Neurocomputing 375 (2020) 32–42

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Neural mechanism underlying risk attitude and probability distortion: One two-stage model of valuation and choice

Hongzhi You[a], Mengya Zhang[b], Da-Hui Wang[b,c,*]

[a] Key Laboratory for NeuroInformation of Ministry of Education, Center for Information in BioMedicine School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China
[b] School of Systems Science, State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China
[c] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 5 April 2019
Revised 30 August 2019
Accepted 12 September 2019
Available online 24 September 2019

Communicated by Witali Dunin-Barkowski

Keywords:
Risk decision
Risk attitude
Probability distortion
Two-stage model

ABSTRACT

Although the neural mechanism underlying risk decision has been extensively investigated, the neural origination of risk attitude and probability distortion need to be further elucidated. In this study, the Rescorla–Wagner model with learning rates a_+/a_- upon gain/loss evaluates the risky outcome and forms the subjective values of risky options through the learning process, and the softmax function of subjective values produces the choice probability between options. Our model demonstrates that risk attitude is determined by the undervaluation/overvaluation of risky outcome, the standard deviation of the subjective value, and the discrimination ability between subjective value. Our model further displays that overweighting/underweighting of small probabilities results from asymmetric learning rates and the discrimination ability between subjective value. These findings suggest that risk attitude and probability distortion share a common neural mechanism.

1. Introduction

The risk attitude and the subjective probability are at the core of economic models of decision making under risk which deals with the trade-off between the higher expected reward of risky option and the sure reward at lower risk. The trade-off can be explicit as the subtraction of the product of the risk attitude and the risk from the mean of reward (return-risk model) [16], or implicit as the maximization of the sum of decision outcomes' utility weighted by subjective probability of the outcome in prospective theory [10]. Neural activities have been linked with the risk attitude, including single neuron activity in orbitofrontal cortex (OFC) [22] and the blood-oxygen-level-dependent (BOLD) signal in orbitofrontal cortex [29], lateral prefrontal cortex [31], anterior insula and nucleus accumbens [12], ventral striatum and anterior insula [23], dopaminergic regions and their targets [20],[32]. It has been shown that the risk attitude can be changed by the manipulation of cortical activity [3],[9],[11] or modulated by dopamine [26] and serotonin [37]. One study showed that the volume of gray matter in the right posterior parietal cortex predict the risk attitude [2]. Meanwhile, electrophysiology recording demonstrated that single neuron in OFC [33], basal ganglia [4] and anterodorsal region of the primate septum [19] may code the probability of reward and the probability distortion may relate with the activity of prefrontal cortex [30],[34] and striatum ([8,38]. However, the neural origination of risk attitude and the probability distortion has not been elucidated. Here, we investigated this issue using a two-stage evaluation and choice framework [7]. Rescorla–Wagner model [21],[28] with asymmetric learning rates upon gain and loss was applied to form the subjective values of risky options, while the probabilistic choice depended on the subjective values. The results of our model indicate that the risk attitude and the probability distortion share the common neural mechanism, i.e. choice based on the subjective value learned from the experience with asymmetric learning rates upon gain and loss. Furthermore, our model exhibits a reasonable behavior: seeking small risk but avoiding large risk.

* Corresponding author at: School of Systems Science, State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China.
E-mail address: wangdh@bnu.edu.cn (D.-H. Wang).

https://doi.org/10.1016/j.neucom.2019.09.021
0925-2312/

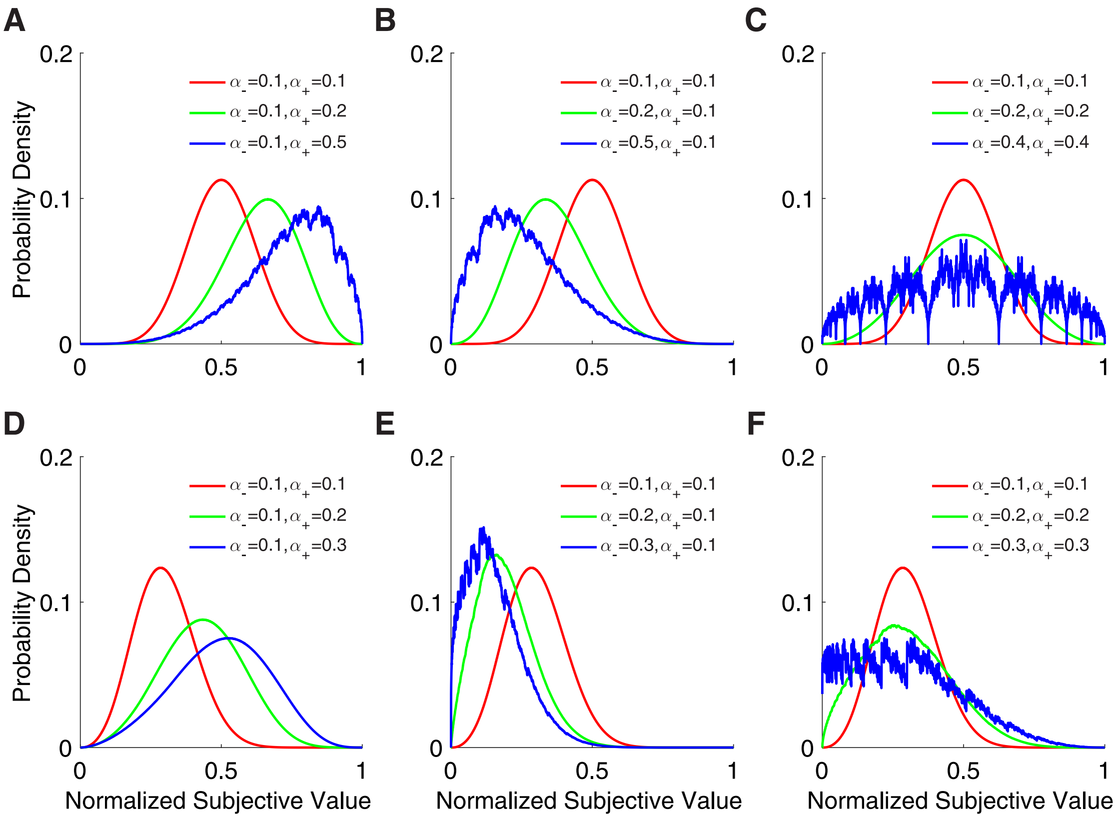


Fig. 1. Distribution of the normalized subjective value. (A) $\alpha_+ > \alpha_-$ and $p = 0.5$; (B) $\alpha_+ < \alpha_-$ and $p = 0.5$; (C) $\alpha_+ = \alpha_-$ and $p = 0.5$; (D) $\alpha_+ > \alpha_-$ and $p = 0.3$; (E) $\alpha_+ < \alpha_-$ and $p = 0.3$; (F) $\alpha_+ = \alpha_-$ and $p = 0.3$. The normalized subjective value is obtained by subtraction of R_- from subjective value and then divided by $R_+ - R_-$.

2. Results

2.1. Valuation process implemented as Rescorla–Wagner model with learning rates upon gain and loss

The valuation process is the first stage of risky decision. Based on the experience, subject learns the risky option's subjective value which can be expressed in physical units of the reference reward at points of equal preference or physical amount of safe reward at choice indifference [39]. If the outcome R_i of *ith* option deviates from the subjective value V_i, the subjective value will be updated according to Rescorla–Wagner model:

$$V_{i,t} = V_{i,t-1} + \alpha(R_{i,t} - V_{i,t-1}) \tag{1}$$

It has been shown that dopaminergic system and lateral habenula involved valuation process. On the one hand, dopaminergic neurons have a peak response to unexpected reward and a small dip response to unexpected loss [24]; on the other hand, some neurons in lateral habenula have a peak response to unexpected loss and a small dip response to unexpected reward [17]. The lateral habenula neurons project to serotonergic neurons in dorsal raphe nucleus and induce a neuromodulation different from that of dopamine [36]. Besides the separation of VTA and LHb, long-term potentiation(LTP) and long-term depotentiation(LTD) could result in an experience-dependent asymmetry effect on input [18]. Therefore, we split the learning process into two parts, one for the situation that outcome is larger than previous subjective value and one for the situation that outcome is smaller than previous subjective value. Thus, the update rule of subjective value can be rewritten as:

$$V_{i,t} = V_{i,t-1} + \begin{cases} \alpha_+(R_{i,t} - V_{i,t-1}) & if \quad R_{i,t} > V_{i,t-1} \\ \alpha_-(R_{i,t} - V_{i,t-1}) & if \quad R_{i,t} < V_{i,t-1} \end{cases} \tag{2}$$

where α_+ is the learning rate if the outcome is larger than previous subjective value and α_- is the learning rate if the outcome is smaller than previous subjective value. If $\alpha_+ = \alpha_-$, our update rule will be reduced to the original version of Rescorla–Wagner model.

We first explored features of the valuation process with separate learning rates α_+ and α_- given stochastic outcome for one risky option. For clarity, we considered one binary stochastic outcome, i.e., a larger reward R_+ at probability p and a smaller reward R_- at probability $1 - p$. The distribution of the subjective value of this risky option follows a dynamic:

$$P_t(V) = pP_{t+1}(\alpha_+R_+ + (1-\alpha_+)V) + (1-p)P_{t+1} \times (\alpha_-R_- + (1-\alpha_-)V) \tag{3}$$

where $P_t(V)$ is the probability that subjective value equals to V at time t. The steady distribution of subjective value can be obtained if we let $P_t(V) = P_{t+1}(V)$. The results in Fig. 1 demonstrate that: (1) the distribution of subjective value skews toward larger reward of the outcome given $\alpha_+ > \alpha_-$, which implies that the mean of the subjective value is larger than the average of the actual reward (Fig. 1A and 1D); (2) the distribution of the subjective value skews toward smaller reward of the outcome given $\alpha_+ < \alpha_-$, which means the average of the subjective value is smaller than the mean of the actual reward (Fig. 1B and 1E); (3) the distribution is smooth given smaller learning rate and non-smooth given larger learning rate; (4) the distribution is narrow given small learning rate but broad given large learning rate, indicating the standard deviation of subjective value, called as subjective risk (SR) in this study, increases with the learning rate (Fig. 1C). The mean of the subjective value (EV) can be worked out as:

$$EV = \frac{\alpha_+R_+p + \alpha_-R_-(1-p)}{\alpha_+p + \alpha_-(1-p)} \tag{4}$$

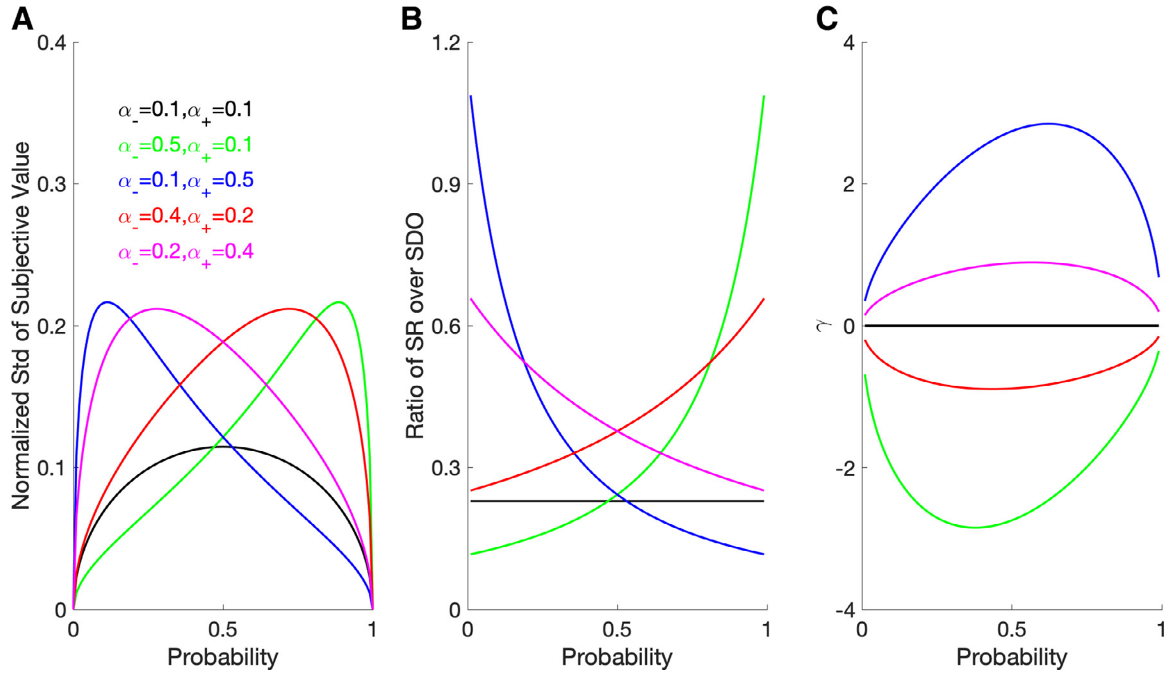


Fig. 2. The effects of learning rates on the subjective risk. (A) The subjective risk normalized by $R_+ - R_-$ as function of probability of receiving reward R_+. (B) The ratio of subjective risk over objective risk $(R_+ - R_-)\sqrt{p(1-p)}$ as function of probability of receiving reward R_+. (C) The subjective gamma ratio.

If $\alpha_+ = \alpha_-$, EV is identical to the real outcome, i.e. $EV = R_+p + R_-(1-p)$, suggesting a faithful perception on the outcome of risky option. Otherwise, $EV > R_+p + R_-(1-p)$ given $\alpha_+ > \alpha_-$ or $EV < R_+p + R_-(1-p)$ given $\alpha_+ < \alpha_-$, indicating an overestimation or underestimation of the outcome of risky option. This is consistent with the distribution of subjective value shown in Fig. 1.

SR can also be worked out as:

$$SR = \frac{\alpha_+\alpha_-(R_+ - R_-)}{\alpha_+p + \alpha_-(1-p)}\sqrt{\frac{p(1-p)}{\left(2\alpha_+ - \alpha_+^2\right)p + \left(2\alpha_- - \alpha_-^2\right)(1-p)}} \quad (5)$$

When $\alpha_+ = \alpha_- = \alpha$, we obtained SR for the original Rescorla–Wagner model: $SR = (R_+ - R_-)\sqrt{\frac{\alpha p(1-p)}{2-\alpha}}$, which is a symmetric function of probability p and reach its maximum at $p = 0.5$. However, the peak of SR skews toward $p = 0$ given $\alpha_+ > \alpha_-$ or skews toward $p = 1$ given $\alpha_+ < \alpha_-$(Fig. 2A). To further clarify the effect of learning rates on SR, we calculate the ratio of SR over the standard deviation of the outcome, i.e., objective risk (OR): $OR = (R_+ - R_-)\sqrt{p(1-p)}$. We find that the ratio is a decreasing function of probability p given $\alpha_+ > \alpha_-$, but an increasing function of probability p given $\alpha_+ < \alpha_-$ (Fig. 2B). Therefore, if $\alpha_+ > \alpha_-$, the subject risk overestimates the risk of the options with lower rewarding probability but underestimates the risk of the options with higher rewarding probability. At the same time, if $\alpha_+ < \alpha_-$, the subject risk underestimates the risk of the options with lower rewarding probability but overestimates the risk of the options with higher rewarding probability.

In traditional financial market, gamma ratio of excess return of risky assets to the standard deviation of the return is often used to make decision on the adjustment of portfolios [25]. Higher gamma ratio indicates a larger excess return per unit of risk or better performance of manager. In this study, to measure how much excessive subjective value per SR, we defined a subjective gamma ratio: $\gamma = \frac{EV-ER}{SR}$, where ER is the average of the actual reward of risky option. Fig. 2C shows the subjective gamma ratio over probability receiving reward. Given $\alpha_+ > \alpha_-$, the subjective gamma is positive and the peak of subjective gamma is larger than 0.5, indicating that subjects overestimate the real reward and are sensitive to the reward at larger probability. Given $\alpha_+ < \alpha_-$, the subjective gamma is negative and reaches its minimum given a probability small than 0.5, suggesting that subjects underestimate the real reward and more sensitive to the reward at small probability.

2.2. Probabilistic choice of the risky option based on subjective value

The risky decision tasks in neuroscience study often include two alternative options, for example option R and option L. The choice probability depends on the difference between divisively normalized subjective values of options [13]:

$$p_R = \left[1 + e^{-\lambda\frac{V_R - V_L}{V_R + V_L + \theta}}\right]^{-1} \quad (6)$$

where p_R is the probability choosing the option R, V_R and V_L are the subjective values of option R and L. λ is the sensitivity of value discrimination. Larger λ leads to bigger difference between probabilities choosing option R and that of option L given fixed subjective values of options. The divisive normalization parameter θ reflects the homeostasis/wealth of subject. Larger θ leads to a poorer discrimination between perceived values of options.

2.3. Behavior of two-stage model on risky task with fixed mean and variant coefficient of variance

McCoy and Platt [15] let monkey chose one option with certain reward or one risky option with random reward by saccade. The outcome of risky option is either smaller or larger reward at half-to-half chance and the mean of reward equals to outcome of certain option 150. The outcome of risky option has six levels of coefficient of variance (CV: the ratio of standard deviation over the mean): 0.0667, 0.1,0.167,0.333,0.5 and 0.667. Their experiments showed that monkeys chose the risky option at higher probability given larger CV, indicating the risk seeking behavior. Particularly, probability of choosing the risky option over previous reward exhibits as a check marker, indicating that gain and loss from the risky option has asymmetric effects on the consequent choice. Here

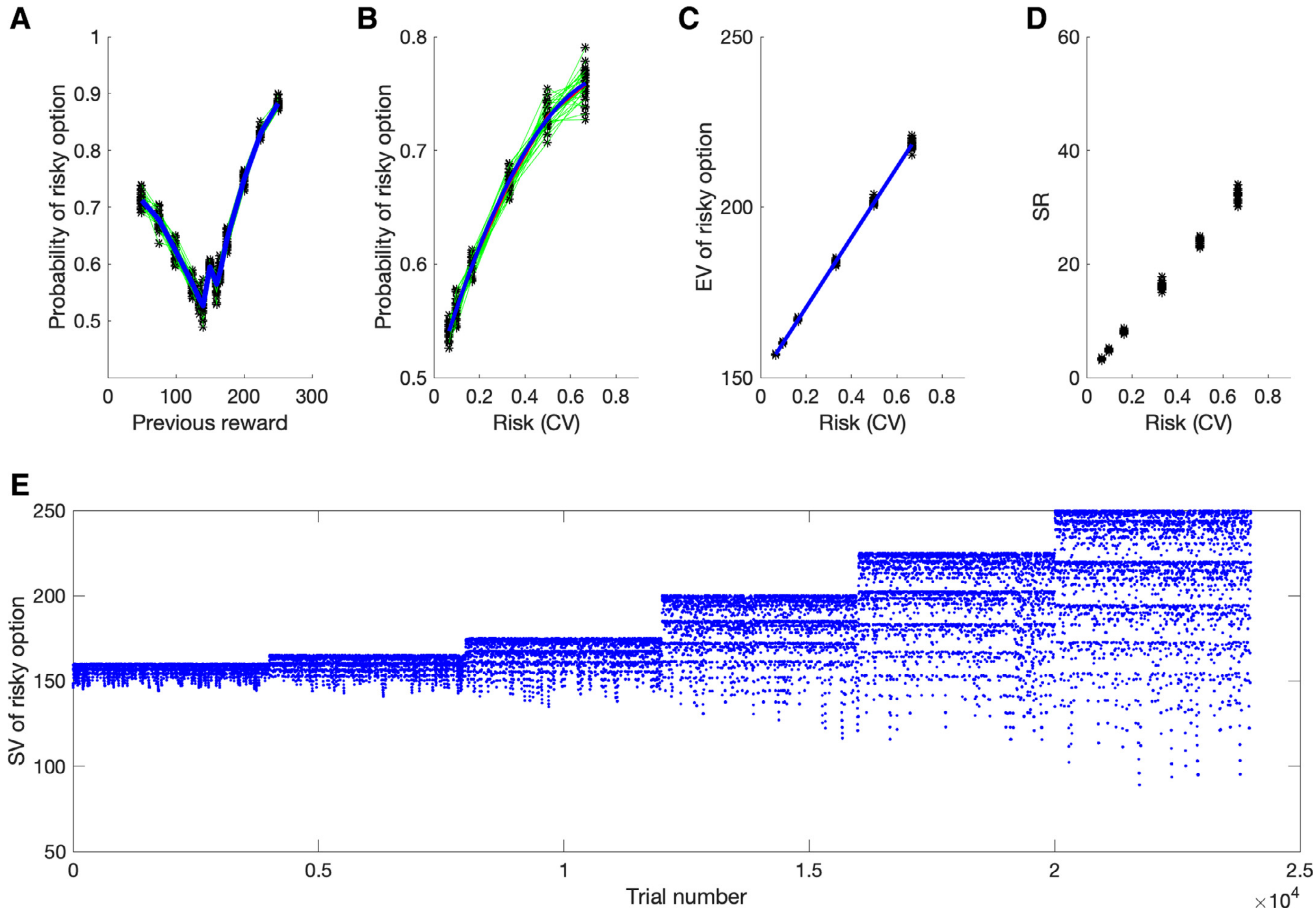


Fig. 3. The risk-seeking behavior captured by two-stage model. (A) The probability of risky choice over previous reward shows an asymmetric check marker. (B) The probability of risk choice as an increasing function of risk. (C) The mean of subjective value of the risky option is an increasing function of risk. (D) The subjective risk is an increasing function of risk. (E) The subjective value of risky option varies with the simulation trials in one session. The parameters are chosen as: $\alpha_+ = 0.8, \alpha_- = 0.15,\ \theta = 300,\ \lambda = 15$. The green lines denote one session while blue line denotes average over sessions.

we simulated this task using our model described as Eqs. (2) and (6). We carried out the simulation with 30 sessions and 3000 trials per session for each risk level.

Our two-stage model captures the risk seeking behavior of monkey observed in McCoy and Platt's experiments (Fig. 3). First, the probability of risky choice decreases with the previous reward when the previous reward is smaller than the certain reward, but the probability of risky option increases with the previous reward if the previous reward is larger than the certain reward. Thus, the plot consists of one left and right branch and the exhibits an asymmetric effect on the probability of choosing risky option in the next trial as shown in McCoy and Platt's experiments (Fig. 3A). Actually, when the previous reward is smaller than the certain reward, the previous reward should be R_- of one risky option and the smaller the previous reward means the larger CV of the risky option. Therefore, the decrease of probability of risky choice with the increase of R_- means a risk-seeking behavior. Second, the probability of choosing risky option increases along with the increasing of risk (Fig. 3B). Third, the subjective value of risky option fluctuates given a fixed level of risk (Fig. 3E), but the average of the subjective value linearly increases with the risk (Fig. 3C). Fourth, SR also linearly increases with the risk (Fig. 3D). The results in Fig. 3C are consistent with our formula for the mean of subjective value and3D are consistent with our formula for subjective risk.

Besides the risk-seeking behavior, the model exhibits rich repertoires of risk behaviors. 1) Risk-avoiding behavior (data are not shown). If EV of risky option is smaller than the reward of certain option due to $\alpha_+ < \alpha_-$, or even EV of risky option equals to the reward of certain option but with large SR, the model exhibits risk-avoiding behavior, i.e., the probability choosing risky option decreases with the increase of CV. 2) Risk-neutral behavior (data are not shown). If SR of risky option is small and EV of risky option equals to the outcome of certain option due to $\alpha_+ = \alpha_-$, the model exhibits a risk-neutral behavior, i.e., risky option and certain option was chosen at half-half chance level. 3) Seeking small risk but avoiding large risk (Fig. 4). We simulate the same task using parameters $\alpha_+ = 0.8, \alpha_- = 0.35,\ \theta = 300,\ \lambda = 15$. We find that the probability of risky option slightly increases with the previous reward when the previous reward is smaller than the certain reward, but sharply increases with the previous reward when the previous reward is larger than the certain reward (Fig. 4A). Given a larger CV of risky option, once the outcome is smaller reward R_-, the subjective value has a bigger drop and could be much smaller than the certain value (Fig. 4E), leading to lower probability to choose the risky option again. Given a smaller CV of risky option, the subjective value has a smaller drop even the outcome is R_- and the subjective value is close to the certain value, resulting a higher probability to choose the risky option. In brief word, the subject seeks a small risk but avoid large risk. EV of risky option increases with

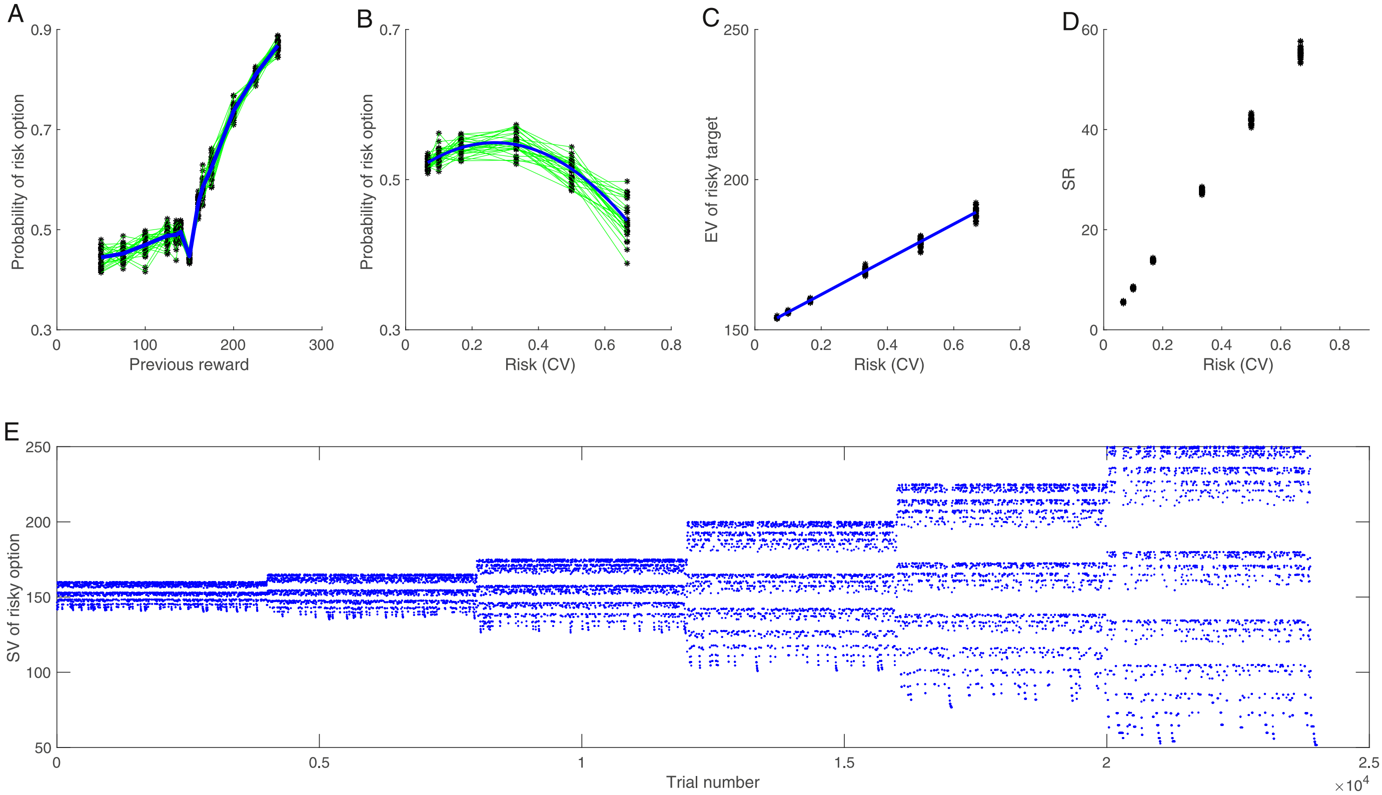


Fig. 4. Seeking small risk but avoiding large risk. (A) The probability of risky choice over previous reward. (B) The probability of risk choice as an increasing function of risk. (C) The mean of subjective value of risky option is an increasing function of risk. (D) The subjective risk is an increasing function of risk. (E) The subjective value of risky option varies with the simulation trials in one session. The parameters are chosen as: $\alpha_+ = 0.8, \alpha_- = 0.35,\ \theta = 300,\ \lambda = 15$. The green lines denote one session while blue line denotes average over sessions.

the risk due to $\alpha_+ > \alpha_-$(Fig. 4C). SR also increases with the risk (Fig. 4D) and the slope is larger than that in Fig. 3D. The raster plots of SV in Fig. 4D is sparser than that in Fig. 3D given large CV, which is consistent with SR. Because of the dynamics of the subjective value of risky option, the probability choosing risky option increases given small risk but decreases with the further increasing of risk (Fig. 4B). This kind of seeking small risk but avoiding large risk behavior is reasonable and quite interesting, but this phenomenon has not been studied in previous researches.

2.4. Behavior of two-stage model on risky tasks with variant mean but fixed coefficient of variance

Yamada et al. [35] investigated the risk attitude of well-trained monkey. The experiment includes four conditions. Each condition has one certain option and one risky option which offers five different risky rewards: zero or nonzero at half-to-half chance and the average magnitude of risky reward ranges from smaller to larger than that of certain option. The magnitude of reward offered by certain option is 60, 120, 180, and 240 for different condition, respectively. Yamada et al. fitted the data from four conditions as whole using choice probability $p_R = [1 + e^{-\beta(Eu_R - Eu_C)}]^{-1}$ based on expected utility function $Eu_R = pv^\alpha$ (where p is the probability of obtaining the offered reward, v is the magnitude of the offered reward). They found that two well-trained monkeys are risk avoiding, i.e., the parameter of utility function $\alpha < 1$.

In this study, two-stage model is used to simulate the tasks in Yamada et al. [35]. The sensitivity of choice and wealth/thirst parameters are set as $\lambda = 20,\ \theta = 240$ but the learning rates are varied. The simulated data are shown as dots in Fig. 5. The simulated data are fitted to choose probability $p_R = [1 + e^{-\beta(Eu_R - Eu_C)}]^{-1}$ and utility function $Eu_R = pv^\alpha$. The two-stage model exhibits rich risky behavior as shown in Fig. 5. First, the model exhibits risk-avoiding behavior that has been observed in Yamada et al. [35] (Fig. 5A). The subjective value of the risky option underestimates the actual reward of risky option if $\alpha_+ < \alpha_-$, as the result, the probability to chose the risky option will be attenuated. To compensate this attenuation, the parameter of utility function must be smaller than one during the data fitting, indicating risk avoiding. In the present study, the learning rates are chosen as $\alpha_+ = 0.09, \alpha_- = 0.078$, which means that the average of the subjective value overestimates the actual reward of the risky option. But, the large fluctuation of the subjective value increases the possibility that instantaneous subjective value is much smaller than the certain reward and cannot be updated. The consequence is to lower the probability of choosing risky option, leading to a risk avoiding behavior (Fig. 5A).

Second, the model also can exhibit risk seeking behavior as observed in McCoy and Platt [15] (Fig. 5B). On the one hand, the model overestimates the value of the risky option due to learning rates $\alpha_+ > \alpha_-(\alpha_+ = 0.09, \alpha_- = 0.058)$. On the other hand, the SR of risky option is comparatively small (Fig. 5 B2 and B3) and the subjective value will not be trapped into big drop of the subjective value. As the result, the probability to choose the risky option was increased. To meet the incremental probability choosing risky option, the parameter of utility function should be larger than one during the data fitting, which implies a risk-seeking behavior.

Third, the model predicts a new type of risk attitude: risk switching (i.e. seeking the small risk but avoiding larger risk) (Fig. 5C). If the standard deviation of the subjective value is small and the average of the subjective value is larger than that of cer-

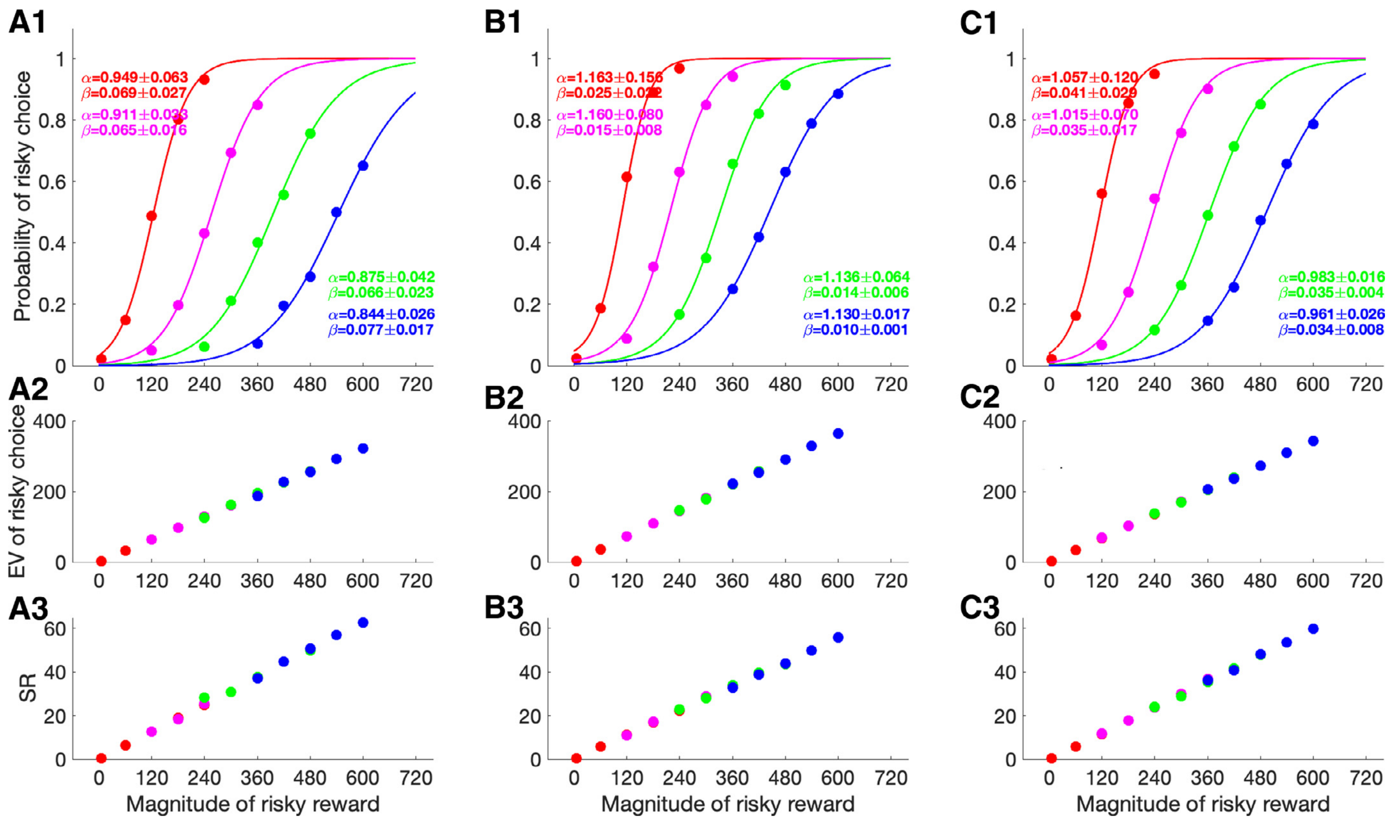


Fig. 5. Simulated risk behavior on tasks with fixed CV but variant average reward. The sensitivity and wealth/thirst parameters are set as: $\lambda = 20$, $\theta = 240$. The learning rates are set as $\alpha_+ = 0.09, \alpha_- = 0.078$ in (A), $\alpha_+ = 0.09, \alpha_- = 0.058$ in (B), and $\alpha_+ = 0.09, \alpha_- = 0.068$ in (C). The dots in figures are simulated data and the lines are fitted choice probability based one expected utility function. The color denotes the magnitude of certain reward, red for 60, magenta for 120, green for 180, and blue for 240. The top panels show the fitted choice function for each condition with different learning rates. Middle panels show the average of subjective value of each risky option. Bottom panels show the subjective risk of each risky option. (A) Risk-avoiding behavior. The utility function parameters $\alpha < 1$ for all conditions (B) Risk-seeking behavior. The utility function parameters $\alpha > 1$ for all conditions. (C) Risk-switching behavior: seeking small risk but avoiding large risk. The utility function parameter changes from $\alpha > 1$ to $\alpha < 1$ along with the increase of the risk.

tain option, the risky option is more likely to be chosen. However, if the standard deviation of the subjective value is large, even the average of the subjective value is larger than that of actual reward, the certain option is more likely to be chosen in void the big loss due to the larger deviation of risky option (see Fig. 5A3, 5B3, and 5C3). To make this clear, we calculate subjective gamma ratio for each risky option under four conditions. Although subjective gamma ratio for each risky option is fluctuating, the subjective gamma ratio can be clustered into three groups and each group of ratios corresponds to one kind of risk attitude. The risk avoiding behavior has a small ratio, while the risk seeking behavior has a big ratio. The subjective gamma ratio for risk switching behavior falls between risk avoiding and risk seeking behavior (Fig. 6A). The average of subjective gamma ratio over four conditions shown in Fig. 6B clearly show a positive relation between γ and α in utility function. At the same time, the γ ratio has not significant difference within same risk attitude but has significant difference between different risk attitude (Fig. 6C).

2.5. Probability distortion shown by two-stage model on choice between two gambles

Human subjects may distort the probability when they face the risky options. Hsu et al. [8] has designed one task to investigate the neural correlates of probability distortion. In their task, human subjects were requested to make a choice between two gambles (p_1, x_1) and (p_2, x_2), where (p_i, x_i) denotes gamble i providing reward x_i at probability p_i. Hsu et al. assumed that subjects' utility function as $U(x) = x^{\alpha_{PD}}$, and the first gamble is chosen at probability $p = [1 + e^{-\lambda(\pi(p_1)U(x_1) - \pi(p_2)U(x_2))}]^{-1}$ with Prelec weighting function: $\pi(p) = 1/\exp[(-\ln p)^{\rho}]$. They found that some of human subject overweight small probability and underweight large probability ($\rho < 1$) while some human subjects underweight small probability and overweight large probability ($\rho > 1$). They also demonstrated the relation between neural activities in the striatum during valuation of monetary gambles with the nonlinearity of probability predicted by prospect theory. However, the origination of the nonlinearity of probability has not been revealed yet. Here we used two-stage model to perform the tasks used in Hsu et al. [8] and fitted the simulated choice data to identical model they used. Three types of learning rates are chosen during the simulation: A) $\alpha_+ = 0.2$, $\alpha_- = 0.3$; B) $\alpha_+ = 0.01$, $\alpha_- = 0.01$; and C) $\alpha_+ = 0.1$, $\alpha_- = 0.02$. The average and the standard deviation of the subjective value of each gamble can be worked out according to formula: $EV = \frac{\alpha_+ x p}{\alpha_+ p + \alpha_-(1-p)}$ and $SR = \frac{\alpha_+ \alpha_- x}{\alpha_+ p + \alpha_-(1-p)} \sqrt{\frac{p(1-p)}{(2\alpha_+ - \alpha_+^2)p + (2\alpha_- - \alpha_-^2)(1-p)}}$ based on the probability p and reward x. Because of learning rates and the probability to receive the reward, the average of subjective value for gambles in Fig. 7B ($\alpha_+ = \alpha_- = 0.01$) is smaller than that in Fig. 7C ($\alpha_+ = 0.01$, $\alpha_- = 0.02$) but larger than that in Fig. 7A ($\alpha_+ = 0.2$, $\alpha_- = 0.3$), the SR in Fig. 7C is larger than that in Fig. 7B but smaller than that in Fig. 7A.

The simulated subjective values of gambles in all tasks shown in the top two rows of Fig. 8 are consistent with the theoretical results in Fig. 7. The average of subjective value of (p_1, x_1) almost

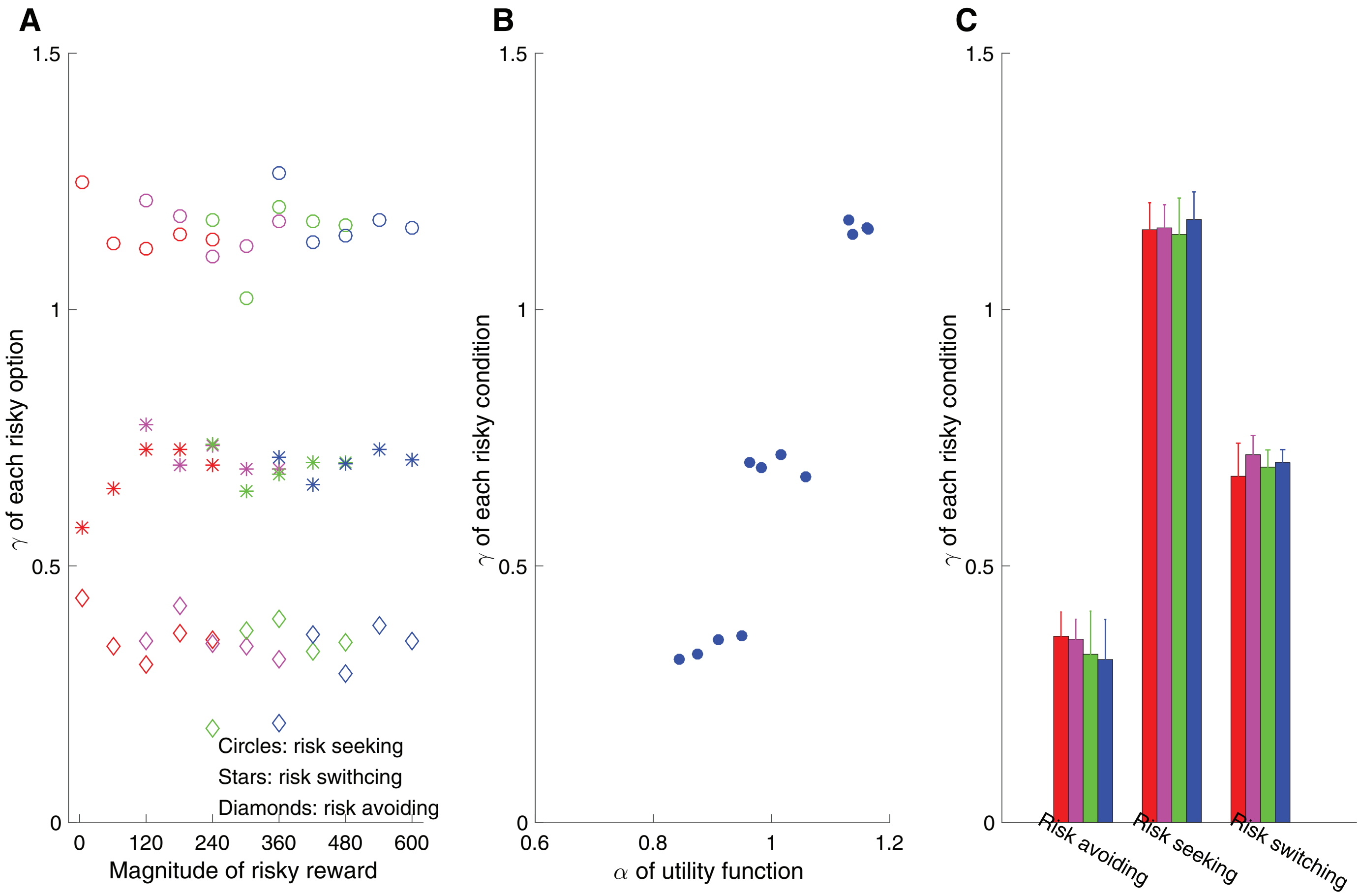


Fig. 6. Subjective gamma ratio and risk attitude. (A) Gamma ratio for each risky option. Circles, stars, and diamonds denote risk seeking, risk avoiding and risk switching attitude, respectively. Color denotes the magnitude of certain reward: red for 60, matagna for 120, green for 180, and blue for 240, respectively. (B) Gamma ratio as a function of α of utility function. (C) Gamma ratio within same risk attitude and between different risk attitude.

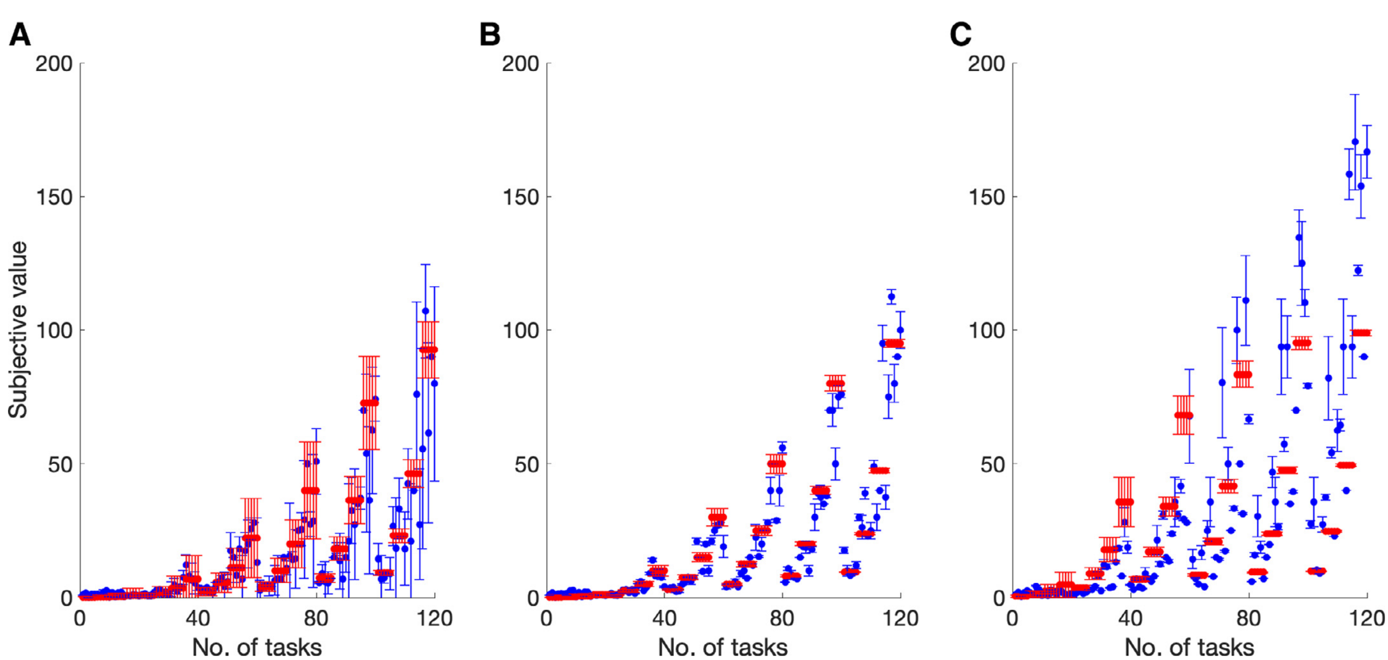


Fig. 7. Subjective value of each gamble. (A) $\alpha_+ = 0.2, \alpha_- = 0.3$;(B) $\alpha_+ = 0.01, \alpha_- = 0.01$; (C) $\alpha_+ = 0.1, \alpha_- = 0.02$. Red for gamble (p_1, x_1) and blue for gamble (p_2, x_2), dots denote the average of subjective value and bars denote the subjective risk.

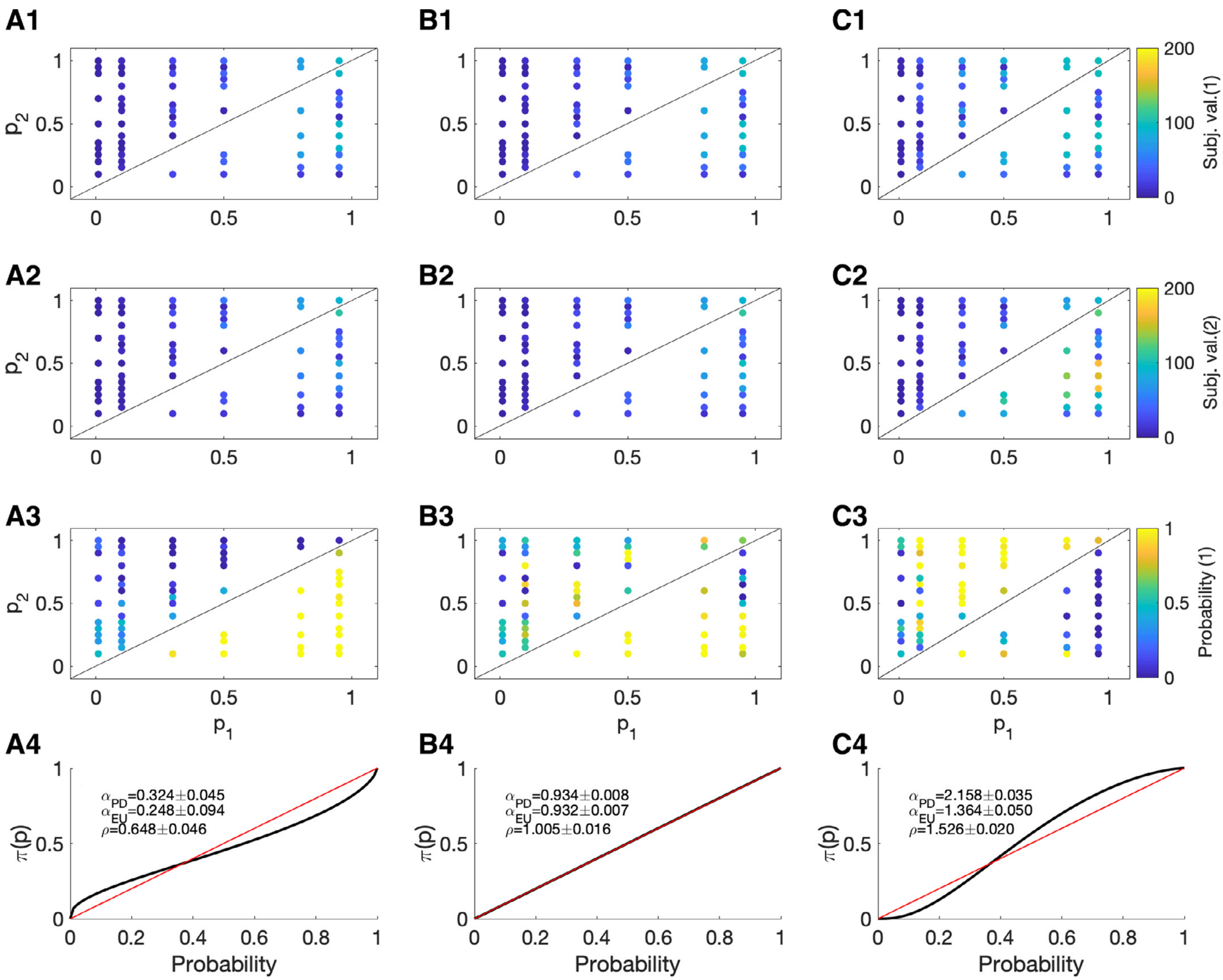


Fig. 8. Weighting of probability inferred from simulated choices. (A) Overweighting small probability but underweighting large probability. (B) Linear weighting of probability. (C) Underweighting small probability but overweight large probability. Black line is weighting function and red line is diagonal line. ρ is parameter of Prelec function $\pi(p) = 1/\exp[(-\ln p)^\rho]$, α_{PD} is parameter of utility function used by Hsu et al. [8], α_{EU} is parameter of utility function in expected utility theory. The first and second row show the subjective value of gamble 1 and gamble 2, respectively. The third row shows the probability to choose gamble 1. The fourth row shows the weighting of probability.

the same as that of (p_2, x_2) (Fig. 8) and gamble (p_1, x_1) has a larger deviation given $p_1 < p_2$. However, the average of subjective value of (p_1, x_1) is larger than that of (p_2, x_2) and the deviation of subjective value of (p_1, x_1) is smaller than that of gamble (p_2, x_2) given $p_1 > p_2$. As consequences, the gamble (p_1, x_1) was chosen more frequently than gamble (p_2, x_2) given $p_1 < p_2$ and the gamble (p_2, x_2) was chosen more frequently than gamble (p_1, x_1) given $p_1 > p_2$ as shown in Fig. 8A2. In Fig. 8B, the average of subjective value of gamble (p_1, x_1) has no big difference with that of gamble (p_2, x_2). The gamble (p_1, x_1) and (p_2, x_2) are almost equally chosen. In Fig. 8C, the average of subjective value of gamble (p_1, x_1) is larger than that of gamble (p_2, x_2) and the deviation of subjective value of gamble (p_1, x_1) is small given $p_1 < p_2$. The average of subjective value of gamble (p_1, x_1) is smaller than that of gamble (p_2, x_2) and the deviation of subjective value of gamble (p_1, x_1) is smaller than that of gamble (p_2, x_2) given $p_1 > p_2$. As results, the gamble (p_1, x_1) was chosen more frequently than gamble (p_2, x_2) given $p_1 > p_2$ and the gamble (p_2, x_2) was chosen more frequently than gamble (p_1, x_1) given $p_1 < p_2$ (Fig. 8C).

The simulated choice probabilities were fitted to the model $p = [1 + e^{-\lambda(\pi(p_1)U(x_1) - \pi(p_2)U(x_2))}]^{-1}$ with utility function $U(x) = x^{\alpha_{PD}}$ and weighting function $\pi(p) = 1/\exp[(-\ln p)^\rho]$ and the fitted weighting functions are shown in the bottom row of Fig. 8. Our model demonstrates three types of probability distortion. The small probability is over weighted and large probability is underweighted with learning rates $\alpha_+ = 0.2$, $\alpha_- = 0.3$ in Fig. 8A4; the small probability is underweighted but large probability is over weighted with learning rates $\alpha_+ = 0.1$, $\alpha_- = 0.02$ in Fig. 8C4; and the probability can be linear weighted given small and symmetric learning rates $\alpha_+ = 0.01$, $\alpha_- = 0.01$ in Fig. 8B4.

Although the risk attitude of subjects has not been reported in Hsu et al. [8], these human subjects should have their own risk attitude. Here, we fitted the simulated choices probability to the expected utility theory: $p = [1 + e^{-\lambda(p_1 U(x_1) - p_2 U(x_2))}]^{-1}$ with $U(x) = x^{\alpha_{EU}}$. The results shown in Fig. 8 demonstrate that risk avoiding subjects may overweight small probability but underweight large probability since $\alpha_{EU} < 1$ and $\rho < 1$ in Fig. 8A3. The risk seeking subjects underweight small probability and overweight large probability since $\alpha_{EU} > 1$ and $\rho > 1$ in Fig. 8C4. Those

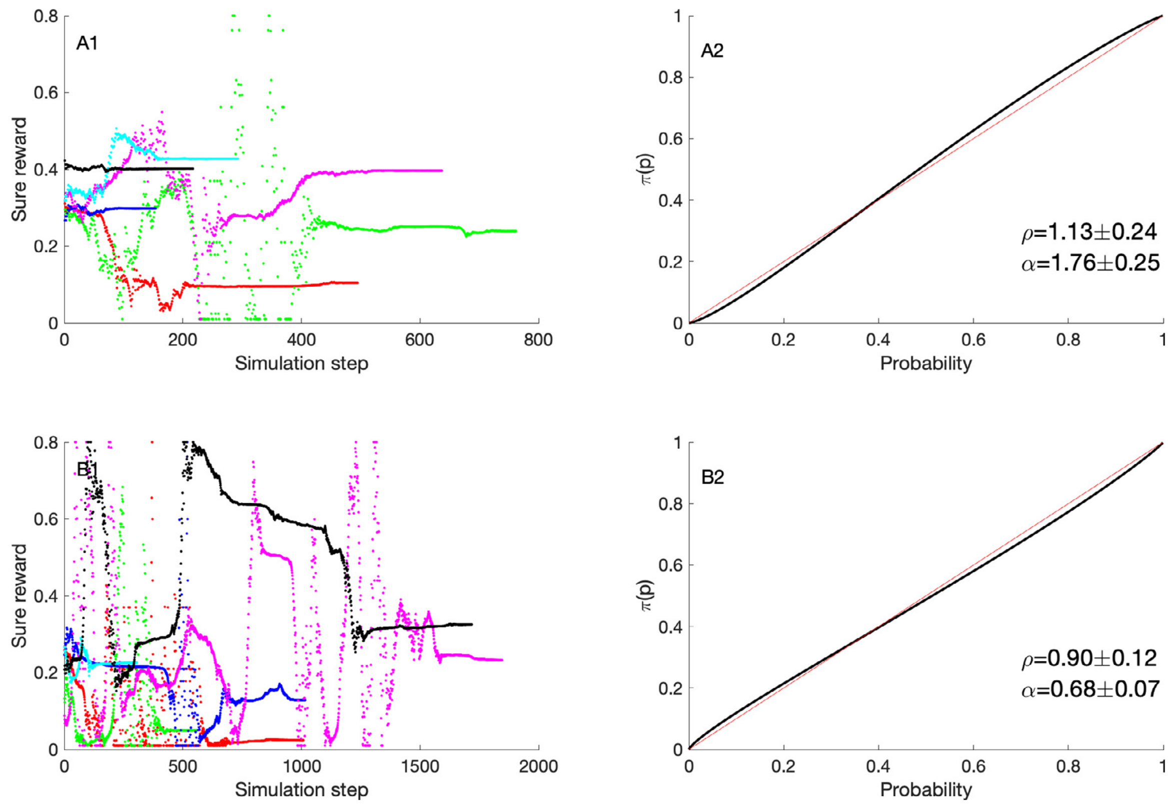


Fig. 9. Weighting of probability in simulated PEST task. (A1). The time course of the sure reward in an example trial of certainty equivalent searching task given $\alpha_+ = 0.05$, $\alpha_- = 0.02$, $\lambda = 4$, and $\beta = 0$. (A2) The weighting function of probability underweighting small probability and overweighting large probability. (B1) The time course of the sure reward in an example trial of certainty equivalent searching task given $\alpha_+ = 0.02$, $\alpha_- = 0.04$, $\lambda = 4$, and $\beta = 0$. (B2) The weighting function overweighting small probability and underweighting large probability.

subjects weighting probability linearly show a risk avoiding behavior ($\alpha_{EU} < 1$ and $\rho \approx 1$) in Fig. 8B4.

2.6. Probability distortion in certainty equivalent searching task using parameter estimation by sequential testing (PEST)

Probability distortion has been revealed in Macaque monkeys [27] as well as in human subjects. Monkeys are requested to make a choice between a risky option and a sure option. In one block of experiment, the risky option provides a reward (RR=0.5) at a probability (0.1,0.25,0.4,0.6,0.75, and 0.9), while the reward of sure option will be adjusted using an adaptive psychometric measurement technique (Parameter Estimation by Sequential Testing, PEST) [14]. If the risky option was chosen, the reward of sure option was increased by ε on the next step. If the sure option was chosen, the reward of sure option was decreased by ε on the next step. There is an upper boundary and lower boundary of the reward of sure option in avoid the explosion in the simulation. Every time two consecutive choices were the same, ε was increased by 20%, and every time the choice was switched from one option to another, ε was increased by 20%. Once ε is smaller than a threshold, the simulation was stopped and the reward of sure option is the certainty equivalent of the risky option. Fig. 9 shows the time course of the simulated reward of sure option. It is easy to see that sure reward fluctuates in early stage of simulation and finally approaches to a stable value (certainty equivalent: CE). The parameters of nonlinear weighting can be obtained by fitting the simulated CE to the formula $CE^\alpha = \pi(p)RR^\alpha$ with $\pi(p) = 1/\exp[(-\ln p)^\rho]$. Fig. 9A show the results given $\alpha_+ = 0.05$, $\alpha_- = 0.02$, $\lambda = 4$, and $\beta = 0$. The certainty equivalents were obtained after hundreds simulation steps, and the average of CEs over 100 trials simulation are 0.0942, 0.2296, 0.3131,0.3777,0.4148, and 0.4587 for reward of risky option at probability 0.1,0.25,0.4,0.6,0.75, and 0.9 respectively. These CEs are larger than the objective average reward from risky option owing to $\alpha_+ > \alpha_-$. The parameter for weighting function is $\rho = 1.13 \pm 0.24$, indicating underweighting small probability and overweighting large probability (Fig. 9A2). The parameter for utility function is $\alpha = 1.76 \pm 0.25$, implying risk seeking behavior. Fig. 9B show the results given $\alpha_+ = 0.02$, $\alpha_- = 0.04$, $\lambda = 4$, and $\beta = 0$. The average CEs are 0.0255, 0.0724, 0.1304, 0.2161, 0.2944, and 0.4350 for reward at probability 0.1,0.25,0.4,0.6,0.75, and 0.9 respectively. These CEs are smaller than objective average of reward from risky option due to $\alpha_+ < \alpha_-$. The parameter for utility function is $\alpha = 0.68 \pm 0.07$, indicating a risk avoiding behavior. The parameter for weighting function is $\rho = 0.90 \pm 0.12$, suggesting an overweighting of small probability and underweighting of large probability. As brief summary, by varying the learning rates, our model can reproduce the observations in [27] and predicts a different type of probability distortion, indicating that learning rates play important roles in the probability perceiving process.

3. Discussion

Risk attitude and probability distortion play central roles in the decision making under risky circumstance. In this study, we investigated their neural basis and found that the valuation process through reinforcement learning from experience is the common neural basis of the risk attitude and the probability distortion. The learning rates in the reinforcement learning determine the average of the subjective value and the subjective gamma ratio ($\gamma = \frac{EV-ER}{SDV}$). If the learning rate for gains a_+ is smaller than the learning rate for losses a_-, the average of subjective value is smaller than that of real reward of risky option and the subjective gamma ratio is smaller than zero, leading to a risk avoiding behavior. When $a_- < a_+$, the average of subjective value is larger than that of real reward of risky option and the subjective gamma ratio is positive. On the one hand, if the subjective gamma ratio is small, the standard deviation of the subjective value is large comparing to the surplus of the subjective value over the real reward, leading to a risk avoiding behavior (Fig. 6); on the other hand, if the subjective gamma ratio is large, the standard deviation of the subjective value is small comparing to the surplus of the subjective value over the real reward, causing a risk seeking behavior. When the subjective gamma ratio is not large enough, the model exhibits a new type of risk behavior: seeking small risk but avoiding large risk.

The learning rates in reinforcement learning determine the SR. If $a_- < a_+$, the SR of risky option receiving reward at small probability will be augmented and the subjective gamma ratio peaks at large probability (Fig. 2). As the result, the gamble with large rewarding probability will be chosen more frequently than the gamble with small reward probability in the task designed by Hsu et al. [8], leading to underweighting of small probability and overweighting of large probability. Actually, this type of probability distortion has long been observed in one comparative study that monkeys increase the probability of highly probable choice and decrease the probability of low one [1]. If $a_- > a_+$, the SR of risky option receiving at large probability is augmented and the subjective gamma ratio has a trough at small probability (Fig. 2). The consequence is that the gamble with small rewarding probability will be chosen more frequently than the gamble with large reward probability in the task designed, implying overweighting of small probability and underweighting large probability. These results imply that the enhancement of VTA neurons leads subjects shift their risk attitude from risk avoiding to risk seeking, while the enhancement of LHb activity will lead subjects shift their risk attitude from risk seeking to risk avoiding.

The traditional reinforcement learning model, which can be looked as our model with symmetric learning rates $a_- = a_+$, can only produce the risk avoiding behavior and cannot explain the risk seeking behavior as in McCoy 2005. Risk sensitive temporal difference model has been proposed to explore risk sensitivity [5],[6],[20] and stated that $a_- > a_+$ leads to risk avoiding behavior and $a_- < a_+$ leads to risk seeking behavior. However, our study found that not only $a_- < a_+$ but also a large subjective gamma ratio can result in risk seeking behavior, even $a_- < a_+$ can lead to a larger average subjective value, the larger SR can cancel out the surplus of subjective value over real reward and lead to risk avoiding behavior (Fig. 6).

We obtained above conclusions based on simulations with good discrimination between values due to large sensitivity parameter. If the discrimination ability on values is impaired by decreasing the sensitivity parameter, the choice will more frequently switch between two options. Considering an extreme condition that $\lambda = 0$, any one of the options will be chosen by half chance. Thus, the certainty equivalent in PEST task converges to a number which is independent of reward receiving probability of risky option, leading to an overweighting of small probability and underweighting of large probability.

Declaration of Competing Interest

None.

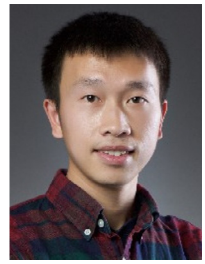
Acknowledgements

This work was supported by NSFC under grant: 31671077, 31271169, 31500863 and BMSTC(Beijing Municipal Science and Technology Commission) under grant: Z171100000117007.

References

[1] M.E. Bitterman, The comparative analysis of learning, Science 188 (1975) 699–709, doi:10.1126/science.188.4189.699.
[2] S. Gilaie-Dotan, A. Tymula, N. Cooper, J.W. Kable, P.W. Glimcher, I. Levy, Neuroanatomy predicts individual risk attitudes, J. Neurosci. 34 (2014) 12394–12401, doi:10.1523/JNEUROSCI.1600-14.2014.
[3] S. Fecteau, A. Pascual-Leone, D.H. Zald, P. Liguori, H. Theoret, S. Paulo Boggio, F. Fregni, Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making, J. Neurosci. 27 (23) (2007) 6212–6218, doi:10.1523/jneurosci.0314-07.2007.
[4] C.D. Fiorillo, P.R. Tobler, W. Schultz, Discrete coding of reward probability and uncertainty by, Dopamine Neurons Sci. 299 (2003) 1898–1902, doi:10.1126/science.1077349.
[5] M.J. Frank, A.A. Moustafa, H.M. Haughey, T. Curran, K.E. Hutchison, Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 16311–16316, doi:10.1073/pnas.0706111104.
[6] M.J. Frank, B.B. Doll, J. Oas-Terpstra, F. Moreno, Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation, Nat. Neurosci. 12 (2009) 1062–1068, doi:10.1038/nn.2342.
[7] P.W. Glimcher, Choice: towards a standard back-pocket model, in: P.W. Glimcher, C.F. Camerer, E. Fehr, R.A. Poldrack (Eds.), Neuroeconomics: Decision Making and the Brain, Academic Press, New York, 2009, pp. 503–521, doi:10.1016/B978-0-12-374176-9.00032-4.
[8] M. Hsu, I. Krajbich, C. Zhao, C.F. Camerer, Neural response to reward anticipation under risk is nonlinear in probabilities, J. Neurosci. 29 (7) (2009) 2231–2237, doi:10.1523/jneurosci.5296-08.2009.
[9] H. Ishii, S. Ohara, P.N. Tobler, K.I. Tsutsui, T. Iijima, Inactivating anterior insular cortex reduces risk taking, J. Neurosci. 32 (45) (2012) 16031–16039, doi:10.1523/jneurosci.2278-12.2012.
[10] D. Kahneman, A. Tversky, Prospect theory: an analysis of decision under risk, Econometrica 47 (2) (1979) 263, doi:10.2307/1914185.
[11] D. Knoch, L.R. Gianotti, A. Pascual-Leone, V. Treyer, M. Regard, M. Hohmann, P. Brugger, Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior, J. Neurosci. 26 (24) (2006) 6469–6472, doi:10.1523/jneurosci.0804-06.2006.
[12] C.M. Kuhnen, B. Knutson, The neural basis of financial risk taking, Neuron 47 (2005) 763–770, doi:10.1016/j.neuron.2005.08.008.
[13] K. Louie, M.W. Khaw, P.W. Glimcher, Normalization is a general neural mechanism for context-dependent decision making, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 6129–6144, doi:10.1073/pnas.1217854110.
[14] D.R. Luce, Utility of Gains and losses: Measurement-Theoretic and Experimental Approached, Lawerence Erlbaum Associates, Mahwah, NJ, 2000, doi:10.4324/9781410602831.
[15] A.N. McCoy, M.L. Platt, Risk-sensitive neurons in macaque posterior cingulate cortex, Nat. Neurosci. 8 (9) (2005) 1220–1227, doi:10.1038/nn1523.
[16] H.M. Markowitz, Portfolio selection, J. Finance 7 (1952) 77–91, doi:10.1111/j.1540-6261.1952.tb01525.x.
[17] M. Matsumoto, O. Hikosaka, Lateral habenula as a source of negative reward signals in dopamine neurons, Nature 447 (2007) 1111, doi:10.1038/nature05860.
[18] M.R. Mehta, M.C. Quirk, M.A. Wilson, Experience-dependent asymmetric shape of hippocampal receptive fields, Neuron 25 (2000) 707–715, doi:10.1016/S0896-6273(00)81072-7.
[19] I.E. Monosov, O. Hikosaka, Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region, Nat. Neurosci. 16 (6) (2013) 756–762, doi:10.1038/nn.3398.
[20] Y. Niv, J. Edlund, P. Dayan, J.P. O'Doherty, Neural prediction errors reveal a risk-sensitive reinforcement learning process in the human brain, J. Neurosci. 32 (2) (2012) 551–562, doi:10.1523/jneurosci.5498-10.2012.
[21] R.A. Rescorla, A.R. Wagner, A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement, in: A.H. Black, W.F. Prokasy (Eds.), Classical Conditioning II: Current Research and Theory, Appleton Century Crofts, New York, 1972, pp. 64–99.
[22] J.D. Roitman, M.F. Roitman, Risk-preference differentiates orbitofrontal cortex responses to freely chosen reward outcomes, Eur. J. Neurosci. 31 (2010) 1492–1500, doi:10.1111/j.1460-9568.2010.07169.x.
[23] S. Rudorf, K. Preuschoff, B. Weber, Neural correlates of anticipation risk reflect risk preferences, J. Neurosci. 32 (47) (2012) 16683–16692, doi:10.1523/jneurosci.4235-11.2012.
[24] W. Schultz, P. Dayan, P.R.;. Montague, A neural substrate of prediction and reward, Science 275 (5306) (1997) 1593–1599, doi:10.1126/science.275.5306.1593.
[25] W.F. Sharpe, The sharpe ratio, J. Portf. Manag. 21 (1) (1994) 49–58, doi:10.3905/jpm.1994.409501.
[26] N.W. Simon, K.S. Montgomery, B.S. Beas, M.R. Mitchell, M.R. LaSarge, I.A. Mendez, C. Banuelos, C.M. Vokes, A.B. Taylor, R.P. Haberman, J.L. Bizon, B. Setlow, Dopaminergic modulation of risky decision-making, J. Neurosci. 31 (48) (2011) 17460–17470, doi:10.1523/JNEUROSCI.3772-11.2011.
[27] W.R. Stauffer, A. Lak, P. Bossaerts, W. Schultz, Economic choices reveal probability distortion in rhesus macaques, J. Neurosci. 35 (7) (2015) 3146–3154, doi:10.1523/jneurosci.3653-14.2015.
[28] R.S. Sutton, A. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998, doi:10.1109/tnn.1998.712192.

[29] P.N. Tobler, J.P. O'Doherty, R.J. Dolan, W. Schultz, Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems, J. Neurophysiol. 97 (2007) 1621–1632, doi:10.1152/jn.00745.2006.
[30] 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 (45) (2008) 11703–11711, doi:10.1523/jneurosci.2870-08.2008.
[31] P.N. Tobler, G.I. Christopoulos, J.P. O'Doherty, R.J. Dolan, W. Schultz, Risk-dependent reward value signal in human prefrontal cortex, Proc. Natl. Acad. Sci. 106 (2009) 7185–7190, doi:10.1073/pnas.0809599106.
[32] S.M. Tom, C.R. Fox, C. Trepel, R.A. Poldrack, The neural basis of loss aversion in decision-making under risk, Science 315 (2007) 515–518, doi:10.1126/science.1134239.
[33] E. van Duuren, G. van der Plasse, J. Lankelma, N.J.M.A. Ruud, M.G.P. Joosten, M.G. Feenstra, M.A. Cyriel, Single-cell and population coding of expected reward probability in the orbitofrontal cortex of the rat, J. Neurosci. 29 (28) (2009) 8965–8976, doi:10.1523/jneurosci.0005-09.2009.
[34] S.W. Wu, M.R. Delgado, L.T. Maloney, The neural correlates of subjective utility of monetary outcome and probability weight in economic and in motor decision under risk, J. Neurosci. 31 (24) (2011) 8822–8831, doi:10.1523/jneurosci.0540-11.2011.
[35] 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 (39) (2013) 15788–15793, doi:10.1073/pnas.1308718110.
[36] L.i. Zhou, M.-Z. Liu, Q. Li, J. Deng, D. Mu, Y.-G. Sun, Organization of functional long-range circuits controlling the activity of serotonergic neurons in the dorsal raphe nucleus, Cell Rep. 18 (12) (2017) 3018–3032, doi:10.1016/j.celrep.2017.08.032.
[37] A.B. Long, C.M. Kuhn, M.L. Platt, Serotonin shapes risky decision making in Monkeys, SCAN 4 (2009) 346–356, doi:10.1093/scan/nsp020.
[38] W.R. Stauffer, A. Lak, P. Bossaert, W. Schultz, Economic choices reveal probability distortion in macaque monkeys, J. Neurosci. 35 (2015) 3146–3154, doi:10.1523/JNEUROSCI.3653-14.2015.
[39] W. Schultz, Electrophysiological correlates of reward processing in dopamine neurons, in: J.-C. Dreher, L. Tremblay (Eds.), Decision Neuroscience, Elsevier Inc., 2017, pp. 21–31, doi:10.1016/B978-0-12-805308-9.00002-6.



Hongzhi You is an associate professor of University of Electronic Science and Technology of China. He works on computational neuroscience and neuromorphic engineering.

Mengya Zhang is a Ph.D candidate in School of Systems Science, Beijing Normal University. Her research focus on computational neuroscience.

Da-Hui Wang is a professor of School of Systems Science Beijing Normal University and State Key Laboratory of Cognitive Neuroscience and Learning of China. He works on neural dynamics underlying decision making, working memory and information processing in brain using nonlinear dynamics.