Report

Current Biology

An Analysis of Decision under Risk in Rats

Highlights

- A novel task enables application of core behavioral economic approaches in rodents
- Like humans, rats exhibit nonlinear utility and probability weighting
- Rats also exhibit trial history effects, consistent with ongoing learning
- A reinforcement learning model incorporating subjective value accounts for the data

Authors

Christine M. Constantinople, Alex T. Piet, Carlos D. Brody

Correspondence

constantinople@nyu.edu

In Brief

Constantinople et al. apply prospect theory, the predominant economic theory of decision-making under risk, to rats. Rats exhibit signatures of both prospect theory and reinforcement learning. The authors present a model that integrates these frameworks, accounting for rats' nonlinear econometric functions and also trial-by-trial learning.

Constantinople et al., 2019, Current Biology *29*, 1–9
June 17, 2019 © 2019 Elsevier Ltd.
https://doi.org/10.1016/j.cub.2019.05.013

CellPress

Please cite this article in press as: Constantinople et al., An Analysis of Decision under Risk in Rats, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.05.013

Current Biology
Report

An Analysis of Decision under Risk in Rats

Christine M. Constantinople,[1,5,6,*] Alex T. Piet,[1,4] and Carlos D. Brody[1,2,3]
[1]Princeton Neuroscience Institute, Princeton University, Washington Road, Princeton, NJ 08544, USA
[2]Department of Molecular Biology, Princeton University, Washington Road, Princeton, NJ 08544, USA
[3]Howard Hughes Medical Institute, Princeton University, Washington Road, Princeton, NJ 08544, USA
[4]Present address: Allen Institute for Brain Science, Westlake Avenue N, Seattle, WA 98109, USA
[5]Present address: Center for Neural Science, New York University, Washington Place, New York, NY 10003, USA
[6]Lead Contact
*Correspondence: constantinople@nyu.edu
https://doi.org/10.1016/j.cub.2019.05.013

SUMMARY

In 1979, Daniel Kahneman and Amos Tversky published a ground-breaking paper titled "Prospect Theory: An Analysis of Decision under Risk," which presented a behavioral economic theory that accounted for the ways in which humans deviate from economists' normative workhorse model, Expected Utility Theory [1, 2]. For example, people exhibit probability distortion (they overweight low probabilities), loss aversion (losses loom larger than gains), and reference dependence (outcomes are evaluated as gains or losses relative to an internal reference point). We found that rats exhibited many of these same biases, using a task in which rats chose between guaranteed and probabilistic rewards. However, prospect theory assumes stable preferences in the absence of learning, an assumption at odds with alternative frameworks such as animal learning theory and reinforcement learning [3–7]. Rats also exhibited trial history effects, consistent with ongoing learning. A reinforcement learning model in which state-action values were updated by the subjective value of outcomes according to prospect theory reproduced rats' nonlinear utility and probability weighting functions and also captured trial-by-trial learning dynamics.

RESULTS

Two key components of prospect theory are *utility* (rewards are evaluated by the subjective satisfaction or "utility" they provide) and *probability distortion* (people often overweight low and underweight high probabilities; Figure 1D). In this theory, subjective value is determined by the shapes of subjects' utility and probability weighting functions.

Learning theories provide an alternative account of subjective value. In animal learning theory, Thorndike's "Law of Effect" described the effect of reinforcers on action selection [8], and Pavlov's subsequent experiments demonstrated how animals learn to associate stimuli with rewards [9]. The Rescorla-Wagner model of classical conditioning formalized how such learning might occur [3–5]. Although these models described how animals might learn associations between stimuli, they were naturally extended to account for learning values from experience [7, 10]. Models of trial-and-error learning from animal learning theory form the basis for reinforcement learning algorithms, including temporal difference learning, which captures temporal relationships between predictors and outcomes [7, 11]. Reinforcement learning provides a powerful framework for value-based decision-making in psychology and neuroscience, in which value estimates are learned from experience and updated trial-to-trial based on prediction errors [7, 12, 13].

Reinforcement learning has had profound impact in part because many of its components have been related to neural substrates [12, 14–17]. However, standard reinforcement learning algorithms dictate that agents learn the *expected value* (volume × probability) of actions or outcomes with experience [6], meaning that they will exhibit linear utility and probability weighting functions. This is incompatible with prospect theory. We found that rats exhibited signatures of both prospect theory and reinforcement learning, and we present an initial attempt to integrate these frameworks. First, we focus on prospect theory.

Most economic studies examine decisions between clearly described lotteries (i.e., "decisions from description"). Studies of risky choice in rodents, however, typically examine decisions between prospects that are learned over time (i.e., "decisions from experience"), which are difficult to reconcile with prospect theory [18–20]. We designed a task in which reward probability and amount are communicated by sensory evidence, eliciting decisions from description rather than experience. This enabled behavioral economic approaches, such as estimating utility functions.

Rats initiated a trial by nose-poking in the center port of a three-port wall. Light flashes were presented from left and right side ports, and the number of flashes conveyed the probability of water reward at each port. Simultaneously, auditory clicks were presented from left and right speakers, and click rate conveyed the volume of water reward baited at each port (Figures 1A and 1B). One port offered a guaranteed or safe reward, and the other offered a risky reward with an explicitly cued probability. The safe and risky ports (left or right) varied randomly. One of four water volumes could be the guaranteed or risky reward (6, 12, 24, 48 μL); risky reward probabilities ranged from 0 to 1, in increments of 0.1 (Figures 1A and 1B).

High-throughput training generated 36 trained rats and many tens of thousands of choices per rat, enabling detailed behavioral

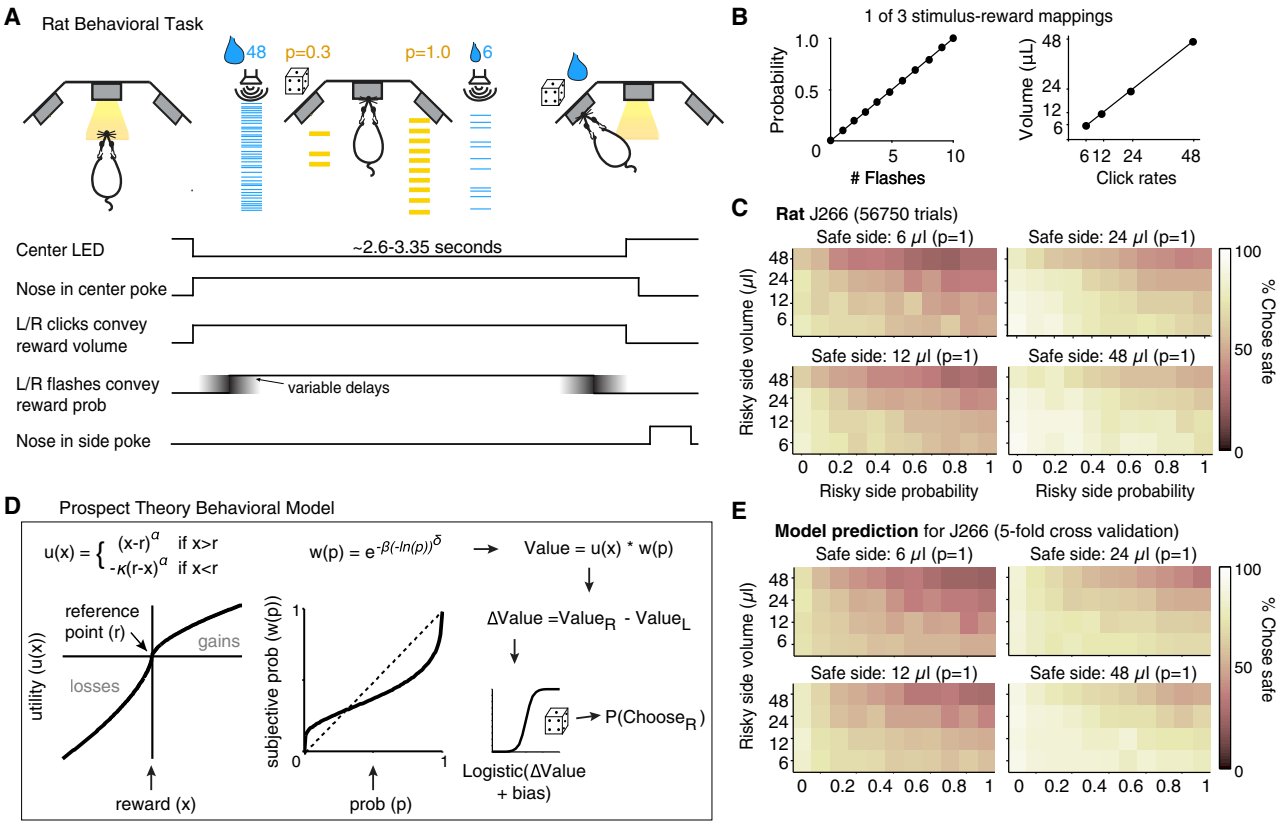

Figure 1. Rats Choose between Guaranteed and Probabilistic Rewards

(A) Behavioral task and timing of task events: flashes cue reward probability (p) and click rates convey water volume (x) on each side. Safe and risky sides are not fixed.

(B) Relationship between cues and reward probability and volume in one task version. Alternative versions produced similar results (Figure S2). There were four possible volumes (6, 12, 24, or 48 μL), and the risky side offered reward probabilities between 0 and 1 in increments of 0.1.

(C) One rat's performance for each of the safe side volumes. Axes are probability and volume of risky options.

(D) A behavioral model inferred the utility and probability weighting functions that best explained rats' choices. See Box 1 for details.

(E) Model prediction for held-out data from one rat, averaged over 5 test sets. See also Figures S1 and S2.

quantification. Rats demonstrated they learned the meaning of the cues by frequently ''opting-out'' of trials offering smaller rewards, leaving the center poke despite incurring a time-out penalty and white-noise sound (Figures S1A–S1C). This indicated that they associated the click rates with water volumes, instead of relying on a purely perceptual strategy. It is possible that opting-out, which persisted despite longer time-out penalties for low-volume trials (STAR Methods), reflected reward-rate maximizing strategies [21–23]. Rats favored prospects with higher expected value (Figures 1C and S1).

We used a standard choice model [24, 25] to estimate each rat's utility and probability weighting functions according to prospect theory (Box 1; Figures 1D and S1). The model predicted rats' choices on held-out data (Figure 1E). It outperformed alternative models, including one imposing linear probability weighting (according to Expected Utility Theory [26]), one that fit linear weights for probabilities and volumes, and several models implementing purely perceptual strategies with sensory noise (Figures S2A–S2F).

Concave utility (the utility function exponent $\alpha < 1$) produces diminishing marginal sensitivity, in which subjects are less sensitive to differences in larger rewards. Rats' median α was 0.54, indicating concave utility, like humans [2] (Figures 2A and S1F). To test for diminishing marginal sensitivity, we compared performance on trials offering guaranteed outcomes of 0 or 24 μL, and 24 or 48 μL (Figures 2B and 2C). Concave utility implies that 24 and 48 μL are less discriminable than 0 and 24 μL (Figure 2C). Indeed, the concavity of the utility function was correlated with reduced discriminability on trials offering 24 and 48 μL (Figures 2D and 2E; p = 1.03e−7, Pearson's correlation). This was true when the guaranteed outcome of 0 included trials offering 24 μL with p = 0 (Figures 2D and 2E), or all volumes with p = 0 (Figures S2G and S2H). Our choice set did not permit analysis of trials offering non-zero rewards, as these trials (24 versus 0, 48 versus 24) were the only ones with equal reward differences. This suggests that rats, like humans, exhibit diminishing marginal sensitivity.

Rats' probability weighting functions revealed overweighting of probabilities (Figure 2F). A logistic regression model that parameterized each probability to predict choice yielded regression weights mirroring the probability weighting functions (Figures 2G and 2H). Control experiments indicated that nonlinear

Box 1. Prospect Theory Behavioral Model

We modeled the probability that the rat chose the right side by a logistic function whose argument was the difference between the subjective value of each option $(V_R - V_L)$ plus a trial history-dependent term. Subjective utility was parameterized as:

$$u(x) = \begin{cases} (x-r)^{\alpha} & \text{if} \quad x > r \\ -\kappa(r-x)^{\alpha} & \text{if} \quad x < r, \end{cases} \tag{1}$$

where α is a free parameter, and x is reward volume. r is the reference point, which determines whether rewards are perceived as gains or losses. We first consider the case where $r = 0$, so

$$u(x) = x^{\alpha}. \tag{2}$$

The subjective probability of each option is computed by:

$$w(p) = e^{-\beta(-ln(p))^{\delta}}, \tag{3}$$

where β and δ are free parameters and p is the objective probability offered. Combining utility and probability yields the subjective value for each option:

$$V_R = u(x_R)w(p_R) \tag{4}$$

$$V_L = u(x_L)w(p_L). \tag{5}$$

These were normalized by the max over trials and transformed into choice probabilities via a logistic function:

$$P(Choose_R) = \iota + \frac{1-2\iota}{1+e^{-\lambda(V_R - V_L)+bias}}, \tag{6}$$

where ι captures stimulus independent variability (lapse rate) and λ determines the sensitivity of choices to the difference in subjective value $(V_R - V_L)$. The bias term was composed of three possible parameters, depending on trial history:

$$bias = \begin{cases} +/-h_1 & \text{if t-1was safe L/R choice} \\ +/-h_2 & \text{if t-1was risky L/R rew} \\ +/-h_3 & \text{if t-1was risky L/R miss.} \end{cases} \tag{7}$$

utility and probability weighting were not due to perceptual errors in estimating flashes and clicks [27] (Figures S2I–S2K).

To evaluate rats' risk attitudes, we measured the *certainty equivalents* (CEs) for all gambles of 48 μL [2, 28, 29]. The certainty equivalent is the guaranteed reward the rat deems equal to the gamble (Figures 2I and 2J). If it is less than the gamble's expected value, that indicates risk aversion: the subject effectively "undervalues" the gamble and will accept a smaller reward to avoid risk (Figure 2K). Conversely, if the certainty equivalent is greater than the gamble's expected value, the subject is risk seeking, and risk neutral if they are equal. Measured certainty equivalents closely matched those predicted from the model, using an analytic expression incorporating utility and probability weighting functions ($CE = w(p)^{1/\alpha}$; STAR Methods. Pearson's correlation 0.96, p = 1.58e−11; Figure 2K). This non-parametric assay further validated the model fits and revealed heterogeneous risk preferences across rats (Figure 2L; Figures S3A–S3C).

Although rats exhibited nonlinear utility and probability weighting, consistent with prospect theory, they also exhibited trial-by-trial learning, consistent with reinforcement learning. Fitting the model to trials following rewarded and unrewarded choices revealed systematic shifts in utility and probability weighting functions: utility functions became less concave and probability weighting functions became more elevated to reflect increased likelihood of risky choice following rewards (Figure 3A). This was consistent across rats, as observed in certainty equivalents from the data (p = 6.49e−5, paired t test) and model (Figure 3B; p = 2.46e−7).

Another feature of human behavior is "reference dependence": people evaluate rewards as gains or losses relative to an internal reference point. It is unclear what determines the reference point [30]; proposals include status quo wealth [1], reward expectation [31, 32], heuristics based on the prospects [33, 34], or recent experience [35, 36].

Rats demonstrated reference dependence by treating smaller rewards as losses. They exhibited win-stay and lose-switch biases: following unrewarded trials, rats were more likely to switch ports (Figure 3C). Surprisingly, most rats exhibited "switch" biases after receiving 6 or 12 μL, consistent with treating these outcomes as losses. The "win or lose" threshold (i.e., reference point) was experience dependent: a separate cohort of rats (n = 3) trained with doubled rewards (12–96 μL) exhibited lose-switch biases after receiving 12 or 24 μL (Figure 3D).

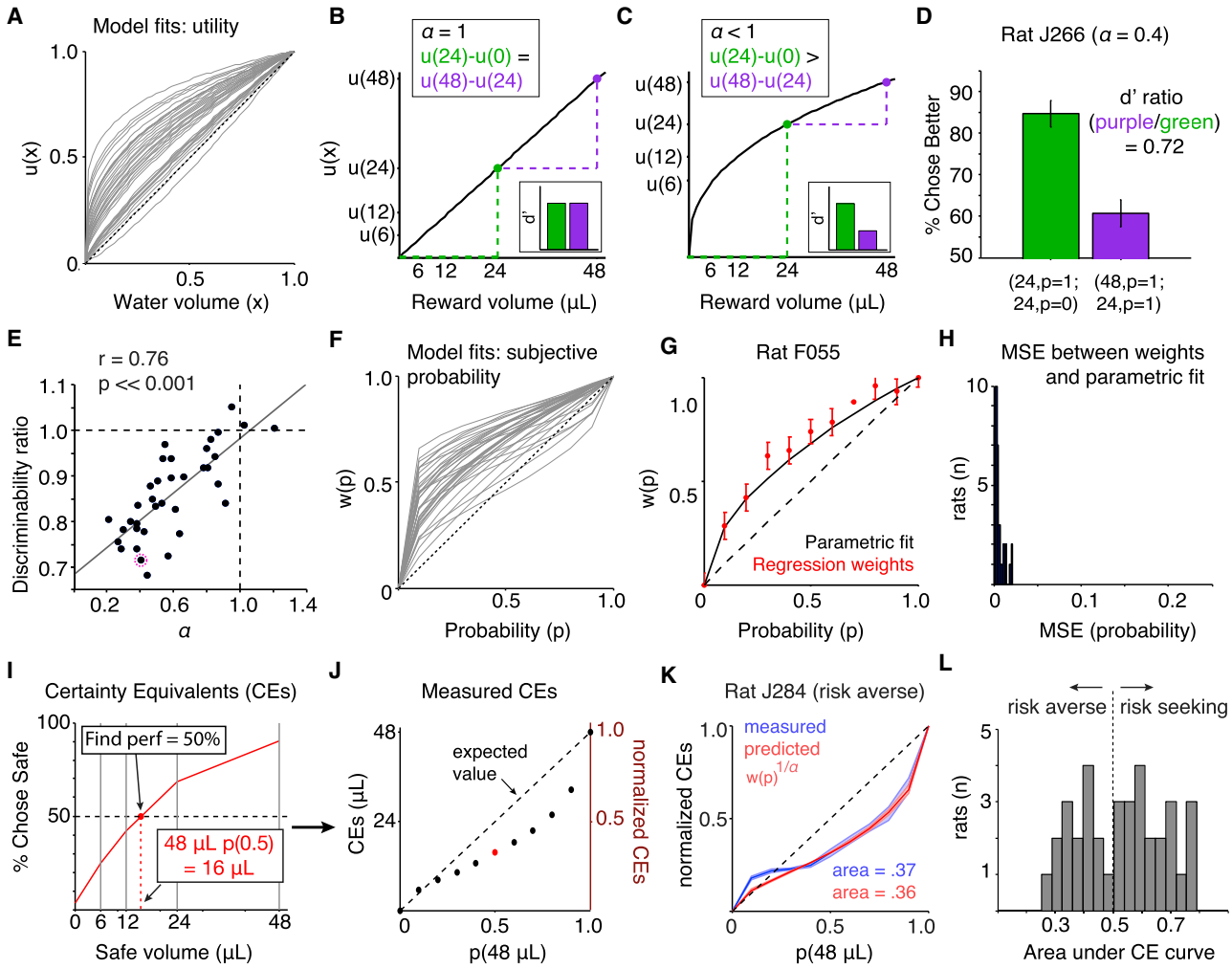

Figure 2. Non-parametric Analyses Confirm Nonlinear Utility and Probability Weighting and Reveal Diverse Risk Attitudes

(A) Model fits of subjective utility functions for each rat, normalized by the maximum volume (48 μL).

(B) Schematic linear utility function: the perceptual distance (or discriminability, d') between 0 and 24 μL is the same as 24 and 48 μL.

(C) Schematic concave utility function: 24 and 48 μL are less discriminable than 0 and 24 μL.

(D) One rat's performance on trials with guaranteed outcomes of 0 versus 24 μL (green) or 24 versus 48 μL (purple). Performance ratio on these trials ("d' ratio") less than 1 indicates diminishing sensitivity. Error bars are binomial confidence intervals.

(E) The concavity of the utility function (α) is significantly correlated with reduced discriminability of larger rewards ($p = 1.03e{-}7$, Pearson's correlation). Pink circle is rat from (D).

(F) Model fits of probability weighting functions.

(G) Weights from logistic regression parameterizing each probability match probability weighting function for one rat. Error bars are SEM for each regression coefficient.

(H) Mean squared error between regression weights and parametric fits for each rat (mean MSE = 0.006, in units of probability).

(I and J) To obtain certainty equivalents, we measured psychometric functions for each probability of receiving 48 μL and estimated the certain volume at which performance = 50%.

(K) Measured (blue) and model-predicted (red) certainty equivalents from one rat indicates systematic undervaluing of the gamble, or risk aversion. Error bars for model prediction are 95% confidence intervals of parameters from 5-fold cross validation. Data are mean ± SEM for left-out test sets.

(L) Distribution of certainty equivalent areas computed using analytic expression from model fits. Measured certainty equivalents were similar (Figure S3C).

See also Figures S2 and S3.

The win or lose threshold was often reward-history dependent (Figure 3E). Therefore, we parameterized a reference point, r, as taking one of two values depending on whether the previous trial was rewarded (see STAR Methods). Rewards less than r were negative (losses). The relative amplitude of losses versus gains was controlled by the parameter κ (Equation 1; Figure 1D; Box 1). Subjective value was reparameterized to include the zero outcome of the gamble, which is a loss when $r > 0$:

$$V_R = u(x_R)w(p_R) + u(0)w(1 - p_R) \quad (8)$$

$$V_L = u(x_L)w(p_L) + u(0)w(1 - p_L) \quad (9)$$

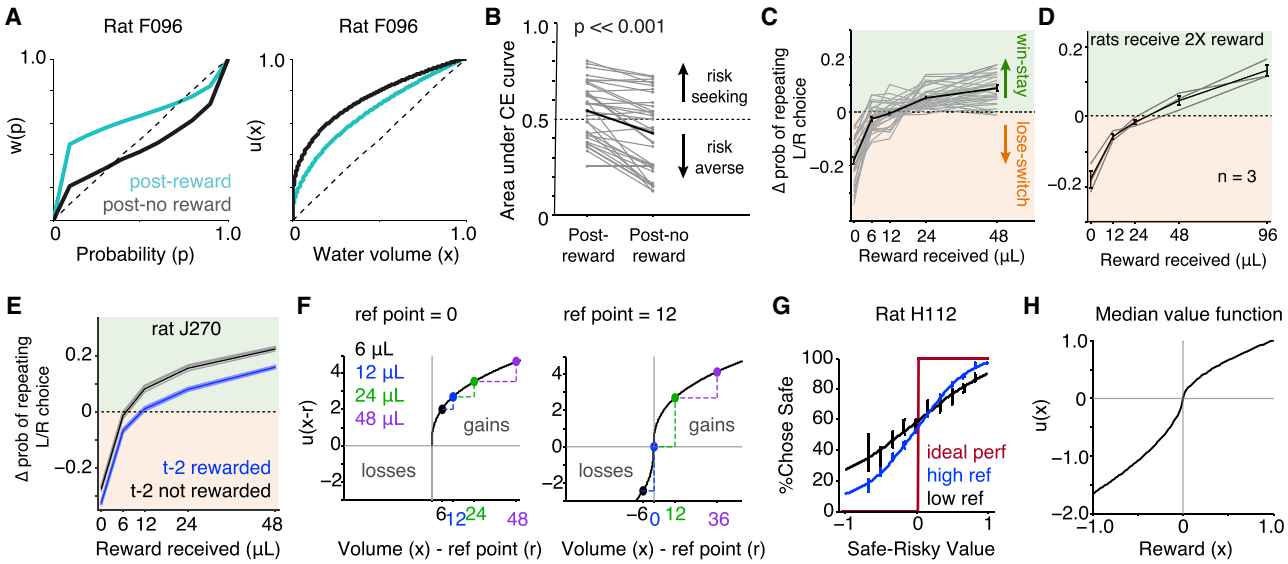

Figure 3. Rats Exhibit Evidence of Trial-by-Trial Learning

(A) Probability weighting function (left) and utility function (right) for one rat from model fit to trials following reward (turquoise) or no reward (black).

(B) Certainty equivalent areas predicted from model fits for all rats following rewarded and unrewarded trials ($p = 2.46e{-}7$, paired t test).

(C) ΔProbability of repeating left or right choices (relative to mean probability of repeating), following each reward. Points above the dashed line indicate an increased probability of repeating ("stay"); those below indicate a decreased probability ("switch"). Black curve is average ±SEM across rats.

(D) A separate cohort of 3 rats was trained with doubled water volumes. They exhibited lose-switch biases following 12 and 24 μL.

(E) Win-stay and lose-switch biases for one rat separated by reward history two trials back.

(F) Schematic illustrating that with concave utility, rewards should be more (less) discriminable when the reference point is high (low).

(G) Psychometric performance from one rat when the inferred reference point was low (black) or high (blue). Red curve is ideal performance.

(H) Value function with the median parameters across rats indicates loss aversion (median $\alpha = 0.6$, $\kappa = 1.7$).

See also Figure S3.

Model comparison (Akaike information criterion, AIC) favored the reference point model for all rats (Figures S3D and S3E). We also parameterized the reference point as reflecting several trials back with an exponential decay, where the time constant was a free parameter (see STAR Methods). For most rats (20/36, 77%), this did not significantly improve model performance compared to the reference point reflecting one trial back, although it was a better fit for a minority of rats with longer integration time constants over trials (Figures S3F–S3H). For the sake of simplicity and because it generally provided a better fit, we focused on the "one-trial back" reference point model. Interestingly, the reference point from the model was not significantly correlated with average reward rate for each rat [37]; this was true regardless of whether opt-out trials were included in estimates of average reward per trial (Pearson's correlation, $p > 0.05$).

With concave utility, rats should exhibit sharper psychometric performance when the reference point is high (and rewards are more discriminable; Figure 3F). Indeed, performance was closer to ideal when the reference point was high (Figure 3G; mean squared error [MSE] between psychometric and ideal performance was 0.143 *low ref* versus 0.122 *high ref*, $p = 3.4e{-}5$, paired t test across rats).

A loss parameter $\kappa > 1$ indicates "loss aversion" or a greater sensitivity to losses than gains (Equation 1). We observed a median κ of 1.66 (Figure 3H). There was variability across rats: 16/36 rats (44%) were not loss averse but were more sensitive to gains ($\kappa < 1$). Still, the median κ across rats suggests similarity to humans (Figure 3H).

Prospect theory does not account for how agents learn subjective values from experience, and we explicitly incorporated trial history parameters to account for trial-by-trial learning (Equations 6 and 7, Figure 4A). To examine learning dynamics, we fit the model to the first and second half of all trials, once rats achieved criterion performance (Figure S4A). There was no significant change in the parameters for the utility or probability weighting functions (Figure S4B). Rats showed a significant increase in the softmax parameter with training, indicating increased sensitivity to value differences, and a decrease in one of the trial history parameters, h_1, indicating reduced win-stay biases (Figure S4).

Reinforcement learning describes an adaptive process in which animals learn the value of states and actions. This framework, however, implies linear utility and probability weighting. We simulated choices of a Q-learning agent, which learned the state-action values of each unique trial type. Fitting the prospect theory model to these simulated choices recovered linear utility and probability weighting functions (regardless of learning rate; Figure 4B). This is expected: since trial sequences were randomized (i.e., each trial was independent), basic reinforcement learning will learn the expected value of each option [6] and will resolve to linear utility and probability weighting functions. We therefore implemented a reinforcement learning model that could accommodate nonlinear subjective utility and probability

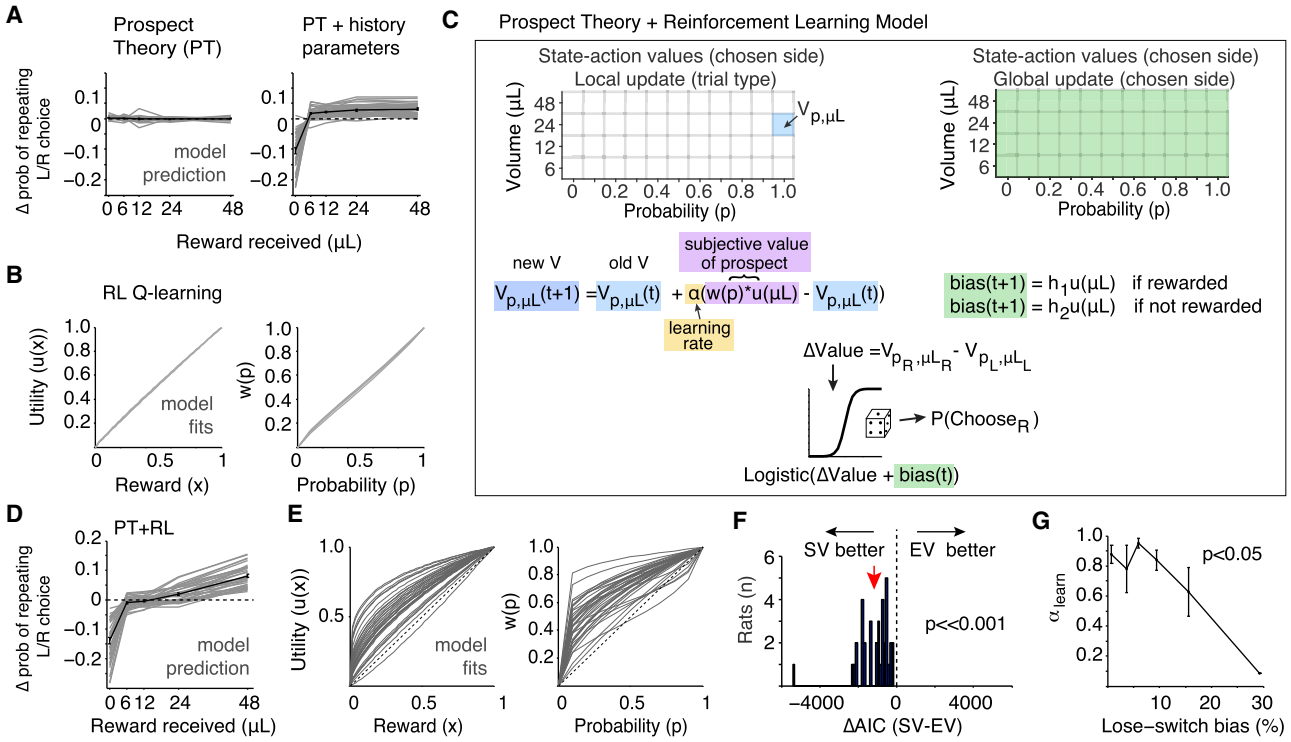

Figure 4. Integrating Prospect Theory and Reinforcement Learning Captures Nonlinear Subjective Functions and Learning

(A) Prospect theory model predictions for each rat, without the trial history parameters (h_1–h_3, see Box 1) does not account for win-stay and lose-switch trial history effects. Inclusion of these parameters accounts for these effects.

(B) Prospect theory model fit to simulated choices from a basic reinforcement learning agent yields linear utility and probability weighting functions over a range of generative learning rates (0.2, 0.4, 0.6, 0.8, 1.0, overlaid).

(C) Schematic of model incorporating prospect theory and reinforcement learning.

(D) The hybrid model described in (C) accounts for win-stay and lose-switch effects.

(E) The model recovers nonlinear utility and probability weighting functions.

(F) Model comparison when the error term used in the model was the subjective value (as shown in C) or the expected value (probability × reward). Red arrow is mean ΔAIC (p = 1.38e−8, paired t test of AIC).

(G) Binned values of rats' lose-switch biases (measured from the data) plotted against the best-fit learning rate, α_{learn}. Pearson's correlation coefficient is −0.37 across rats (p = 0.026).

See also Figure S4.

weighting but also learning over trials (Figure 4C [10, 38]). The model assumed that the rats learned the value of left and right choices for each unique combination of probability (p) and reward (μL) according to the following equation:

$$V_{p,\mu L}(t+1) = V_{p,\mu L}(t) + \alpha_{learn}(w(p)u(x) - V_{p,\mu L}(t)) \qquad (10)$$

where α_{learn} is the learning rate parameter, and $w(p)$ and $u(x)$ are parameterized as in Equations 2 and 3. The learned values of the right and left prospects on each trial were transformed into choice probabilities via a logistic function (see STAR Methods). We also implemented a global bias for left and right choices depending on reward history (Figures 4C and 4D). In this model, utility and probability weighting functions were exclusively used for learning or updating values, whereas *choice* depended on the learned values on each trial. Although the parameters of the utility and probability weighting functions were free parameters in this model, we recovered parameter values identical to the prospect theory model (Figure 4E; Figure S4C). Importantly, a reinforcement learning model in which the expected value (EV) was the error signal driving learning underperformed compared to the model incorporating subjective value according to prospect theory (Figure 4F; p = 1.38e−8, paired t test of AIC). Finally, each rats' learning rate (α_{learn}) was negatively correlated with the magnitude of their lose-swich biases, suggesting an inverse relationship between learning dynamics governing gradual improvements in task performance, and trial-by-trial learning, which is deleterious to performance when trials are independent [27, 39, 40]. Rats with slower dynamics (lower learning rate) showed more prominent trial history effects, whereas rats with rapid learning showed reduced trial history biases (Figure 4G).

DISCUSSION

There is a strong foundation for using animal models to study the cost-benefit calculations underlying economic theory [41]. In foraging studies, animals either exploit a current option for reward or explore a new one and often maximize their rate of reward [42–45]. Rodents exhibit complex aspects of economic decision-making, including regret [46, 47] and sensitivity to sunk costs [48, 49]. Here, we applied prospect theory to rats.

Like humans, rats exhibit nonlinear concave utility for gains, probability distortion, reference dependence, and, frequently, loss aversion. Nearly all rats exhibited concave utility, which produced diminishing marginal sensitivity. In contrast, most studies in monkeys have reported convex utility [24, 50–53] (but see [25]). In Expected Utility Theory, concave utility indicates risk aversion [26]. However, in prospect theory, concave utility can coincide with risk-seeking behavior due to the elevation of the probability weighting function [28, 29].

Our rats differ from primates in that they do not appear to underweight moderate and high probabilities [1, 24, 28]. The "inverted-S" shape of the probability weighting function may reflect diminishing sensitivity relative to two reference points that, in the probability domain, correspond to 0 and 1 [2, 28]. The rats, either due to the task or species differences, may not treat certainty as a reference point.

Reward history modified rats' risk preferences, producing shifts in utility and probability weighting functions. The extent to which risk preferences are stable traits is an area of active research [54, 55]. Recent work suggests a general or stable component of risk attitudes, but variability across domains (e.g., finance, recreation [55]). In foraging tasks, risk preferences reflect food availability and/or energy budget in a variety of species [44, 56]. Here, we document dynamic risk preferences; these trial-by-trial dynamics are not likely driven by physiological factors (e.g., energy budget) but may reflect dynamic internal or cognitive states mediated by reinforcement learning.

We found evidence for reference dependence, in which rats' treatment of outcomes as gains or losses reflected their reward history. Studies in several species, including capuchin monkeys [57], have also suggested reference dependence. Starlings and locusts prefer options that previously were encountered under greater hunger, presumably because those rewards were perceived to have greater reference-dependent value [58–60]. Rats modulate their approach speed for rewards depending on previously experienced reward amounts [61, 62]. Regret, which reflects a post-decision valuation of a choice relative to an unchosen alternative, may be a reference-dependent computation; for regret, the reference point would be the counterfactual prospect [63].

While variable, the median loss parameter (κ) across rats indicated loss aversion, which has been documented in capuchin monkeys [64]. We note that we did not examine losses by taking reinforcers away from the animal. However, several decision theories [32, 65] posit that rewards less than the reference point are losses. Loss aversion in humans is remarkably variable [66] and possibly domain specific [67, 68]. The nature of loss aversion is intriguing: is it a constant, a psychological trait similar to risk preferences [55], or an emergent property of constructing preferences [69]?

Prospect theory, animal learning theory, and reinforcement learning are complementary frameworks for studying decision-making (but see [70]). Reinforcement learning and animal learning theory are principally concerned with how subjects learn values over experience and use those learned values to make decisions. Prospect theory, in contrast, does not address learning but *describes* nonlinear distortions that account for the decision. We propose a simple approach for integrating these frameworks, in which animals learn the values of actions associated with task states, but the reward prediction error driving learning is in units of subjective value according to prospect theory [10, 38]. This hypothesis is consistent with studies of dopamine neurons, which are thought to instantiate reward prediction errors in the brain in a temporal-difference learning algorithm [6, 7, 13, 14]. Conditioned stimuli predicting rewards with different probabilities or magnitudes have been shown to elicit phasic dopamine responses reflecting the value of the expected reward [71–73]. In delay discounting tasks, the phasic dopamine response reflects discounted value of delayed rewards in monkeys and rats [74, 75]. Finally, recent work has shown that dopamine reward prediction errors reflect the shape of monkeys' measured utility functions [76]. The hypothesis of a reward prediction error in units of subjective value (perhaps according to prospect theory in the case of explicitly described lotteries) is also conceptually related to studies of homeostatic reinforcement learning, in which internal state influences subjective valuation [77]. This hypothesis bridges animal learning theory, reinforcement learning, and economic concepts of subjective value. A key topic of future research should address how subjective estimates of value arise: are they innate, learned early in life, or constantly evolving over the lifespan?

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subjects
- METHOD DETAILS
  - Behavioral training
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral model
  - Alternative models
  - Psychometric curves
  - Logistic regression to compare regressors to probability weighting functions
  - Certainty equivalents
  - Behavioral model with reference point
  - Behavioral model integrating Prospect Theory and Reinforcement Learning
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.05.013.

ACKNOWLEDGMENTS

The authors thank Paul Glimcher, Kenway Louie, Mike Long, Cristina Savin, David Schneider, Kevin Miller, Ben Scott, Mikio Aoi, Matthew Lovett-Barron, Cristina Domnisoru, Alejandro Ramirez, and members of the Brody lab for helpful discussions and comments on the manuscript. We thank J. Teran, K. Osorio, L. Teachen, and A. Sirko for animal training. This work was funded in part by a K99/R00 award from NIMH (MH111926 to C.M.C.).

AUTHOR CONTRIBUTIONS

All authors provided feedback on analyses and the manuscript. C.M.C. designed and performed all experiments, analyzed the data, and wrote the initial draft of the paper. A.T.P. and C.D.B. provided guidance for modeling and analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 7, 2018
Revised: March 6, 2019
Accepted: May 1, 2019
Published: May 30, 2019

REFERENCES

1. Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica *47*, 263.
2. Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. J. Risk Uncertain. *5*, 297–323.
3. Bush, R.R., and Mosteller, F. (2006). A mathematical model for simple learning. In Selected Papers of Frederick Mosteller, S.E. Fienberg, and D.C. Hoaglin, eds. (Springer), pp. 221–234.
4. Bush, R.R., and Mosteller, F. (2006). A model for stimulus generalization and discrimination. In Selected Papers of Frederick Mosteller, S.E. Fienberg, and D.C. Hoaglin, eds. (Springer), pp. 235–250.
5. Rescorla, R.A., and Wagner, A.R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black, and W.F. Prokasy, eds. (Appleton-Century-Crofts), pp. 64–99.
6. Glimcher, P.W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. USA *108* (*Suppl 3*), 15647–15654.
7. Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: An Introduction (The MIT Press).
8. Thorndike, E.L. (1911). Animal Intelligence: Experimental Studies (The Macmillan Company).
9. Pavlov, P.I. (2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. Ann. Neurosci. *17*, 136–141.
10. Glimcher, P.W. (2010). Foundations of Neuroeconomic Analysis (Oxford University Press).
11. Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.
12. Lee, D., Seo, H., and Jung, M.W. (2012). Neural basis of reinforcement learning and decision making. Annu. Rev. Neurosci. *35*, 287–308.
13. Niv, Y. (2009). Reinforcement learning in the brain. J. Math. Psychol. *53*, 139–154.
14. Bornstein, A.M., and Daw, N.D. (2011). Multiplicity of control in the basal ganglia: Computational roles of striatal subregions. Curr. Opin. Neurobiol. *21*, 374–380.
15. van der Meer, M.A.A., and Redish, A.D. (2011). Ventral striatum: a critical look at models of learning and evaluation. Curr. Opin. Neurobiol. *21*, 387–392.
16. Averbeck, B.B., and Costa, V.D. (2017). Motivational neural circuits underlying reinforcement learning. Nat. Neurosci. *20*, 505–512.
17. Ito, M., and Doya, K. (2011). Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. Curr. Opin. Neurobiol. *21*, 368–373.
18. Hertwig, R., and Erev, I. (2009). The description-experience gap in risky choice. Trends Cogn. Sci. *13*, 517–523.
19. Barron, G., and Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. J. Behav. Decis. Making *16*, 215–233.
20. Erev, I., and Roth, A.E. (2014). Maximization, learning, and economic behavior. Proc. Natl. Acad. Sci. USA *111* (*Suppl 3*), 10818–10825.
21. Herrnstein, R.J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. J. Exp. Anal. Behav. *4*, 267–272.
22. Heyman, G.M., and Duncan Luce, R. (1979). Operant matching is not a logical consequence of maximizing reinforcement rate. Anim. Learn. Behav. *7*, 133–140.
23. Gallistel, C.R., Mark, T.A., King, A.P., and Latham, P.E. (2001). The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. J. Exp. Psychol. Anim. Behav. Process. *27*, 354–372.
24. Stauffer, W.R., Lak, A., Bossaerts, P., and Schultz, W. (2015). Economic choices reveal probability distortion in macaque monkeys. J. Neurosci. *35*, 3146–3154.
25. Yamada, H., Tymula, A., Louie, K., and Glimcher, P.W. (2013). Thirst-dependent risk preferences in monkeys identify a primitive form of wealth. Proc. Natl. Acad. Sci. USA *110*, 15788–15793.
26. von Neumann, J., and Morgenstern, O. (2007). Theory of Games and Economic Behavior (Princeton University Press).
27. Scott, B.B., Constantinople, C.M., Erlich, J.C., Tank, D.W., and Brody, C.D. (2015). Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats. eLife *4*, e11308.
28. Gonzalez, R., and Wu, G. (1999). On the shape of the probability weighting function. Cognit. Psychol. *38*, 129–166.
29. Abdellaoui, M., Bleichrodt, H., and L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. J. Risk Uncertainty *36*, 245.
30. Barberis, N. (2013). Thirty years of prospect theory in economics: A review and assessment J. Econ. Perspect *27*, 173–196.
31. Köszegi, B., and Rabin, M. (2007). Reference-dependent risk attitudes. Am. Econ. Rev. *97*, 1047–1073.
32. Köszegi, B., and Rabin, M. (2006). A model of reference-dependent preferences. Q. J. Econ. *121*, 1133–1165.
33. Bleichrodt, H., Pinto, J.L., and Wakker, P.P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. Manage. Sci. *47*, 1498–1514.
34. van Osch, S.M.C., van den Hout, W.B., and Stiggelbout, A.M. (2006). Exploring the reference point in prospect theory: gambles for length of life. Med. Decis. Making *26*, 338–346.
35. Khaw, M.W., Glimcher, P.W., and Louie, K. (2017). Normalized value coding explains dynamic adaptation in the human valuation process. Proc. Natl. Acad. Sci. USA *114*, 12696–12701.
36. Hunter, L.E., and Gershman, S.J. (2018). Reference-dependent preferences arise from structure learning. bioRxiv. https://doi.org/10.1101/252692.
37. Constantino, S.M., and Daw, N.D. (2015). Learning the opportunity cost of time in a patch-foraging task. Cogn. Affect. Behav. Neurosci. *15*, 837–853.
38. Niv, Y., Edlund, J.A., Dayan, P., and O'Doherty, J.P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J. Neurosci. *32*, 551–562.
39. Akrami, A., Kopec, C.D., Diamond, M.E., and Brody, C.D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature *554*, 368–372.
40. Busse, L., Ayaz, A., Dhruv, N.T., Katzner, S., Saleem, A.B., Schölvinck, M.L., Zaharia, A.D., and Carandini, M. (2011). The detection of visual contrast in the behaving mouse. J. Neurosci. *31*, 11351–11361.
41. Kagel, J.H., Battalio, R.C., and Green, L. (1995). Economic Choice Theory: An Experimental Analysis of Animal Behavior (Cambridge University Press).
42. Charnov, E.L. (1976). Optimal foraging, the marginal value theorem. Theor. Popul. Biol. *9*, 129–136.

43. Kacelnik, A. (1984). Central place foraging in starlings (Sturnus vulgaris). I. Patch residence time. J. Anim. Ecol. *53*, 283–299.
44. Stephens, D.W., and Krebs, J.R. (1986). Foraging Theory (Princeton University Press).
45. Ollason, J.G. (1980). Learning to forage—optimally? Theor. Popul. Biol. *18*, 44–56.
46. Steiner, A.P., and Redish, A.D. (2014). Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. Nat. Neurosci. *17*, 995–1002.
47. Sweis, B.M., Thomas, M.J., and Redish, A.D. (2018). Mice learn to avoid regret. PLoS Biol. *16*, e2005853.
48. Sweis, B.M., Abram, S.V., Schmidt, B.J., Seeland, K.D., MacDonald, A.W., 3rd, Thomas, M.J., and Redish, A.D. (2018). Sensitivity to "sunk costs" in mice, rats, and humans. Science *361*, 178–181.
49. Wikenheiser, A.M., and David Redish, A. (2012). Sunk costs account for rats' decisions on an intertemporal foraging task. BMC Neurosci. *13*, 63.
50. McCoy, A.N., and Platt, M.L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. Nat. Neurosci. *8*, 1220–1227.
51. Hayden, B.Y., and Platt, M.L. (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. Curr. Biol. *17*, 49–53.
52. So, N.Y., and Stuphorn, V. (2010). Supplementary eye field encodes option and action value for saccades with variable reward. J. Neurophysiol. *104*, 2634–2653.
53. Chen, X., and Stuphorn, V. (2018). Inactivation of medial frontal cortex changes risk preference. Curr. Biol. *28*, 3709.
54. Schildberg-Hörisch, H. (2018). Are risk preferences stable? J. Econ. Perspect. *32*, 135–154.
55. Frey, R., Pedroni, A., Mata, R., Rieskamp, J., and Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. Sci. Adv. *3*, e1701381.
56. Kacelnik, A., and Bateson, M. (1996). Risky theories—The effects of variance on foraging decisions. Am. Zool. *36*, 402–434.
57. Lakshminarayanan, V.R., Keith Chen, M., and Santos, L.R. (2011). The evolution of decision-making under risk: Framing effects in monkey risk preferences. J. Exp. Soc. Psychol. *47*, 689–693.
58. Pompilio, L., and Kacelnik, A. (2005). State-dependent learning and suboptimal choice: when starlings prefer long over short delays to food. Anim. Behav. *70*, 571–578.
59. Pompilio, L., Kacelnik, A., and Behmer, S.T. (2006). State-dependent learned valuation drives choice in an invertebrate. Science *311*, 1613–1615.
60. Marsh, B. (2004). Energetic state during learning affects foraging choices in starlings. Behav. Ecol. *15*, 396–399.
61. Crespi, L.P. (1942). Quantitative variation of incentive and performance in the white rat. Am. J. Psychol. *55*, 467.
62. Zeaman, D. (1949). Response latency as a function of the amount of reinforcement. J. Exp. Psychol. *39*, 466–483.
63. Krähmer, D., and Stone, R. (2011). Anticipated regret as an explanation of uncertainty aversion. Econom. Theory *52*, 709–728.
64. Chen, M.K., Keith Chen, M., Lakshminarayanan, V., and Santos, L.R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. J. Polit. Econ. *114*, 517–537.
65. Gul, F. (1991). A theory of disappointment aversion. Econometrica *59*, 667–686.
66. Sayman, S., and Öncüler, A. (2005). Effects of study design characteristics on the WTA–WTP disparity: A meta analytical framework. J. Econ. Psychol. *26*, 289–312.
67. Dhar, R., and Wertenbroch, K. (1999). Consumer choice between hedonic and utilitarian goods. J. Marketing Res *XXXVII*, 60–71.
68. Heath, T.B., Ryu, G., Chatterjee, S., McCarthy, M.S., Mothersbaugh, D.L., Milberg, S., and Gaeth, G.J. (2000). Asymmetric competition in choice and the leveraging of competitive disadvantages. J. Consum. Res. *27*, 291–308.
69. Johnson, E.J., Gächter, S., and Herrmann, A. (2006). Exploring the Nature of Loss Aversion. IZA Discussion Papers 2015. Institute for the Study of Labor (IZA).
70. Plonsky, O., and Erev, I. (2017). Learning in settings with partial feedback and the wavy recency effect of rare events. Cognit. Psychol. *93*, 18–43.
71. Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science *299*, 1898–1902.
72. Morris, G., Arkadir, D., Nevet, A., Vaadia, E., and Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. Neuron *43*, 133–143.
73. Tobler, P.N., Fiorillo, C.D., and Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. Science *307*, 1642–1645.
74. Kobayashi, S., and Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. J. Neurosci. *28*, 7837–7846.
75. Day, J.J., Jones, J.L., Wightman, R.M., and Carelli, R.M. (2010). Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. Biol. Psychiatry *68*, 306–309.
76. Stauffer, W.R., Lak, A., and Schultz, W. (2014). Dopamine reward prediction error responses reflect marginal utility. Curr. Biol. *24*, 2491–2500.
77. Keramati, M., and Gutkin, B. (2014). Homeostatic reinforcement learning for integrating reward collection and physiological stability. eLife *3*.. Published online December 2, 2014. 10.7554/eLife.04811.
78. Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science *340*, 95–98.
79. Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., and Brody, C.D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature *520*, 220–223.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Rat: Long Evans	Taconic	RRID: RGD_1566430
Rat: Sprague Dawley	Taconic	RRID: RGD_1566440
Rat: Long Evans	Hilltop	http://www.hilltoplabs.com/
Rat: Long Evans	Harlan	RRID: RGD_5508398
Rat: Pvalb-iCre	University of Missouri RRRC	RRID: RGD_10412329
Software and Algorithms		
MATLAB	MathWorks	RRID: SCR_001622
Behavioral control software	Bcontrol	http://brodywiki.princeton.edu/bcontrol/index.php/Main_Page

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Christine Constantinople (constantinople@nyu.edu). Transgenic (Pvalb-iCre)2Ottc rats (n = 5) were obtained by an MTA from University of Missouri RRRC.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

A total of 39 male rats between the ages of 6 and 24 months were used for this study, including 35 Long-evans and 4 Sprague-Dawley rats (*Rattus norvegicus*). The Long-evans cohort also included LE-Tg (Pvalb-iCre)2Ottc rats (n = 5) made at NIDA/NIMH and obtained from the University of Missouri RRRC (transgenic line 0773). These are BAC transgenic rats expressing Cre recombinase in parvalbumin expressing neurons. Animal use procedures were approved by the Princeton University Institutional Animal Care and Use Committee (IACUC #1853) and carried out in accordance with National Institutes of Health standards.

Rats were typically housed in pairs or singly; rats that trained during the day were housed in a reverse light cycle room. Some rats trained overnight, and were not housed with a reverse light cycle. Access to water was scheduled to within-box training, 2-4 hours per day, usually 7 days a week, and between 0 and 1 hour ad lib following training.

METHOD DETAILS

Behavioral training

Rats were trained in a high-throughput facility using a computerized training protocol. Rats were trained in operant training boxes with three nose ports. When an LED from the center port was illuminated, the animal could initiate a trial by poking his nose in that port; upon trial initiation the center LED turned off. While in the center port, rats were continuously presented with a train of randomly timed clicks from a left speaker and, simultaneously, a different train of clicks from a right speaker. The click trains were generated by Poisson processes with different underlying rates [78, 79]; the rates conveyed the water volume baited at each side port. After a variable pre-flash interval ranging from 0 to 350ms, rats were also presented with light flashes from the left and right side ports; the number of flashes conveyed reward probability at each port. Each flash was 20ms in duration; flashes were presented in fixed bins, spaced every 250ms, to avoid perceptual fusion of consecutive flashes [27]. After a variable post-flash delay period from 0 to 500ms, the end of the trial was cued by a go sound and the center LED turning back on. The animal was then free to choose the left or right center port, and potentially collect reward.

The trials were self-paced: on trials when rats did not receive reward, they were able to initiate another trial immediately. However, if rats terminated center fixation prematurely, they were penalized with a white noise sound and a time out penalty. Since rats disproportionately terminated trials offering low volumes, we scaled the time out penalty based on the minimum reward offered. The time out penalties were adjusted independently for each rat to minimize terminated trials (as an example, several rats were penalized with 6 s time-outs for terminating trials offering a minimum of 6μL, 4.5 s for terminating trials offering a minimum of 12μL, 3 s for terminating trials offering a minimum of 24μL, and 1.5 s for terminating trials offering a minimum of 48μL).

In this task, the rats were required to reveal their preference between safe and risky rewards. To determine when rats were sufficiently trained to understand the meaning of the cues in the task, we evaluated the "efficiency" of their choices as follows. For each training session, we computed the average expected value per trial of an agent that chose randomly, and a perfect expected value maximizer, or an agent that always chose the side with the greater expected value. We compared the expected value per trial from the rat's choices relative to these lower and upper bounds. Specifically, the efficiency was calculated as follows:

$$efficiency = 0.5\frac{rat_{EV/trial} - rand_{EV/trial}}{EV\ max_{EV/trial} - rand_{EV/trial}} + 0.5 \tag{11}$$

The threshold for analysis was the median performance of all sessions minus 1.5 times the interquartile range of performance across the second half of all sessions. Once performance surpassed this threshold, it was typically stable across months. Occasional days with poor performance were usually due to hardware malfunctions in the rig. Days in which performance was below threshold were excluded from analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral model

We fit a behavioral model separately for each rat (see Box 1 for description of the model). We used MATLAB's constrained minimization function fmincon to minimize the sum of the negative log likelihoods with respect to the model parameters. 20 random seeds were used in the maximum likelihood search for each rat; parameter values with the maximum likelihood of these seeds were deemed the best fit parameters. When evaluating model performance (e.g., Figure 1E), we performed 5-fold cross-validation and evaluated the predictive power of the model on the held-out test sets.

We initially evaluated three different parametric forms of the probability weighting function, the one- and two-parameter Prelec models and the linear in log-odds model (see below) [24, 28]. We compared the different parametric forms using Akaike Information Criterion (AIC), $AIC = 2k + 2nLL$, where k is the number of parameters, and nLL is the negative log likelihood of the model. AIC favored the two-parameter Prelec model for nearly all rats, although some rats were equally well-fit by the linear in log-odds model (data not shown). Therefore, we implemented the two-parameter Prelec model.

$$\text{One-parameter Prelec}: \ w(p) = e^{-(ln(p))^{\delta}}, \tag{12}$$

where p is the true probability, and δ is a free parameter. δ controls the curvature of the weighting function; its crossover point is fixed at $1/e$.

$$\text{Two-parameter Prelec}: \ w(p) = e^{-\beta(-ln(p))^{\delta}}, \tag{13}$$

Where p is the true probability, β and δ are free parameters. δ primarily controls the curvature and β primarily controls the elevation of the weighting function.

$$\text{Linear in log-odds}: w(p) = \frac{\delta p^{r}}{\delta p^{r} + (1-p)^{r}}, \tag{14}$$

where p is the true probability and γ and δ are free parameters. γ primarily controls the curvature of the weighting function and δ controls the elevation.

Alternative models

We compared the prospect theory model to a number of alternative models. The Expected Utility Theory model (EUT) has the same form as the prospect theory model, except that the subjective value on each side is the product of objective probability and subjective utility (see Figure S2A):

$$V_R = u(x_R)p_R \tag{15}$$

$$V_L = u(x_L)p_L. \tag{16}$$

The linear weighting model fit different weights to flashes and clicks, before combining them (see Figure S1G):

$$u(x) = cx \tag{17}$$

$$w(p) = dp, \tag{18}$$

where c and d are constants, x is the click rate and p is the number of flashes presented to the animal on each side. The value on each side is the product of linearly weighted flashes and clicks:

$$V = u(x)w(p). \tag{19}$$

We next included sensory noise as part of a perceptual strategy. Previously, we have used a signal-detection theory (SDT) model to estimate rats' perceptual variability (noise) in estimating numbers of flashes and clicks; we found they exhibit a property called scalar variability, meaning that the standard deviation in their estimate grows linearly with the mean [27]. We implemented four different signal-detection theory models that instantiated scalar noise, according to this work. The models differ in the decision rules they apply. The models assume that on each trial, the rats' estimate of the number of flashes (clicks) on each side is a random variable drawn from a normal distribution, the mean of which corresponds to the actual number of flashes (clicks) presented to the animal. According to scalar variability, the standard deviation is linearly related to the number of flashes (clicks). There are two free parameters that define this linear relationship; we fit separate linear scaling relationships to the estimation of clicks and flashes:

$$\sigma_F = m_F x + b_F \tag{20}$$

$$\sigma_C = m_C x + b_C \tag{21}$$

Where m_F, m_C, b_F, b_C, are free parameters. x is the number of flashes (clicks) presented to the rat on each side. For the first two SDT models, we compute choice probabilities based on the flash difference (ΔF), and click difference (ΔC) separately, where these choice probabilities are calculated as follows, according to [27]:

$$P(\text{went right}|\Delta F) = \int_0^\infty N\left(R_F - L_F, \sqrt{\sigma_{R_F}^2 + \sigma_{L_F}^2}\right) d(R_F - L_F) \tag{22}$$

$$P(\text{went right}|\Delta C) = \int_0^\infty N\left(R_C - L_C, \sqrt{\sigma_{R_C}^2 + \sigma_{L_C}^2}\right) d(R_C - L_C) \tag{23}$$

$R_F(R_C)$ and $L_F(L_C)$ are the number of right and left flashes (clicks) presented to the rat on each trial, and the σ terms are the noise terms defined in Equations 20 and 21. One model (SDT_1) assumes that the rat's choice is given by the average of these probabilities (see Figure S2C). Another model (SDT_2) assumes that the rat's choice is given by the most informative cue on each trial (the choice probability most different from 0.5; see Figure S2D).

Alternatively, it's possible that the rats combine the noisy estimates of flashes and clicks on each side. Therefore, we evaluated two additional models parameterized as follows:

$$P(\text{went right}|\text{right ev}) = \int_0^\infty N\left(R_F + R_C, \sqrt{\sigma_{R_F}^2 + \sigma_{R_C}^2}\right) d(R_F + R_C) \tag{24}$$

$$P(\text{went right}|\text{left ev}) = -\int_0^\infty N\left(L_F + L_C, \sqrt{\sigma_{L_F}^2 + \sigma_{L_C}^2}\right) d(L_F + L_C) \tag{25}$$

One model (SDT_3) assumes that the rat's choice is given by the average of these probabilities (Figure S2E), and the other (SDT_4) assumes that the rat's choice is given by the most informative side on each trial (the choice probability most different from 0.5; Figure S2F).

Psychometric curves

We measured rats' psychometric performance when choosing between the safe and risky options. For these analyses, we excluded trials where both the left and right side ports offered certain rewards. We binned the data into 11 bins of the difference in the subjective value (inferred from the behavioral model) of the safe minus the risky option. Psychometric plots show the probability that the subjects chose the safe option as a function of this difference (see Figure S1D). We fit a 4-parameter sigmoid of the form:

$$P(\text{choose}_S) = y_0 + \frac{1 - 2a}{1 + e^{(-b(Vs - V_R - x_0))}}, \tag{26}$$

where y_0, a, b, and x_0 were free parameters. Parameters were fit using a gradient-descent algorithm to minimize the mean square error between the data and the sigmoid, using the sqp algorithm in MATLAB's constrained optimization function fmincon.

Logistic regression to compare regressors to probability weighting functions

We fit a logistic regression model with a separate regressor for each probability the rat may have been offered (0 to 1 in 0.1 increments), plus a constant term. To compare the regressors to the parametric fits, we normalized the regressors for each probability by subtracting the minimum and dividing by the maximum regressor value, so they ranged from 0 to 1 (Figure 2G). We computed the mean square error between these normalized regressor values and the probability weighting functions (Figure 2H). The model was fit using MATLAB's function glmfit.

Certainty equivalents

Non-parametric estimate

We estimated rat's certainty equivalents by evaluating their psychometric performance (%Chose risky) for each gamble of 48 μL, and estimating the value of the psychometric curve at which performance was at 50% (Figure 2I). To do this, we fit a line to the two points of the psychometric curve above and below chance level using MATLAB's regress.m function, and interpolated the value of that line that would correspond to 50%.

Analytic expression for CE from the model fits

We compared our estimates of rats' certainty equivalents from their behavioral data to an analytic expression from the subjective probability and utility functions we obtained from the model. We define the certainty equivalent, $\tilde{x}$, as the guaranteed reward equal to a gamble, x with probability p. In the case of linear probability weighting, we express this as follows:

$$\begin{aligned} px^{\alpha} &= \tilde{x}^{\alpha} \\ ln(p) + \alpha ln(x) &= \alpha ln(\tilde{x}) \\ ln(p) &= \alpha ln\left(\frac{\tilde{x}}{x}\right) \\ \frac{1}{\alpha} ln(p) &= ln\left(\frac{\tilde{x}}{x}\right) \\ p^{\frac{1}{\alpha}} &= \frac{\tilde{x}}{x} \end{aligned} \tag{27}$$

For nonlinear probability weighting, substituting $w(p)$ for p yields an analytic expression for the certainty equivalent from the exponent of the utility function (α) and the probability weighting function (also see [29]).

Behavioral model with reference point

The behavioral model with the reference point (see Figure 3) was similar to the behavioral model described above, except for elaborations of the subjective utility function $u(x)$ and subjective value (V_R, V_L). We modified the subjective utility function to include a dynamic reference point, r, below which value was treated negatively (as a loss). The relative amplitude of losses versus gains was controlled by the scale parameter κ.

$$u(x) = \begin{array}{ll} (x-r)^{\alpha} & \text{if} \quad x > r \\ -\kappa(r-x)^{\alpha} & \text{if} \quad x < r \end{array} \tag{28}$$

where, as before, α is the exponent of the utility function, and x is the offered reward. We also reparameterized subjective value. The risky prospect offers two possible outcomes: x with probability p, and 0 with probability $1-p$. In the absence of a reference point, the zero reward outcome $(0, 1-p)$ does not influence choice $(0^{\alpha} = 0)$. However, if $r > 0$, the zero reward outcome can be perceived as a loss. Therefore, in the reference point model, subjective value was reparameterized to incorporate this possible outcome of the gamble:

$$V_R = u(x_R)w(p_R) + u(0)w(1-p_R) \tag{29}$$

$$V_L = u(x_L)w(p_L) + u(0)w(1-p_L) \tag{30}$$

We parameterized the reference point, r, to take on two discrete values depending on whether the previous trial was rewarded or not. There were two additional free parameters, y and m that could account for asymmetric effects of rewarded and unrewarded trials:

$$r(t) = \begin{cases} m & \text{if t} - 1 \text{ was rewarded} \\ y & \text{if t} - 1 \text{ was not rewarded.} \end{cases} \tag{31}$$

We constrained $r > 0$.

Behavioral model integrating Prospect Theory and Reinforcement Learning

This behavioral model was similar to the prospect theory model, except that $w(p)$ and $u(x)$ were used to update the subject's value of each unique trial type based on experience. There were separate state-action value matrices for left and right choices. The entry of each matrix corresponded to a unique trial type, for each unique probability p and reward volume μL,

$$V_{p,\mu L}(t+1) = V_{p,\mu L}(t) + \alpha_{learn}(w(p)u(x) - V_{p,\mu L}(t)), \tag{32}$$

where α_{learn} is an additional free parameter fit by the model. $w(p)$ and $u(x)$ are parameterized as they were in the prospect theory model, according to Equations 2 and 3 in the main text.

We also included a global bias for the entire left or right value matrix that reflected reward history as follows:

$$\text{bias} = \begin{cases} \beta_{win} u(x) & \text{if t was rewarded} \\ \beta_{loss} u(x) & \text{if t was rewarded} \end{cases}, \quad (33)$$

where $u(x)$ corresponds to the subjective utility of the chosen reward volume. Choice probabilities were computed according to Equation 6 in the main text.

DATA AND SOFTWARE AVAILABILITY

Behavioral data are available upon request by contacting the Lead Contact, Christine Constantinople (constantinople@nyu.edu).