinergic interneurons in the dorsal striatum in associative learning of environmental events. *J. Neural Transm.* **125**, 501–513 (2018).
41. N. Kettlewell, F. Rijsdijk, S. Siribaddana, A. Sumathipala, A. Tymula, H. Zavos, N. Glozier, Civil war, natural disaster and risk preferences: Evidence from Sri Lankan twins (IZA Discussion Paper No. 11901, 2018).
42. J. Guo, A. Tymula, Waterfall illusion in risky choice–Exposure to outcome-irrelevant gambles affects subsequent valuation of risky gambles. *Eur. Econ. Rev.* **139**, 103889 (2021).
43. M. Hsu, I. Krajbich, C. Zhao, C. F. Camerer, Neural response to reward anticipation under risk is nonlinear in probabilities. *J. Neurosci.* **29**, 2231–2237 (2009).
44. P. N. Tobler, G. I. Christopoulos, J. P. O'Doherty, R. J. Dolan, W. Schultz, Neuronal distortions of reward probability without choice. *J. Neurosci.* **28**, 11703–11711 (2008).
45. B. R. Eisenreich, B. Y. Hayden, J. Zimmermann, Macaques are risk-averse in a freely moving foraging task. *Sci. Rep.* **9**, 15091 (2019).
46. A. Nioche, S. Bourgeois-Gironde, T. Boraud, An asymmetry of treatment between lotteries involving gains and losses in rhesus monkeys. *Sci. Rep.* **9**, 10441 (2019).
47. A. Fujimoto, T. Minamimoto, Trait and state-dependent risk attitude of monkeys measured in a single-option response task. *Front. Neurosci.* **13**, 816 (2019).
48. D. R. Cavagnaro, M. A. Pitt, R. Gonzalez, J. I. Myung, Discriminating among probability weighting functions using adaptive design optimization. *J. Risk Uncertain.* **47**, 255–289 (2013).
49. J. C. Houk, J. L. Adams, A. G. Barto, *Models of Information Processing in the Basal Ganglia*, J. C. Houk, J. L. Davis, D. G. Beiser, Eds. (The MIT Press, 1995), pp. 249–270.
50. S. Ferrari-Toniolo, P. M. Bujold, F. Grabenhorst, R. Baez-Mendoza, W. Schultz, Non-human primates satisfy utility maximization in compliance with the continuity axiom of expected utility theory. *J. Neurosci.* **41**, 2964–2979 (2021).
51. W. Genest, W. R. Stauffer, W. Schultz, Utility functions predict variance and skewness risk preferences in monkeys. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8402–8407 (2016).
52. T. Caraco, S. Martindale, T. S. Whitham, An empirical demonstration of risk-sensitive foraging preferences. *Anim. Behav.* **28**, 820–830 (1980).
53. F. Brito e Abreu, A. Kacelnik, Energy budgets and risk-sensitive foraging in starlings. *Behav. Ecol.* **10**, 338–345 (1999).
54. H. Yamada, Hunger enhances consistent economic choices in non-human primates. *Sci. Rep.* **7**, 2394 (2017).
55. A. Pastor-Bernier, A. Stasiak, W. Schultz, Orbitofrontal signals for two-component choice options comply with indifference curves of revealed preference theory. *Nat. Commun.* **10**, 4885 (2019).
56. X. Chen, V. Stuphorn, Inactivation of medial frontal cortex changes risk Preference. *Curr. Biol.* **28**, 3114–3122.e4 (2018).
57. W. R. Stauffer, A. Lak, S. Kobayashi, W. Schultz, Components and characteristics of the dopamine reward utility signal. *J. Comp. Neurol.* **524**, 1699–1711 (2016).
58. H. Yamada, K. Louie, A. Tymula, P. W. Glimcher, Free choice shapes normalized value signals in medial orbitofrontal cortex. *Nat. Commun.* **9**, 162 (2018).
59. K. Enomoto, N. Matsumoto, H. Inokawa, M. Kimura, H. Yamada, Topographic distinction in long-term value signals between presumed dopamine neurons and presumed striatal projection neurons in behaving monkeys. *Sci. Rep.* **10**, 8912 (2020).
60. H. Inokawa, N. Matsumoto, M. Kimura, H. Yamada, Tonically active neurons in the monkey dorsal striatum signal outcome feedback during trial-and-error search behavior. *Neuroscience* **446**, 271–284 (2020).
61. B. Garcia, F. Cerrotti, S. Palminteri, The description–experience gap: A challenge for the neuroeconomics of decision-making under uncertainty. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **376**, 20190665 (2021).
62. B. Greiner, Subject pool recruitment procedures: Organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* **1**, 114–125 (2015).
63. D. Prelec, The probability weighting function. *Econometrica* **66**, 497–527 (1998).
64. W. M. Goldstein, H. J. Einhorn, Expression theory and the preference reversal phenomena. *Psychol. Rev.* **94**, 236–254 (1987).
65. G. W. Harrison, E. E. Rutström, *Handbook of Experimental Economics Results*, in *Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods*, (Elsevier, 2008), vol. 1, pp. 752–767.
66. K. Burnham, D. Anderson, Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Method Res.* **33**, 261–304 (2004).
67. K. Amit, S. Vitalie, P. W. Peter, An experimental test of prospect theory for predicting choice under ambiguity. *J. Risk Uncertain.* **48**, 1–17 (2014).

Acknowledgments: We wish to express appreciation to R. Tajiri, M. Nejime, V. Sihui, Y. Yabana, Y. Suwa, and S. Nishino for technical assistance. We thank N. Hagura for comments on the analysis via a social networking service. Monkey FU was provided by NBRP "Japanese Monkeys" through the National Bio Resource Project of MEXT, Japan. **Funding:** This study was supported by JSPS KAKENHI grant numbers JP:15H05374 and 21H02797, Takeda Science Foundation, Council for Addiction Behavior Studies, Narishige Neuroscience Research Foundation, Moonshot R&D JPMJMS2294 (H.Y.), and ARC DP190100489 (A.T.). **Author contributions:** H.Y., A.T., and X.W. designed the study. H.Y., X.W., and Y.I. conducted experiments. M.M. and T.K. conducted a part of experiments. A.T. and H.Y. developed the new model and analytical tools. A.T., X.W., and H.Y. analyzed the data. A.T., H.Y., X.W., and J.K. evaluated the results. A.T. and H.Y. wrote the manuscript. All authors edited and approved the final manuscript. A.T. and H.Y. completed the entire process for the economic and neuroscientific aspects, respectively. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Monkey experiment: Raw data: data_monkeyfu.txt; data_monkeysun.txt. Analysis codes: Code_RLSimu.r for the reinforcement learning model simulation, StataCode_get_results.do and StataCode_structural_models.do for the statistical tests. Human experiment: Raw data: HumData_all_session_full.csv. Analysis codes: Humget_results_paper.do, Humopen_dataset_hu_paper.do, Humstructural_models_paper.do for the statistical tests.

Submitted 9 September 2022
Accepted 14 April 2023
Published 19 May 2023
10.1126/sciadv.ade7972