ARTICLE IN PRESS

Progress in Neurobiology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Progress in Neurobiology

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

Original Research Article

Sunk cost effects hinge on the neural recalibration of reference points in mental accounting

Jiashu Wang [a], Bo Zhang [a,b], Shi Liang [a,b], Jian Li [a,c,*]

[a] School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China
[b] Peking-Tsinghua Center for Life Sciences, Beijing, 100871, China
[c] IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China

ARTICLE INFO

Keywords:
Mental accounting
Sunk cost
Windfall
Dorsolateral prefrontal cortex
Chosen value
Medial orbitofrontal cortex

ABSTRACT

The context of reinforcement history drastically influences human value-based choices. Mental accounting theory concerns how prior outcomes are perceived, combined and assigned into specific "mental" accounts to influence subsequent decisions but remains agnostic about the underlying computational and neural mechanisms. In a two-stage sequential decision-making task, we found previously incurred costs and bonuses biased subjects' choices in the opposite directions with similar magnitudes. Such effects were consistent with a computational model where the reference point was recalibrated by prior gains and losses encoded in the ventral striatum activities. Moreover, individual's susceptibility to prior outcomes was captured by the response of the dorsolateral prefrontal cortex and its functional connectivity with the medial orbitofrontal cortex, whose activity tracked the value of the chosen option. Our findings provide both behavioral and neural evidence of how sunk costs, benefits, and prospects are integrated within the mental accounting framework to influence choice behavior.

1. Introduction

In the tale of Le Petit Prince, "It is the time you have wasted for your rose that makes your rose so important", the fox told the little prince (Saint-Exupery, 1987). The time wasted for the rose is termed sunk cost in behavioral economics and decision science, and how the sunk cost exerts its influence in value-based decision making remains one of the prominent research questions (Thaler, 1985). Recent developments in these fields suggest that contrary to standard economic theory, how options are represented explicitly and implicitly can both bias choice selections, such as the well-studied framing effect and mental accounting effect, respectively (Kahneman and Tversky, 1979; Soman and Cheema, 2001; Thaler, 1985, 1999). Mental accounting refers to a set of cognitive processes to "organize, evaluate, and keep track of financial activities" in a manner similar to that of the financial accounting. More specifically, mental accounting constrains the way financial outcomes, both experienced and prospective, are perceived, integrated and grouped into specific accounts and it is hypothesized to be heavily involved in the value-based decision making and responsible for numerous economic anomalies (Thaler, 1999). This is especially the case in the multiple stage decision problem where the prior outcome is registered but has not been psychologically realized until the end of the experiment. For example, one of the most well-known manifestation of the mental accounting theory is the "sunk cost" effect and it refers to the fact that subjects are frequently influenced by previously experienced and non-recoverable losses and gains when making decisions (Arkes and Blumer, 1985; Staw, 1976; Thaler, 1980), rather than solely focusing on the marginal costs and benefits. Among other factors, sunk cost may engender effects such as the commitment escalation where people are more likely to continue an endeavor after incurring a substantial amount of costs ("throwing good money after bad") (Arkes and Ayton, 1999; Haller and Schwabe, 2014; Sleesman et al., 2012; Staw, 1981; Thaler, 1999).

Extant mental accounting theory tends to ascribe the sunk cost effect to the way prior outcomes and marginal prospects are integrated in option valuation processes and the commitment escalation therefore emerges from the hypothesis that subjects evaluate different prospects by linearly combining sunk and marginal costs/benefits before the valuation transformation. Due to the convexity of loss value function assumed by the prospect theory, people are more likely to continue an endeavor after the occurrence of sunk cost relative to no cost (Arkes and

* Corresponding author at: School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China.
E-mail address: leekin@gmail.com (J. Li).

https://doi.org/10.1016/j.pneurobio.2021.102178
Received 14 January 2021; Received in revised form 12 September 2021; Accepted 20 September 2021
Available online 23 September 2021
0301-0082/© 2021 Elsevier Ltd. All rights reserved.

Blumer, 1985; Bazerman et al., 1984; Fujino et al., 2016; Staw, 1976; Thaler, 1980, 1999; Tversky and Kahneman, 1981). However, this predominant theory failed to reconcile other literatures indicating that the sunk cost sometimes also leads to the de-escalation of commitment effect (Garland et al., 1990; Heath, 1995; Heath and Soll, 1996; McCain, 1986; Roth et al., 2015). For instance, an interesting example is the observation that movie-goers are more reluctant to purchase another ticket after they accidentally lost the original one than after losing an equivalent amount of money, possibly due to the calculation that the movie is not worth the price of two movie tickets (Kahneman and Tversky, 1984; Thaler, 1999).

More critically, the underlying neural implementation regarding how prior outcomes are perceived and integrated to exert their aggregate effect in subsequent decision-making processes under the mental accounting framework remains unclear (Duxbury, 2012; Heath, 1995; Soman and Cheema, 2001; Thaler and Johnson, 1990). One suggestion is that prior costs and benefits are processed in the brain in a manner similar to that of the marginal gain and loss prospects and the neural representations of prior costs and benefits also engage brain activities in the reward processing structures, such as the ventral striatum (Tom et al., 2007). Furthermore, the orbitofrontal cortex (OFC) might encode an explicit preference signal by integrating inputs that represent prior and marginal outcome values (Hare et al., 2009; Kable and Glimcher, 2007; Rangel et al., 2008). The cognitive control network in the brain, such as the lateral prefrontal cortex, has been shown to modulate the valuation process to achieve high-order goals (Crockett et al., 2017; Dixon and Christoff, 2014; Hare et al., 2009) and to be involved in implementing the "do not waste" rule, a high-order goal driving the commitment escalation effect, in recent efforts to study the sunk cost effect using neuroimaging and transcranial direct current stimulation (tDCS) techniques (Bogdanov et al., 2017; Fujino et al., 2016; Haller and Schwabe, 2014).

We used functional magnetic resonance imaging (fMRI) to test how humans evaluate and integrate prior costs and benefits while facing risky gambles. To this end, we designed a more naturalistic two-step decision task where subjects first decided whether to incur a certain amount of non-recoverable cost (on their own will) before making a second choice about purchasing a lottery ticket. Behaviorally, we found prior costs and benefits parametrically decreased and increased the commitment of current endeavor, respectively. This behavioral pattern was consistent with a computational mental accounting model where the reference point was recalibrated by the prior outcomes. Furthermore, individual differences of the sunk cost effect were captured by the model parameter of relative weight subjects assigned to the option values. Importantly, our neuroimaging results showed that the trial-by-trial neural activities in the ventral striatum tracked the magnitudes of prior costs and benefits when lottery purchase decision was contemplated, thus supporting the reference point adjustment hypothesis. On the other hand, individual difference of the sunk cost effect was encoded by the activities of the dorsolateral prefrontal cortex (dlPFC) and superior parietal lobule (SPL). Moreover, the medial OFC (mOFC) responses were associated with the value of chosen option and were subject to the functional connectivity modulation from dlPFC and SPL across participants. Together, these results provide both behavioral and neural evidence for a computational mental accounting model that captured individual differences of the direction and magnitude of sunk cost effects.

2. Materials and methods

2.1. Participants

Thirty-five healthy right-handed volunteers (17 females; mean age: 21.6; age range: 19–26 years old; SD = 2.1) participated in this study. All participants were free of the neurological and psychiatric history. This experiment was conducted in accordance with the protocol approved by the Ethics Committee of School of Psychological and Cognitive Sciences, Peking University and all the subjects provided written informed consents before the experiment.

2.2. Task and design

As described in the main text (Fig. 1), in our task each trial consisted of at most two choices. At the beginning of each trial, the face value of the lottery ticket (50 % chance of winning) and entrance fee (displayed in red) were presented on the computer screen and subjects had up to 5 s to decide whether they would be willing to pay the entrance fee to earn the opportunity to purchase the displayed lottery ticket later in the trial. When no choice was made within the 5 s window, a warning message would appear to remind subjects to respond faster. The trial ended if subjects decided not to pay the entrance fee; otherwise, the trial proceeded and an independent amount of windfall (green) appeared on the screen for 2−4 s before the lottery ticket price tag (red) was revealed and subjects had up to 5 s to make their second choice (whether to purchase the ticket at the listed price). All the information (sunk cost, windfall, ticket price and ticket face value) related to ticket purchase were on the computer screen during the 2nd decision to avoid potential confounds in working memory capacity differences across subjects. A randomly jittered inter-trial interval (ITI) with mean of 2 s (1−3 s) was introduced before the beginning of the next trial. The remaining reaction time of subjects' last choice (up to 5 s) was also added to the next trial's ITI and therefore the whole range of ITI was 1−8 s.

In our design, the face value and the probability of winning for the lottery ticket were fixed at ¥20 and 50 %, respectively. Furthermore, we independently manipulated the amounts of the entrance fee, windfall and lottery ticket price. The entrance fee had five different levels: [0, 1.5, 3, 4.5, 6], and the windfall had 5 levels [0, 0.5, 2.5, 4.5, 6.5] with windfall = 0 repeated twice to ensure enough trials to estimate choice sensitivity to the lottery ticket price without the influence of sunk costs or benefits. Finally, we set the lottery ticket price independently at 6 different levels: [2, 4, 6, 8, 10, 12]. For those trials where the entrance fee was 0, we set the lottery ticket price at [5, 7, 9, 11, 13, 15] such that subjects chose to purchase lottery tickets roughly 50 % of the time according to our pilot study. In summary, we had 5 (entrance fee amount) × 6(windfall amount) × 6(lottery ticket price) = 180 trials for each subject and in 12 trials both the entrance fee and windfall were 0. Subjects received a brief practice session of 8 trials before the real task and the 180 trials were equally divided into two scanning sessions with screen locations of "Pay" and "Quit" (corresponding to the left- and right-hand button box submission) randomized across subjects. Also, each subject was endowed with ¥18 (maximum amount they could have lost in one trial) before the experiment and was told that only one randomly selected trial would be honored for real money to ensure incentive compatibility. Monetary outcomes would not carry over from trial-to-trial and subjects should treat each trial as a separate event with the ¥18 endowment. In addition, subjects also received ¥80 for their participation. On average, subjects received ¥110.

Since we aimed to study how prior outcomes influence subjects' subsequent choices, we focused on trials where subjects paid the entrance fee (valid trials, Supplementary Fig. 1). Across subjects the average number of valid trials was 150.1 (SD = 24.6), and in the valid trials the windfall and entrance fee did not show significant correlation (mean r = -0.002, one-sample t-test t_{34} = -0.168, P = 0.868).

2.3. Procedures

The experiment was implemented in Psychtoolbox 3 (PTB, http://www.psychtoolbox.org/) in MATLAB®R2010b. All stimuli were presented via an MRI compatible screen and subjects indicated their choices by pressing buttons on two button boxes in the scanner (one for each hand). The average time to finish the experiment was 43 min and each scanning session lasted around 20 min.

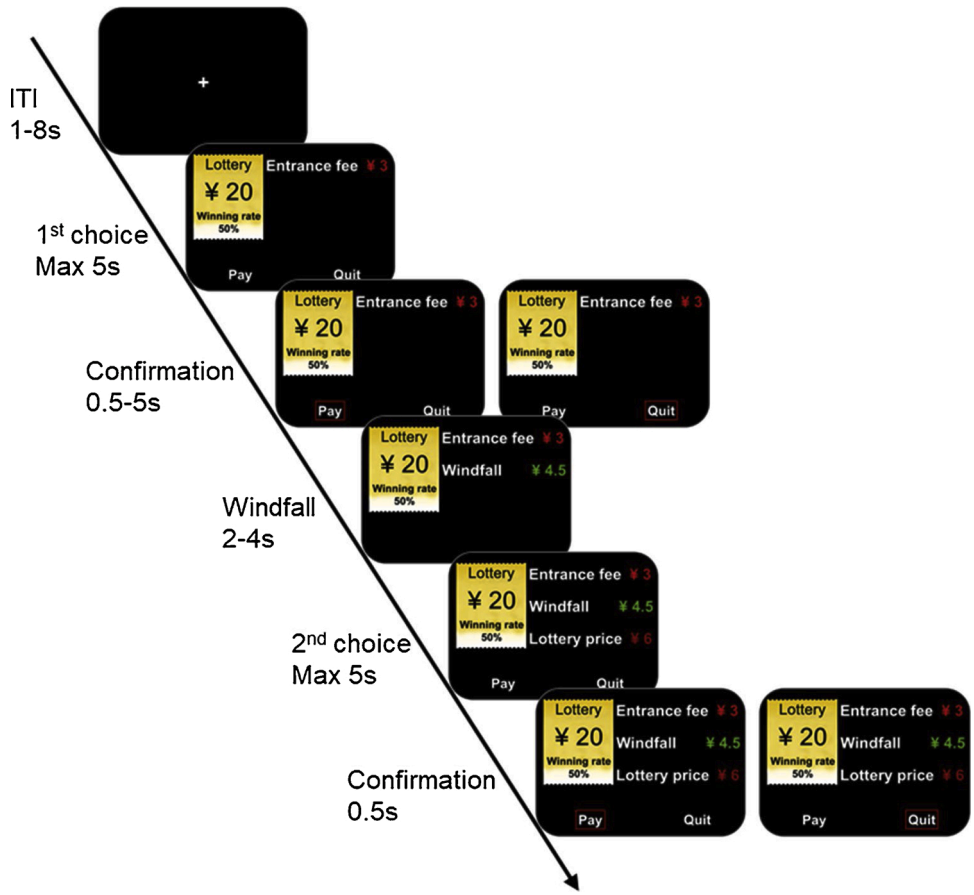

Fig. 1. Experimental procedures. Each trial consisted of two stages of choices. Subjects first had to decide whether to pay the entrance fee to earn the right to purchase a lottery ticket later in the trial. After paying the fee, an independent amount of windfall appeared on the screen before lottery ticket price was revealed and subjects could either purchase the ticket at listed price or quit the trial with the entrance fee and windfall as sunk costs and benefits, respectively.

2.4. Mental accounting model

The classic prospect theory typically defined the reference point as the '*status quo*'; whereas previous research in the mental accounting literature proposed to incorporate prior outcomes into prospect valuation, effectively shifting subject's internal reference point (Thaler, 1980, 1999). We adopted this approach and augmented the model with an additional parameter to capture subject's relative weight assigned to the values of purchase and opt-out options (M1). More formally, the difference in subjective values between purchase and opt-out options was defined as (assuming no probability distortion):

$$\Delta V_i = (1-\kappa)*\left[\sum_j P_j * V(O_{i,j}) + V(-S_i + W_i - Pr_i)\right] - \kappa * V(-S_i + W_i) \tag{1}$$

Where the indexes i indicates trial number and j indicates the potential outcome; P_j is the lottery winning probability; and $O_{i,j}, S_i, W_i$ and Pr_i refer to the potential outcome, the magnitude (absolute value) of the sunk cost (entrance fee), windfall and ticket price at trial i, respectively; and κ is the escalation parameter to capture the weighted option value difference between the option value of purchase (first term) and opt-out (second term) calculation. The value function $V(\cdot)$ in the above formula follows the traditional prospect theory format:

$$V(X) = \begin{cases} X^{\beta}, \ when\ X \geq 0 \\ -\lambda*(-X)^{\beta}, \ when\ X < 0 \end{cases} \tag{2}$$

Where parameter β is the subject's risk attitude in the gain and loss domain and the parameter λ is the loss aversion coefficient. Subjects' choice on each trial is associated with the weighted difference between option values:

$$P_i(Purchase) = (1 + e^{-\mu * \Delta V_i})^{-1} \tag{3}$$

Where μ is the inverse temperature parameter controlling choice consistency.

We took a two-step approach in estimating all the parameters in our model (β, λ, κ, μ). First, we estimated β for each subject by pooling all the trials where there was no entrance fee nor windfall (Subjects chose to pay the ¥0 entrance fee almost 100 % of the trials, Supplementary Fig. 1A). Under such condition, our task reduced into a standard risky choice task: whether subjects were willing to purchase a gamble (¥20, 50 %; ¥0, 50 %) with the listed price. Since there was no entrance fee nor windfall involved, the option value of opt-out was 0 in this situation, and we used a simplified formula (Eq. 4) here according to Eq. 3:

$$P_i(Purchase) = (1 + e^{-\mu * \left(\sum_j P_j * V(O_{i,j}) - V(Pr_i)\right)})^{-1} \tag{4}$$

Across subjects, the mean of the estimated β is 0.830, consistent with previous literatures (Kahneman and Tversky, 1979; Sokol-Hessner et al., 2009; Tversky and Kahneman, 1981). Next, we further estimated the other three parameters (λ, κ, μ) from the remaining trials where entrance fees or windfalls were not zero, using a hierarchical Bayesian algorithm (HBA) implemented in R (see below). Similar results were obtained when the model was fitted to the subjects' behavioral data all at one step.

We also tested the above model against a variety of alternative models. Firstly, we tested the classic prospect theory model where the

reference point was defined as the '*status quo*' (M2). In such a baseline model, the option value difference was calculated as follows:

$$\Delta V_i = \left[\sum_j P_j * V(O_{i,j}) + V(-Pr_i)\right] - 0 \quad (5)$$

Where the symbols had the same meanings as in Eq. 1 and the last term 0 indicated the option value of opting out. It is obvious that since the entrance fee and windfall are not present in the above equation, such a model cannot capture the effect of sunk costs.

We also compared our candidate model (M1) with the one introduced by the original mental accounting paper (Thaler, 1980). The only difference between these two models was the introduction of the escalation parameter κ in our model (M1). In the original mental accounting model (M3), the option value difference was calculated as follows:

$$\Delta V_i = \left[\sum_j P_j * V(O_{i,j}) + V(-S_i + W_i - Pr_i)\right] - V(-S_i + W_i) \quad (6)$$

Where the symbols followed the same definition as in Eq. 1. Such a model shifts subjects' reference points and effectively places them in the loss domain and therefore can account for the commitment escalation effect (Thaler, 1980).

Lastly, given that the model-free regression analysis showed that the effects of entrance fees and windfalls were not significantly different in terms of their magnitude (Fig. 2), it is possible that the loss aversion coefficient λ is not necessary. To test such a hypothesis, we simplified our candidate model (M1) by omitting parameter λ in the value function (M4):

$$V(X) = \begin{cases} X^{\beta}, \ when\ X \geq 0 \\ -(-X)^{\beta}, \ when\ X < 0 \end{cases} \quad (7)$$

The same choice function and value function (except for M4) were applied to all the models tested.

To fit models to subjects' choice behavior, we adopted the hierarchical Bayesian analysis (HBA) approach (Gelman and Hill, 2006). Under this framework, individual-level parameters (for example, β, λ, κ, μ in our mental accounting model) were samples from group level normal distributions $N(\mu_0, \sigma^2)$. To avoid the bias to the posterior distribution when sample size is small, normal and half-Cauchy distributions were used to specify the priors of the group-level means ($\mu_0 \sim$ Normal (0, 1)) and standard deviations ($\sigma \sim$ Cauchy (0, 5)), respectively. In the model fitting, all parameters were properly transformed according to their ranges (see Table S1 for details). Posterior inference for the model was performed using standard Markov-Chain Monte Carlo (MCMC) sampling methods implemented in rStan (v2.18.2) (Charpentier et al., 2017). A total of 20,000 samples were drawn after a burn-in period of 20,000 samples with four chains (5,000 samples after 5,000 burn-in samples for each chain). All model parameters had $\widehat{R}$ values less than 1.01, indicating MCMC chains had converged well (Ahn et al., 2017). The Deviance Information Criterion (DIC) was used in the Bayesian model comparison (BMC) to assess the performance of all the models and we reported the protected exceedance probabilities (PXP) in Table S1 (Berg et al., 2004; Stephan et al., 2009). It is worth noting that in the winning model (M1), there is a moderate but significant correlation between escalation parameter κ and choice consistency parameter μ (r = -0.491, p = 0.003). However, all the results reported remained similar after controlling for μ as a confounding covariate.

To test the validity of our mental accounting model (M1), we also performed model parameter recovery and cross-validation prediction analyses. For the parameter recovery analysis, we first selected all subjects' valid trials (where participants paid entrance fees) as the whole dataset trials. For each parameter, we selected 10 values uniformly in a range slightly larger than the corresponding model fitting results ($\kappa \in [0,\ 0.9]$; $\lambda \in [0.6,\ 1.5]$; $\beta \in [0.4,\ 1.3]$; $\mu \in [1,\ 10]$). We then simulated subjects' choices based on the combination of parameters we specified as true parameter values ($\kappa_0, \lambda_0, \beta_0,\ \mu_0$) and fitted the mental accounting model (M1) in each of these datasets using the maximum likelihood estimation (MLE) method to obtain the recovered parameters ($\kappa_1, \lambda_1, \beta_1,\ \mu_1$), respectively. Next, to test the model adequacy, we also performed the cross-validation prediction analysis. Specifically, for each subject, we randomly selected half of the valid trials to estimate the model parameters (estimation trials) and then used the estimated model parameters to predict subjects' choices in the remaining other half valid trials (test trials). This process was repeated 20 times for each subject and we measured the predictive accuracy by calculating the prediction accuracy index (proportion of correctly predicted choices in the test trials) for all the candidate models described previously.

2.5. Behavioral statistical analysis

Statistical analyses of behavioral data were performed with R v3.5.2 (http://www.r-project.org) and MATLAB® R2010b and our behavioral analysis focused only on the valid trials. Package 'lme4′ in R (https://cran.r-project.org/web/packages/lme4) was used to fit the linear mixed-effects regression model (Bates et al., 2015). In our analyses, the dependent variable was subject's 2nd choice and the independent variables included the entrance fee, windfall, lottery ticket price and the interaction term between the entrance fee and windfall (Fig. 2A). Each independent variable had a fixed effect across all subjects and a random effect for each subject. All p-values reported were two-tailed.

2.6. MRI data acquisition

Our MRI scanning was performed on a Siemens Prisma 3 T scanner at the Center for MRI Research of Peking University. The blood oxygen level dependent (BOLD) signals were collected by the T2*-weighted echo-planar images (EPIs) sequence with the repetition time (TR) of 2 s, echo time (TE) of 30 ms and the flip angle of 90 ° together with the matrix size = 64 × 64 and the field-of-view = 224 mm. The in-plane resolution was 3.5 × 3.5 mm with a total of 33 slices and the slice thickness of 3.5 mm, and the interslice gap of 0.7 mm. To reduce the fMRI signal dropout in the basal frontal area and the medial temporal regions, a tilted plane of acquisition method was used such that all slices were acquired 30 ° clockwise to the anterior commissure-posterior commissure axis and the phase encoding direction was set from the posterior to the anterior (Deichmann et al., 2003). The experiment was split into two sessions with a brief break of ~60 s between sessions. A T1-weighted MP-RAGE image was also obtained for each subject with the following parameters: TR = 2530 ms, TE = 2.98 ms, flip angle = 7 °, matrix size = 256 × 256, field-of-view = 256 mm, 192 slices/stab, slice thickness = 1 mm, and no interslice gap. A field map using a gradient echo (GRE) sequence was also collected for each subject and was later used to correct the magnetic field distortion.

2.7. fMRI data preprocessing

Functional imaging data were analyzed using SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK). The EPI images were first slice timing corrected using the midpoint slice as the reference slice. The images were then realigned and corrected for field distortion using field maps collected earlier. They were further co-registered to the T1-images after which T1-images were segmented using default tissue probability maps and normalized to the standard Montreal Neurological Institute (MNI) template via the 12-parameter affine transformation. Normalized functional images were then resampled into the 2 × 2 × 2 mm resolution. Finally, a Gaussian kernel of 6 mm full-width at half-maximum (FWHM) was applied to the normalized images for spatial smoothing.

2.8. General linear model (GLM) definition

For statistical analyses, we focused on the psychological processes when subjects decided between lottery purchase and opt-out choices and constructed four different GLMs with different parametric regressors at the onset of ticket purchase decisions. First, GLM1 was established to investigate how the general sunk cost (entrance fee – windfall) was represented in the brain activity and we constructed the general sunk cost regressor for the whole brain analysis (Fig. 3A). Next, we set up GLM2 to specifically test the neural correlate of chosen option value, a variable standard in value-based decisions, by including chosen option value predicted by the model (Eq. 1) as a parametric regressor (Fig. 4A). To control for the potential confounds, we also orthogonalized chosen option value against general sunk cost (Fig. 4B). We further defined a choice difficulty regressor as the negative absolute value of the option value difference (GLM3, Supplementary Fig. 3B). Finally, we tested how the idiosyncratic escalation parameter κ was represented in the brain across subjects. We used $V(-S_i + W_i)$ as the parametric regressor (Eq. 1) and searched the whole brain for neural correlates of κ with beta images generated from above 1st-level analysis (GLM4, Fig. 5A).

For the above four GLMs, we also included other regressors-of-no-interest to control for the overall variance in the fMRI data. These regressors included the onset of the entrance fee revelation, modulated by the entrance fee amount; subjects' first decision onset in each trial; the onset of the windfall revelation, modulated by the amount of windfall; time onset of the ticket price revelation, modulated by the amount of ticket price as well as six head movement parameters.

For all these analyses, we convolved regressors specified in the above GLM design matrix with SPM12′s canonical hemodynamic response function (HRF) and computed parameter estimates for each subject and took these estimates to the group level random-effect analysis for statistical inference. For all the whole-brain analyses, we tested for statistical significance using whole-brain correction ($P < 0.05$, false discovery rate (FDR)-corrected at the cluster level, initial cluster-forming height threshold $Z > 2.6$).

2.9. Region-of-interest analyses

We used region-of-interest (ROI) analysis to illustrate the individual difference and the time course of brain activations. For the time series analysis (Figs. 3B and 4 C), the ROI analysis was performed to further test whether the results (Figs. 3A and 4 A, B) depended on the specific shape of the HRF. To achieve this, BOLD signals extracted from each ROI were firstly realigned to the ticket purchase decision onset across trials and then detrended and resampled with 100 ms temporal resolution before regressed against the variables of interest using the linear interpolation method ('interp1′ function in Matlab). In Fig. 3B, general sunk cost was the only regressor for the ROI BOLD signals (Blue). In a separate regression, we included both the entrance fee (Red) and windfall (Yellow) to test their unique contribution to the ROI BOLD responses. In Fig. 4C, we included both chosen option value (red) and general sunk cost (blue) as separate regressors in the regression analysis. Group averaged effect sizes (beta values from the linear regression) at each time point were then reported with their standard errors across subjects. For all the other ROI analyses (Figs. 4D, 5 and 6), ROIs were defined based on the corresponding GLM contrasts and separate hypotheses were tested from brain signals in the ROIs.

2.10. Psychophysiological interaction analyses

we examined functional connectivity from seed regions of the left dlPFC and rSPL identified in Fig. 5A to test which brain regions' activity was associated with these brain areas in a κ dependent manner. To this end, we first extracted the deconvolved time series from the left dlPFC and rSPL as two separate seed regions (Fig. 5A). The psychophysiological interaction (PPI) term was then calculated as the element-by-element product of the time courses from each brain area and a vector coding for the lottery purchase choice (psychological vector, 1 for the onset of lottery purchase choice, -1 for the fixation onset in each trial). For each subject, we created PPI models that contained 3 regressors: physiological, psychological and PPI vectors. Finally, the beta images of PPI from each participant were entered into the whole brain analysis to search for the brain correlates of PPI that correlated with the escalation parameter κ across subjects.

3. Results

3.1. Influence of prior costs and benefits on lottery purchase decision

After paying the entrance fee in each trial, subject could either purchase the ticket with the listed price or quit (Fig. 1). We independently manipulated the amount of entrance fee, windfall as well as lottery ticket price to delineate their individual effect on subject's lottery purchasing behavior. Critically, when subjects contemplated about lottery ticket purchase in each trial, their entrance fee (sunk cost) and windfall (sunk benefit) were already "sunk" (non-recoverable). Therefore, this design provided us with the unique opportunity to quantitatively study how sunk costs/benefits would influence subjects' risky choices.

Behaviorally, subjects were sensitive to the magnitude of the entrance fee: their willingness to participate in a certain trial significantly decreased as the amount of entrance fee increased ($\beta = -1.867$, $P < 0.001$, 95 % confidence interval (CI) = [-2.238, -1.497]) and they on average paid entrance fee in 83.4 % of the trials (Supplementary Fig. 1A). We then assessed the impact of sunk costs (entrance fee), sunk benefits (windfall) and lottery ticket price on lottery ticket purchase using the mix-effect logistic regression. As expected, the ticket price was negatively correlated with subjects' purchase behavior ($\beta = -0.970$, $P < 0.001$, 95 % CI = [-1.098, -0.841]), indicating that subjects were sensitive to the price of the lottery ticket. Surprisingly, our subjects showed an overall de-escalation effect of entrance fees ($\beta = -0.290$, $P < 0.001$, 95 % CI = [-0.448, -0.133]): as the magnitude of entrance fees increased, participants were less likely to purchase the ticket, in stark contrast to the commonly observed escalation effect (Soman and Cheema, 2001; Thaler, 1980, 1985, 1999) but consistent with the mental accounting theory that included both prior and future outcomes in the same category of expenditure (Heath, 1995; Thaler, 1999). On the other hand, higher windfall prompted people to purchase the lottery more often ($\beta = 0.366$, $P < 0.001$, 95 % CI = [0.210, 0.523]; Fig. 2A), consistent with the well-documented house money effect (Thaler and Johnson, 1990), and the interaction between the entrance fee and windfall was not significant ($\beta = -0.005$, $P = 0.618$, 95 % CI = [-0.025, 0.015]). Across subjects, there was a substantial amount of heterogeneity of the entrance fee effect on choice behavior (Fig. 2B): 8 of 35 subjects showed escalation effects of the entrance fee whereas the others had de-escalation effects. Furthermore, the entrance fee and windfall seemed to exert comparable (paired *t*-test, $t_{34} = 1.426$, $P = 0.163$; Fig. 2C inset) but opposite ($r = -0.744$, $P < 0.001$, 95 % CI = [-0.863, -0.545]; Fig. 2C) effects on subjects' choices. This result suggested that contrary to the over-weighting of losses relative to gains typical in prospects valuation, past costs (entrance fee) and benefits (windfall) tended to weigh similarly and it was in line with previous studies that indicated the sunk costs and sunk benefits could be integrated into the same mental account (Soman and Cheema, 2001). We thus combine the entrance fee and windfall and use the term 'general sunk cost' (entrance fee – windfall) in the following modeling analysis. The effect of general sunk cost on subject's choices remained similar to that of the simple entrance fee across subjects (Supplementary Fig. 1, B and C). To rule out the possibility that subjects' sensitivity towards general sunk cost was caused the lack of attention to the fixed gamble lottery face value, we also conducted a pilot study where we simultaneously manipulated entrance fee, windfall, lottery price and, lottery face value (see Supplementary information). Consistent with the results from

the main study, subjects' lottery purchase behavior was negatively associated with the entrance fee and lottery price but positively related to the windfall (Supplementary Fig. 1D). In the meantime, subjects were more likely to gamble with increasing lottery face value, indicating they were sensitive to the gamble lottery (Supplementary Fig. 1D).

3.2. Behavioral analysis using mental accounting framework

To account for the individual difference in subjects' choice behavior, we developed a quantitative mental accounting model to integrate the entrance fee and windfall into the valuation process of prospects, as suggested previously (Kahneman and Tversky, 1979; Thaler, 1980). The key idea in our model is conceptualized in the subject specific escalation parameter κ to capture the relative weight subjects assign to the choice options when option values are compared against each other. The option value difference is a weighted sum of opt-in and opt-out option values (Crockett et al., 2017). Under classic prospect theory where the reference point is assumed to be the status quo, such a parameter would not influence a subject's behavior since the opt-out option (reference option) value is always 0. However, under our mental accounting framework, the value of the reference option is recalibrated by the general sunk cost. Intuitively, when κ is close to 0, subjects mainly focus on the prospect of purchasing the lottery ticket and the sunk costs are simply added to the ticket price to augment the overall costs, thus rendering subjects less likely in lottery purchases (de-escalation effect) (Heath, 1995; Thaler, 1999). When κ approaches 0.5, however, both option values are weighed similarly and subject's choice is contingent on the comparison between the values of the purchase prospect and the reference option, as suggested by previous literatures (Thaler, 1980). As κ further increases towards 1, subjects mainly concern about the prospect of opt-out option and therefore tend to purchase tickets more often under the higher sunk cost condition (escalation effect) due to the loss frame induced by the reference recalibration of the sunk cost. The introduction of such escalation parameter κ captures subjects' tendency in the commitment escalation and is thus termed the escalation parameter. Similar ideas about relative weighting on option valuation have also been applied in recent eye-tracking research on value-based decisions (Krajbich et al., 2010) and neuroimaging studies of sunk cost effects (Fujino et al., 2016; Haller and Schwabe, 2014).

We fitted our model to subjects' behavioral data using a hierarchical Bayesian algorithm (HBA) and found that the escalation parameter κ estimated from each subject positively correlated with general sunk cost coefficient from aforementioned behavioral regression analysis ($r = 0.905$, $P < 0.001$, 95 % CI = [0.819, 0.952]; Fig. 2D and Supplementary Fig. 2), indicating that κ indeed captured the direction and magnitude of the general sunk cost effect. Interestingly, individual escalation parameter κ also positively and negatively correlated with the effects of the entrance fee and windfall, respectively (Supplementary Fig. 2B), suggesting that κ also reflected separate sunk costs and benefits effects. Finally, subjects' reaction time (RT) was positively correlated with choice difficulty, which was defined as the negative absolute value of the option value difference ($r = 0.340$, $t_{34} = 14.772$, $P < 0.001$, Supplementary Fig. 3A). We then tested the performance of our model against a variety of alternative models and confirmed its superior performance in explaining choice behavior across participants (Supplementary Table 1). Furthermore, the model parameter recovery analysis showed that all the model parameters could be recovered well from simulated choice dataset with sufficient true parameter ranges in our mental accounting model (M1, Supplementary Fig. 2D). Likewise, the cross-validation analysis also demonstrated that our model (M1) performed the best among all the candidate models in terms of predictive accuracy ($Ps < 0.001$, Supplementary Fig. 2E).

3.3. Sunk cost representation and option evaluation in the brain

Our mental accounting model is predicated on the assumption that people lodge prior outcomes and integrate them with future prospects when facing risky choices. To test this hypothesis at the neural level, we

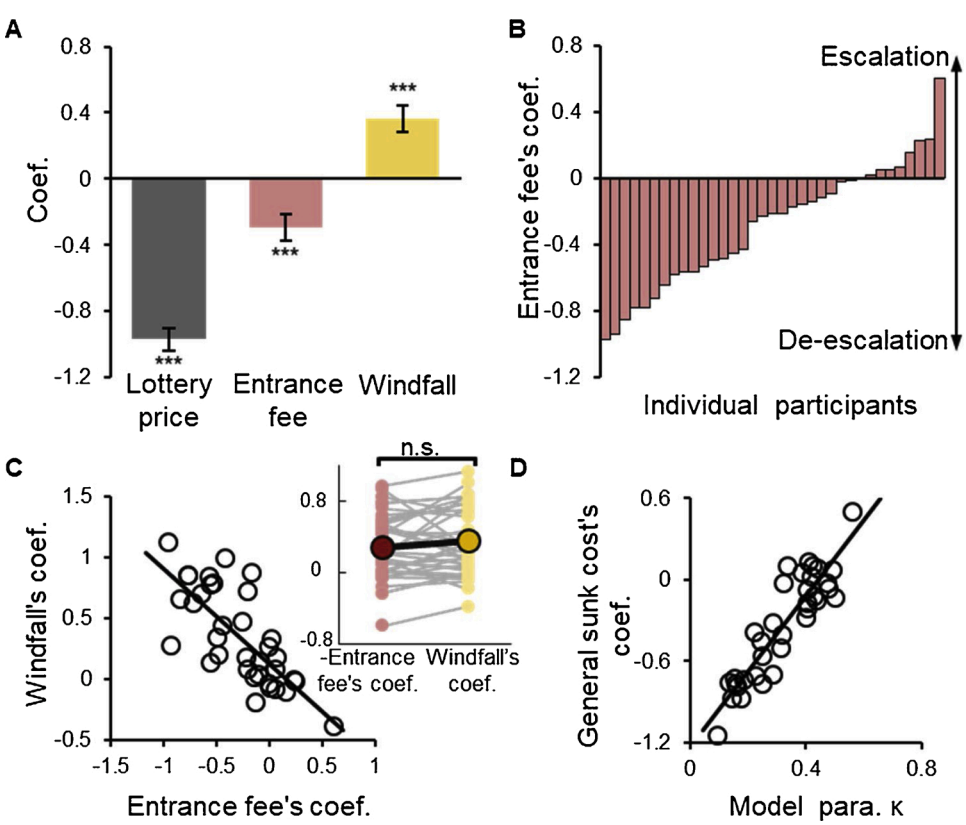

Fig. 2. Behavioral results. (**A**) Logistic regression on subjects' second choices. Error bars represented SEM. *** $P < 0.001$. (**B**) Individual difference of entrance fee's effects in Fig. 2A. The sign of the coefficient indicates whether the entrance fee had an escalation (> 0) or de-escalation (< 0) effect. (**C**) The coefficients of entrance fee and windfall were negatively correlated across subjects. Inset: The entrance fee and windfall had comparable but opposite effects on the lottery purchase choice. n.s., not significant. (**D**) Model estimated escalation parameter κ captured the direction and magnitude of the general sunk cost effects.

constructed a general linear model (GLM) to identify brain areas whose blood oxygen level-dependent (BOLD) activities correlated with general sunk cost at the lottery decision onset. We found that the BOLD activity in the ventral striatum (VST) was negatively correlated with the general sunk cost (whole brain FDR-corrected $P < 0.05$; Fig. 3A and Supplementary Table 2). A time series analysis including the entrance fee and windfall as separate regressors also revealed that the VST BOLD signals were negatively and positively correlated with the entrance fee and windfall, respectively (Fig. 3B) and such correlations did not depend on the specific shape of the HRF. Previous studies have showed that both experienced and prospective gains and losses engaged activities in the brain reward processing structures such as the striatum and the orbitofrontal cortex (Tom et al., 2007). Our imaging results thus corroborated behavioral results by illustrating that the prior outcome information was registered in the striatum activities while subjects faced lottery choice; even though according to the classic economic theory, such consideration was not necessary.

We further reasoned that if our mental accounting model was implemented at the neural level, then BOLD signals in the brain valuation regions should encode participants' internal representation of preferred option value at the decision onset. This was indeed the case: Using the model predicted chosen option value as the parametric regressor, we found significant positive correlations between activities of both the striatum and the medial orbitofrontal cortex (mOFC) and chosen option values (whole brain FDR-corrected $P < 0.05$; Fig. 4A and Supplementary Table 3). Since the chosen option value was mildly correlated with the general sunk cost and the striatum had been shown to encode general sunk cost (Fig. 3), we further included the general sunk cost as a regressor of no interest in the above GLM. The activities of mOFC still positively correlated with the chosen option value after controlling for the effect of the entrance fee and windfall (Fig. 4, B and C). Interestingly, the mOFC seemed to serve a dual function: in addition to encoding the chosen option value across subjects, an exploratory regression analysis showed that its activity was also negatively correlated with escalation parameter κ (r = -0.352, P = 0.038, 95 % CI = [-0.613, -0.021]; Fig. 4D, orange). However, the same pattern was not observed in the VST (r = 0.141, P = 0.418, 95 % CI = [-0.201, 0.453]; Fig. 4D, gray; interactions between the escalation parameter κ and brain regions (mOFC and VST): β = -0.048, P = 0.032, 95 % CI = [-0.092, -0.004]).

Furthermore, we found BOLD signals in the bilateral insula and anterior cingulate cortex (ACC) positively correlated with the choice difficulty (whole brain FDR-corrected $P < 0.05$; Supplementary Fig. 3B and Supplementary Table 4), consistent with the roles of these two brain areas in conflict monitoring and adjustment in cognitive control (Kerns et al., 2004; Shenhav et al., 2016).

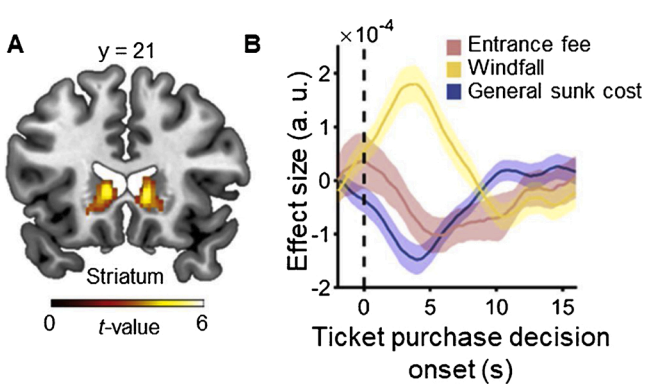

Fig. 3. Neural responses to the sunk cost and windfall. (**A**) The magnitude of the general sunk cost negatively correlated with BOLD signals in the ventral striatum (FDR-corrected $P < 0.05$). (**B**) Time course of the regression coefficients between the VST activity and the general sunk cost as the only regressor (blue), or with the entrance fee (red) and windfall (yellow) as separate regressors. a.u., arbitrary units, mean ± SEM.

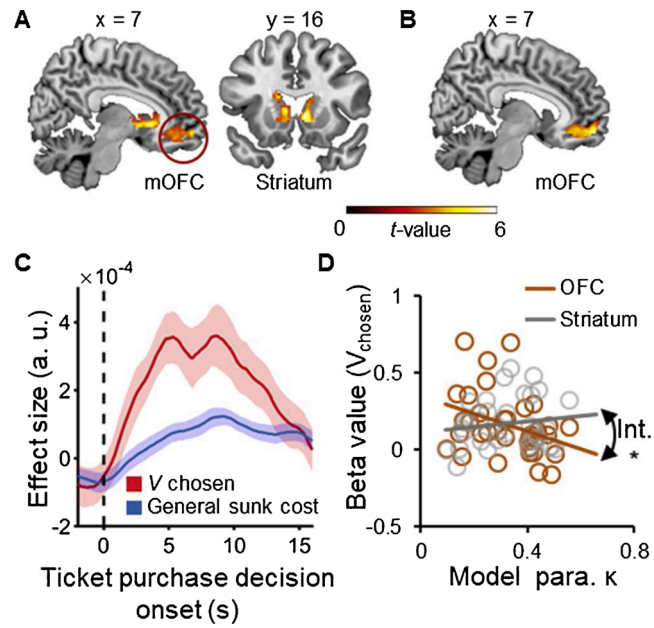

Fig. 4. Neural representation of the chosen option value. (**A**) The chosen option value was represented in the activities of brain areas including the striatum and the mOFC (FDR-corrected $P < 0.05$). (**B**) BOLD signals in the mOFC still positively correlated with the chosen option value after controlling for the effect of general sunk cost (FDR-corrected $P < 0.05$). (**C**) The time course of the regression coefficient between the mOFC activity and the chosen option value (red) and general sunk cost (blue) as separate regressors. a.u., arbitrary units, mean ± SEM. (**D**) Exploratory analysis showed that the mOFC activities negatively correlated with the escalation coefficient κ (orange), which was absent in the striatum activities (gray), and the escalation parameter κ and brain regions (mOFC vs. Striatum) interaction was significant. * $P < 0.05$.

3.4. Neural accounts of the individual susceptibility to the sunk cost

As shown above, our model hypothesized that individual differences in subjects' susceptibility to the sunk cost were embodied in the escalation parameter κ. We then set out to identify the neural representation associated with the parameter κ across subjects. Specifically, we performed a whole brain analysis to identify brain areas whose reference option value related neural activities showed significant correlation with κ across subjects. We found significant negative correlation between BOLD activities in the left dlPFC and right SPL (rSPL) and κ (whole brain FDR-corrected $P < 0.05$; Fig. 5A and Supplementary Table 5). Further region-of-interest (ROI) analysis confirmed that subjects with smaller κ showed stronger reference option value related activities in the left dlPFC (r = -0.658, $P < 0.001$, 95 % CI = [-0.813, -0.416]; Fig. 5B, orange) and rSPL (r = -0.621, $P < 0.001$, 95 % CI = [-0.791, -0.363]; Fig. S4A, orange). However, such correlation was not observed between the escalation parameter κ and activities in the striatum (r = -0.022, P = 0.899, 95 % CI = [-0.353, 0.313]; Fig. 5B and Supplementary Fig. 4A, gray), previously identified to encode the general sunk cost (Fig. 3A). Moreover, this brain activation and κ relationship differed significantly between the striatum and the left dlPFC (interaction β = -0.074, $P < 0.001$, 95 % CI = [-0.109, -0.040]; Fig. 5B), and between the striatum and the rSPL (interaction β = -0.079, $P < 0.001$, 95 % CI = [-0.118, -0.040]; Supplementary Fig. 4A). These results suggested that activities of the striatum only registered prior costs and benefits but were not involved in representing individual difference in the escalation tendency. Instead, we speculated that activities of the dlPFC and SPL might exert their effects by modulate brain processes in the valuation areas such as mOFC and striatum in accordance with subjects' idiosyncratic escalation parameter κ, based on previous literatures indicating that the dlPFC and SPL are heavily involved in functions such as goal maintenance and cognitive control.

3.5. Functional connectivity between the cognitive control and valuation systems

To test the above hypothesis, we next investigated whether the left dlPFC and rSPL, presumably indicating the level of cognitive control, and mOFC exhibited task-related functional connectivity; and furthermore, whether this connectivity was also associated with individual differences of the escalation parameter κ. We performed psychophysiological interaction (PPI) analyses separately using the left dlPFC and the rSPL as seed regions (see Methods for details). These analyses revealed that the functional connectivity between the left dlPFC (as well as the rSPL) and the mOFC was indeed correlated with the individual parameter κ (whole brain FDR-corrected $P < 0.05$; Fig. 6A, Supplementary Fig. 4B and Supplementary Table 6). Moreover, the mOFC area identified by the PPI analysis and previous chosen option value GLM analysis share significant overlap (Fig. 6A and Supplementary Fig. 4B). Further ROI analysis confirmed such correlation, with smaller κ corresponded with more negative interaction (inhibition) between the left dlPFC and the mOFC ($r = 0.484$, $P = 0.003$, 95 % CI = [0.180, 0.704]; Fig. 6B, green), and between the rSPL and the mOFC ($r = 0.472$, $P = 0.004$, 95 % CI = [0.165, 0.696]; Supplementary Fig. 4C, green). Although the striatum was also involved in encoding chosen option value in our task, we did not find κ related functional connectivity between the left dlPFC and the striatum ($r = 0.210$, $P = 0.226$, 95 % CI = [-0.132, 0.508]; Fig. 6B, blue), or between the rSPL and the striatum ($r = 0.288$, $P = 0.093$, 95 % CI = [-0.050, 0.567]; Supplementary Fig. 4C, blue). Importantly, the ROI analysis revealed that there was a significant interaction effect between the escalation parameter κ and brain regions (striatum and the mOFC) in terms of their functional connectivity with the left dlPFC (interaction $\beta = 0.267$, $P = 0.011$, 95 % CI = [0.062, 0.472], Fig. 6B) and the right SPL(interaction $\beta = 0.179$, $P = 0.031$, 95 % CI = [0.017, 0.342]; Supplementary Fig. 4C), suggesting a functional segregation between these two areas.

4. Discussion

It is well established that human choices are susceptible to the contexts where the decisions are made (Cox and Kable, 2014; Gluth et al., 2017; Kahneman and Tversky, 1979; Louie et al., 2013; Rigoli et al., 2016; Soltani et al., 2012; Tversky, 1972). Classic decision models such as the prospect theory focused on how the prospects were explicitly framed relative to the reference point, whereas other frameworks were more concerned with how the decision problems were internally represented (Gluth et al., 2017; Louie et al., 2013; Rigoli et al., 2016; Soltani et al., 2012). For example, the mental accounting theory emphasized the manner in which prior and marginal outcomes were lodged, internally registered, and integrated to bias choice preference (Kahneman and Tversky, 1979; Thaler, 1985, 1999).

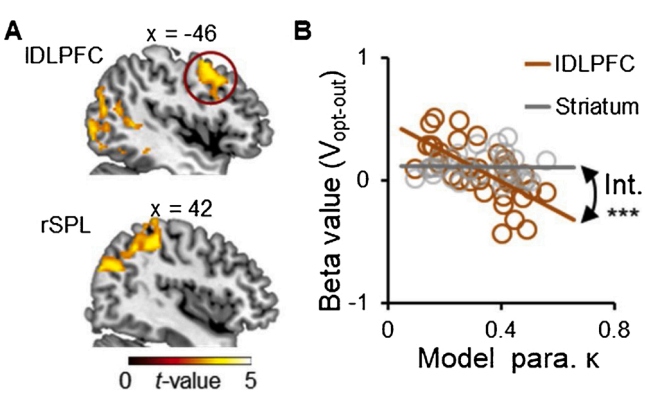

Fig. 5. Neural representation of the escalation parameter κ. (**A**) The opting out option value related brain activities in the rSPL and left dlPFC negatively correlated with the escalation parameter κ (FDR-corrected $P < 0.05$). (**B**) Opting out option value related left dlPFC activities were negatively associated with smaller κ (orange), which was absent in the activity pattern in the striatum (gray), and the brain regions (lDLPFC and Striatum) x parameter κ interaction was also significant. *** $P < 0.001$.

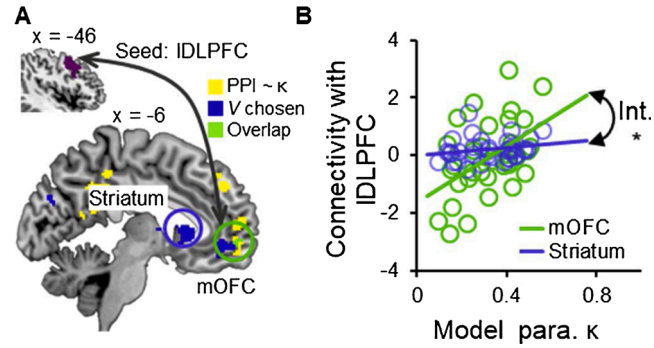

Fig. 6. Connectivity between the dlPFC and mOFC associated with escalation parameter κ. (**A**) The decision-specific connectivity between the left dlPFC and the medial prefrontal cortex was positively correlated with model parameter κ (yellow). And the above brain areas overlapped with those defined by the chosen value (blue) regressor at the mOFC (green). (**B**) Smaller κ corresponded with more negative interaction between the left dlPFC and the mOFC (green), which was absent in the connectivity between dlPFC and striatum (blue), and the escalation parameter κ and brain regions (mOFC and Striatum) interaction was significant. * $P < 0.05$.

Extant data support the idea that prior outcomes are lodged in the mental account but how these outcomes are registered and influence subjects' choice preference remain controversial in humans and other species (Duxbury, 2012; Heath, 1995; Sweis et al., 2018; Thaler, 1980). Although the majority of literatures show the escalation of commitment effect of the sunk cost, other research provides opposite results (de-escalation effect), especially when prior and marginal costs are of the same modalities (Heath, 1995; Kahneman and Tversky, 1984; Thaler, 1999) and belong to the same ongoing progress (Roth et al., 2015). Our behavioral results also revealed an overall de-escalation effect of sunk costs with significant individual differences (Fig. 2B). Such results are consistent with the mental budget theory proposed previously, where subjects either set a mental budget implicitly based on the expected benefit, or, explicitly due to strict budget constraint for specific account (Heath, 1995). In our experiment, subjects were endowed with ¥18 to play the game before each trial and the expected value for the lottery was easy to calculate. Therefore, it is possible that the de-escalation effect of sunk cost we observed was caused by subjects' reluctance to exceed both the explicit mental budget (¥18) and the implicit budget (expected value of the lottery) by tracking their total costs (sunk cost + lottery price). Interestingly, we also found a parallel escalation effect of prior benefits (windfall), both of which could be captured by how the reference point was recalibrated by the general sunk costs in our computational model. In our experiment, windfall served two purposes: encouraging subjects to voluntarily incur sunk cost (the expectation of windfalls if entering the game) such that the sunk cost effect can be delineated from the agency effect (Thaler, 1980), and independently investigating the prior benefit effects. We designed the experiment such that prior costs and benefits were independent and their distinct effects can be examined. The linear regression analysis and our modeling approach showed that the effects of prior costs and benefits have opposite signs but comparable magnitudes (Fig. 2C inset, Supplementary Fig. 2C). Previous research also showed that the reminder of past positive experience biased subjects' decisions for reward (Bornstein et al., 2017; Bornstein and Norman, 2017). Our results indicated that both positive and negative past experiences weigh on participants' choices with similar magnitudes, in stark contrast to the loss aversion phenomena when people evaluate the prospect of losses and gains (Kahneman and Tversky, 1979; Martino et al., 2006). Additionally, our

results also differed from the optimism bias (Sharot et al., 2007), confirmation bias (Doll et al., 2011) and the asymmetric reinforcement learning (Dorfman et al., 2019; Lefebvre et al., 2017; Niv et al., 2012) literatures where positive (congruent) events tended to have a bigger effect than the negative (incongruent) ones. One possibility might concern the nature of the mental accounting. The mental accounting theory postulates that after the joint outcomes (positive & negative) occur, the outcomes can be integrated first before being subscribed to a usually non-linear value function if the joint outcomes belong to the same mental account, instead of being evaluated separately first and yielding different effects based on the outcome valence or congruency. In our model, due to the similar magnitude but opposite direction effects of the entrance fee and windfall (Fig. 2C inset), these two variables were directly integrated in the value difference function (Eq. 1). According to the hedonic editing rules (Thaler, 1985), it is possible that participants may prefer more nuanced value integration or segregation rules when it comes to evaluating the option value for purchasing the ticket than the function we proposed (Eq. 1). More dedicated studies are required to answer such questions in the future. For example, whether the general sunk cost and the lottery price should be evaluated separately, or whether the general sunk cost should be taken into account when considering the lottery valuation itself. Nevertheless, the inclusion of sunk cost and sunk benefit in the valuation function seems to be necessary and our model provides a benchmark against which future models can be tested.

Additionally, in our quantitative mental accounting model, the escalation parameter κ captured the relative weights subjects assigned between "paying for the lottery ticket" and "opt-out" option values. Such an extension to the classic mental accounting theory allowed us to account for the individual differences across subjects, especially the de-escalation sunk cost effect. It is worth mentioning that the rescaling of the option values by the parameter κ, as implemented in our model, differs dramatically from simply scaling the magnitude of the sunk costs. Rescaling the size of sunk cost only changes the magnitude of the sunk cost effect in the direction of commitment escalation (Thaler, 1980). The introduction of the escalation parameter κ, however, allowed our model to yield opposite effects of the sunk costs (escalation or de-escalation), depending on the idiosyncratic value of the parameter κ for individual subject (Haller and Schwabe, 2014; Heath, 1995; Soman and Cheema, 2001; Thaler, 1980). Our model was further validated by subjects' choice behavior under the influence of sunk benefits (Fig. 2A and Supplementary Fig. 2C). There are two critical assumptions in our model. First, in contrast to the marginal utility theory, we hypothesized that prior costs and benefits were lodged into the mental account and altered the reference point. Second, this recalibration of the reference point provided a hinge upon which the escalation parameter κ could act (the opt-out value is always 0 under the marginal utility theory), and hence determined the direction and amplitude of the sunk cost effects. Similar ideas have been proposed in previous reinforcement learning and decision-making literatures to reflect the imbalanced cognitive processing from different information sources (Crockett et al., 2017; Krajbich et al., 2010; Li and Daw, 2011). Our model was also consistent with recent neuroimaging research that proposed the sunk cost effect was determined by the degree to which subjects followed pre-established principles such as the rule of "do not waste" (Fujino et al., 2016; Haller and Schwabe, 2014; Staw and Ross, 1989). In our task, opting out of the game after incurring sunk cost would violate such principle and the parameter κ could be interpreted as controlling how likely subjects stick to the principle of "not appearing wasteful", thus engendering the escalation or de-escalation effect.

Consistent with our behavioral results, sunk costs and benefits were encoded in the brain activities of the ventral striatum, which receives heavy projections from midbrain dopaminergic neurons and plays a critical role in prospect valuation, reinforcement learning and reward processing (Kable and Glimcher, 2009; Rangel et al., 2008). Interestingly, VST activity only tracked the sunk cost and benefit information, but not the degree to which subject weighed on this information (parameter κ). In line with this observation, although striatum responses were also involved in representing the chosen option value, they were not related to the individual difference of suck cost effects. Neither did the striatum embody the individual (de-)escalation propensity in its connectivity with the cognitive control network brain regions such as the dlPFC and SPL (Fig. 6 and Supplementary Fig. 4). Taken together, these results suggested that striatum only registered sunk cost information and further represented the chosen option value across subjects, but did not carry information about individual difference in the option evaluation, a function presumably undertaken by the mOFC.

The mOFC seems to be the ideal brain structure to integrate the inputs from ventral striatum and other brain regions to encode the decision value. Indeed, BOLD responses in the mOFC tracked the chosen option value in our task, consistent with previous literatures suggesting the mOFC as the neural integrator for subjective value and choice preference (Hare et al., 2009; Kable and Glimcher, 2007). In contrast to that of the ventral striatum, the neural representation of the chosen option value in the mOFC also tracked individual differences in sunk cost effects, suggesting that escalation parameter κ was not only involved in the mOFC valuation calculation, but also directly related to the neural representation itself. Furthermore, across individuals, the escalation parameter κ was also correlated with BOLD signals in the left dlPFC and rSPL, two critical brain components of the cognitive control network (Cole et al., 2013; Esterman et al., 2009) that are heavily involved in self-control, norm compliance and more broadly, goal-directed behavior (Buckholtz, 2015; Hare et al., 2009; Ruff et al., 2013). Importantly, these two brain areas also showed decision-specific functional connectivity with the mOFC, but not the ventral striatum, in an escalation parameter dependent manner, further indicating the functional segregation between the ventral striatum and the mOFC (Fig. 6A and Supplementary Fig. 4B). Therefore, it is tempting to hypothesize that the ventral striatum is responsible for lodging the sunk cost information and conveying such information for option valuation; whereas the mOFC is mainly for decision valuation. Also, due to the anatomical and functional projections it receives from the cognitive control network, the mOFC activities have also been suggested to relate to behavioral traits (Levy and Glimcher, 2011; Li et al., 2019; Tom et al., 2007). Such conjecture is also consistent with previous theories that the value-based decision might involve intricate interactions between cognitive control brain areas such as dlPFC and valuation system regions such as OFC (Aron et al., 2004; Haller and Schwabe, 2014; Hare et al., 2009). Finally, our results also resonated well with recent studies showing that the transcranial direct current stimulation (tDCS) over the dlPFC increased the impact of past expenses on subsequent decision-making (Bogdanov et al., 2017) and highlighted the pivotal role of the mOFC as the final gateway for value-based choices (Montague et al., 2004; Rangel et al., 2008).

5. Conclusions

In the current study, with a two-step choice task, we provided both the computational and functional imaging evidence to support the operation of a quantitative mental accounting model to capture the effects of prior outcomes on subjects' decision-making, both within and across subjects. To our knowledge, this is the first study to provide neural computational interpretation of sunk cost effects under the mental accounting framework. At the group level, prior costs and benefits de-escalated and escalated participants' commitment to the lottery purchase behavior, respectively. Such behavioral effects were captured by the novel escalation parameter κ in our mental accounting model. Corresponding to the behavioral effect, we showed that prior costs and benefits were represented by the ventral striatum BOLD activities. Such a neural representation of sunk costs and benefits was then integrated into the calculation of the chosen option value, which was encoded by the mOFC activities. The left dlPFC and rSPL activities, on the other hand, carried information about the escalation parameter κ and influenced the

direction and magnitude of sunk cost effects via their functional connectivity with the mOFC. Such a model, together with its neural instantiation, would also shed lights to future studies such as whether and how prior outcomes can be deposited into different mental accounts, especially when they are of different modalities (Heath, 1995), and under what circumstances these accounts are fungible (Thaler, 1999; Thaler and Johnson, 1990) to allow for more flexible decision making.

Author contributions

J.L. conceived the idea. J.L., B.Z. and S.L. designed and conducted experiments. J.W. and J.L. performed data analysis and computational modeling. J.W. and J.L. wrote the paper. All authors discussed the results and implications and commented on the manuscript at all stages.

Data availability

All data related to this paper are available upon reasonable request.

Funding

This work was supported by the Chinese National Science foundation (grant # 31871140, 32071090) and National Program on Key Basic Research Project (grant # 2015CB559200).

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. The Peer Review Overview and Supplementary data

The Peer Review Overview and Supplementary data associated with this article can be found in the online version, at doi:https://doi.org/10.1016/j.pneurobio.2021.102178.

References

Ahn, W.-Y., Haines, N., Zhang, L., 2017. Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. Comput. Psych. 1, 24–57.
Arkes, H.R., Ayton, P., 1999. The sunk cost and concorde effects: are humans less rational than lower animals? Psychol. Bull. 125, 591–600.
Arkes, H.R., Blumer, C., 1985. The psychology of sunk cost. Organ. Behav. Hum. Decis. Process. 35, 124–140.
Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. Trends Cogn. Sci. 8, 170–177.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67.
Bazerman, M.H., Giuliano, T., Appelman, A., 1984. Escalation of commitment in individual and group decision making. Organ. Behav. Hum. Perform. 33, 141–152.
Berg, A., Meyer, R., Yu, J., 2004. Deviance information criterion for comparing stochastic volatility models. J. Bus. Econ. Stat. 22, 107–120.
Bogdanov, M., Ruff, C.C., Schwabe, L., 2017. Transcranial stimulation over the dorsolateral prefrontal cortex increases the impact of past expenses on decision-making. Cereb. Cortex 27, 1094–1102.
Bornstein, A.M., Norman, K.A., 2017. Reinstated episodic context guides sampling-based decisions for reward. Nat. Neurosci. 20, 997–1003.
Bornstein, A.M., Khaw, M.W., Shohamy, D., Daw, N.D., 2017. Reminders of past choices bias decisions for reward in humans. Nat. Commun. 8, 1–9.
Buckholtz, J.W., 2015. Social norms, self-control, and the value of antisocial behavior. Curr. Opin. Behav. Sci. 3, 122–129.
Charpentier, C.J., Aylward, J., Roiser, J.P., Robinson, O.J., 2017. Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. Biol. Psychiatry 81, 1014–1022.
Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. Nat. Neurosci. 16, 1348–1355.
Cox, K.M., Kable, J.W., 2014. BOLD subjective value signals exhibit robust range adaptation. J. Neurosci. 34, 16533–16543.
Crockett, M.J., Siegel, J.Z., Kurth-Nelson, Z., Dayan, P., Dolan, R.J., 2017. Moral transgressions corrupt neural representations of value. Nat. Neurosci. 20, 879–885.
Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R., 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430–441.
Dixon, M.L., Christoff, K., 2014. The lateral prefrontal cortex and complex value-based learning and decision making. Neurosci. Biobehav. Rev. 45, 9–18.
Doll, B.B., Hutchison, K.E., Frank, M.J., 2011. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J. Neurosci. 31, 6188–6198.
Dorfman, H.M., Bhui, R., Hughes, B.L., Gershman, S.J., 2019. Causal inference about good and bad outcomes. Psychol. Sci. 30, 516–525.
Duxbury, D., 2012. Sunk costs and sunk benefits: a re-examination of re-investment decisions. Br. Account. Rev. 44, 144–156.
Esterman, M., Chiu, Y.C., Tamber-Rosenau, B.J., Yantis, S., 2009. Decoding cognitive control in human parietal cortex. Proc. Natl. Acad. Sci. U. S. A. 106, 17974–17979.
Fujino, J., Fujimoto, S., Kodaka, F., Camerer, C.F., Kawada, R., Tsurumi, K., Tei, S., Isobe, M., Miyata, J., Sugihara, G., Yamada, M., Fukuyama, H., Murai, T., Takahashi, H., 2016. Neural mechanisms and personality correlates of the sunk cost effect. Sci. Rep. 6, 1–10.
Garland, H., Sandefur, C.A., Rogers, A.C., 1990. De-escalation of commitment in oil exploration: when sunk costs and negative feedback coincide. J. Appl. Psychol. 75, 721–727.
Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge university press, Cambridge, England.
Gluth, S., Hotaling, J.M., Rieskamp, J., 2017. The attraction effect modulates reward prediction errors and intertemporal choices. J. Neurosci. 37, 371–382.
Haller, A., Schwabe, L., 2014. Sunk costs in the human brain. Neuroimage 97, 127–133.
Hare, T.A., Camerer, C.F., Rangel, A., 2009. Self-control in decision-making involves modulation of the vmPFC valuation system. Science. 324, 646–649.
Heath, C., 1995. Escalation and de-escalation of commitment in response to sunk costs: the role of budgeting in mental accounting. Organ. Behav. Hum. Decis. Process. 62, 38–54.
Heath, C., Soll, J., 1996. Mental budgeting and consumer decisions. J. Consum. Res. 23, 40–52.
Kable, J.W., Glimcher, P.W., 2007. The neural correlates of subjective value during intertemporal choice. Nat. Neurosci. 10, 1625–1633.
Kable, J.W., Glimcher, P.W., 2009. The neurobiology of decision: consensus and controversy. Neuron 63, 733–745.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–292.
Kahneman, D., Tversky, A., 1984. Choices, values, and frames. Am. Psychol. 39, 341–350.
Kerns, J.G., Cohen, J.D., MacDonald, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S., 2004. Anterior cingulate conflict monitoring and adjustments in control. Science. 303, 1023–1026.
Krajbich, I., Armel, C., Rangel, A., 2010. Visual fixations and the computation and comparison of value in simple choice. Nat. Neurosci. 13, 1292–1298.
Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., Palminteri, S., 2017. Behavioural and neural characterization of optimistic reinforcement learning. Nat. Hum. Behav. 1, 1–9.
Levy, D.J., Glimcher, P.W., 2011. Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. J. Neurosci. 31, 14693–14707.
Li, J., Daw, N.D., 2011. Signals in human striatum are appropriate for policy update rather than value prediction. J. Neurosci. 31, 5504–5511.
Li, W., Wang, H., Xie, X., Li, J., 2019. Neural mediation of greed personality trait on economic risk-taking. Elife 8, 1–15.
Louie, K., Khaw, M.W., Glimcher, P.W., 2013. Normalization is a general neural mechanism for context-dependent decision making. Proc. Natl. Acad. Sci. 110, 6139–6144.
Martino, B.De, Kumaran, D., Seymour, B., Dolan, R.J., 2006. Frames, biases, and rational decision-making in the human brain. Science. 313, 684–688.
McCain, B.E., 1986. Continuing investment under conditions of failure: a laboratory study of the limits to escalation. J. Appl. Psychol. 71, 280–284.
Montague, P.R., Hyman, S.E., Cohen, J.Y., 2004. Computational roles for dopamine in behavioural control. Nature 431, 760–767.
Niv, Y., Edlund, J.A., Dayan, P., O'Doherty, J.P., 2012. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J. Neurosci. 32, 551–562.
Rangel, A., Camerer, C., Montague, P.R., 2008. A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. 9, 545–556.
Rigoli, F., Rutledge, R.B., Dayan, P., Dolan, R.J., 2016. The influence of contextual reward statistics on risk preference. Neuroimage 128, 74–84.
Roth, S., Robbert, T., Straus, L., 2015. On the sunk-cost effect in economic decision-making: a meta-analytic review. Bus. Res. 8, 99–138.
Ruff, C.C., Ugazio, G., Fehr, E., 2013. Changing social norm compliance with noninvasive brain stimulation. Science 342, 482–484.
Saint-Exupery, Ade, 1987. Le petit prince. Gallimard, Paris.
Sharot, T., Riccardi, A.M., Raio, C.M., Phelps, E.A., 2007. Neural mechanisms mediating optimism bias. Nature 450, 102–105.
Shenhav, A., Cohen, J.D., Botvinick, M.M., 2016. Dorsal anterior cingulate cortex and the value of control control. Nat. Neurosci. 19, 1286–1291.
Sleesman, D.J., Conlon, D.E., McNamara, G., Miles, J.E., 2012. Cleaning up the big muddy: A meta-analytic review of the determinants of escalation of commitment. Acad. Manag. J. 55, 541–562.
Sokol-Hessner, P., Hsu, M., Curley, N.G., Delgado, M.R., Camerer, C.F., Phelps, E.A., 2009. Thinking like a trader selectively reduces individuals' loss aversion. Proc. Natl. Acad. Sci. 106, 5035–5040.
Soltani, A., de Martino, B., Camerer, C., 2012. A range-normalization model of context-dependent choice: a new model and evidence. PLoS Comput. Biol. 8, 1–15.

Soman, D., Cheema, A., 2001. The effect of windfall gains on the sunk-cost effect. Mark. Lett. 12, 51–62.
Staw, B.M., 1976. Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. Organ. Behav. Hum. Perform. 16, 27–44.
Staw, B.M., 1981. The escalation to course of commitment of action. Acad. Manag. Rev. 6, 577–587.
Staw, B.M., Ross, J., 1989. Understanding behavior in escalation situations. Science 246, 216–220.
Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., Friston, K.J., 2009. Bayesian model selection for group studies. Neuroimage 46, 1004–1017.
Sweis, B.M., Abram, S.V., Schmidt, B.J., Seeland, K.D., MacDonald, A.M., Thomas, M.J., Redish, A.D., 2018. Sensitivity to "sunk costs" in mice, rats, and humans. Science 361, 178–181.
Thaler, R.H., 1980. Toward a positive theory of consumer choice. J. Econ. Behav. Organ. 1, 39–60.
Thaler, R.H., 1985. Mental accounting and consumer choice. Mark. Sci. 4, 199–214.
Thaler, R.H., 1999. Mental accounting matters. J. Behav. Decis. Mak. 12, 183–206.
Thaler, R.H., Johnson, E.J., 1990. Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice. Manage. Sci. 36, 643–660.
Tom, S.M., Fox, C.R., Trepel, C., Poldrack, R.A., 2007. The neural basis of loss aversion in decision-making under risk. Science 315, 515–518.
Tversky, A., 1972. Elimination by aspects: a theory of choice. Psychol. Rev. 79, 1–19.
Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211, 453–458.