Behavioral/Cognitive

Transcranial Stimulation over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used to Resolve Exploration–Exploitation Trade-Offs

Anjali Raja Beharelle, Rafael Polanía, Todd A. Hare,* and Christian C. Ruff*
Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, 8006 Zurich, Switzerland

Optimal behavior requires striking a balance between exploiting tried-and-true options or exploring new possibilities. Neuroimaging studies have identified different brain regions in humans where neural activity is correlated with exploratory or exploitative behavior, but it is unclear whether this activity directly implements these choices or simply reflects a byproduct of the behavior. Moreover, it remains unknown whether arbitrating between exploration and exploitation can be influenced with exogenous methods, such as brain stimulation. In our study, we addressed these questions by selectively upregulating and downregulating neuronal excitability with anodal or cathodal transcranial direct current stimulation over right frontopolar cortex during a reward-learning task. This caused participants to make slower, more exploratory or faster, more exploitative decisions, respectively. Bayesian computational modeling revealed that stimulation affected how much participants took both expected and obtained rewards into account when choosing to exploit or explore: Cathodal stimulation resulted in an increased focus on the option expected to yield the highest payout, whereas anodal stimulation led to choices that were less influenced by anticipated payoff magnitudes and were more driven by recent negative reward prediction errors. These findings suggest that exploration is triggered by a neural mechanism that is sensitive to prior less-than-expected choice outcomes and thus pushes people to seek out alternative courses of action. Together, our findings establish a parsimonious neurobiological mechanism that causes exploration and exploitation, and they provide new insights into the choice features used by this mechanism to direct decision-making.

Key words: brain stimulation; decision-making; frontal lobes; prediction error; reward; tDCS

Significance Statement

We often must choose whether to try something new (exploration) or stick with a proven strategy (exploitation). Balancing this trade-off is important for survival and growth across species because, without exploration, we would perseverate with the same strategies and never discover better options. Which brain mechanisms are responsible for our ability to make these decisions? We show that applying different types of noninvasive brain stimulation over frontopolar cortex causes participants to explore more or less in uncertain environments. These changes in exploration reflect how much participants focus on expected payoffs and on memory of recent disappointments. Thus, our results characterize a neural mechanism that systematically incorporates anticipated rewards and past experiences to trigger exploration of alternative courses of action.

Introduction

To make optimal choices, we often have to shift away from actions with known outcomes (e.g., relying on a current fuel source) and test untried options that may potentially maximize long-term success (e.g., devoting resources to pursue alternative sources of energy). Managing this quandary between exploiting well-known choice options and exploring alternative opportunities is a central aspect of behavior for most species (e.g., Krebs et al., 1978; Keasar, 2002; Ame et al., 2004; Ben Jacob et al., 2004; Pratt and Sumpter, 2006; Hayden et al., 2011b) because exploration is critical for organisms to harvest resources optimally for growth and reproduction (Watkinson et al., 2005). When con-

Received June 17, 2015; revised Sept. 3, 2015; accepted Sept. 15, 2015.

Author contributions: A.R.B., T.A.H., and C.C.R. designed research; A.R.B. and R.P. performed research; A.R.B., R.P., T.A.H., and C.C.R. analyzed data; A.R.B., R.P., T.A.H., and C.C.R. wrote the paper.

This work was supported by Swiss National Science Foundation Grant 143443 to T.A.H. and C.C.R. and Marie Curie International Incoming Fellowship PIIF-GA-2012-327196 to A.R.B. We thank Adrian Etter for programming support, Michele Lorenzi for assistance with data collection and analysis, and Daphna Shohamy for insightful discussions.

The authors declare no competing financial interests.

*T.A.H. and C.C.R. contributed equally to this work as co-senior authors.

Correspondence should be addressed to Dr. Anjali Raja Beharelle, Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland. E-mail: anjali.rajabeharelle@econ.uzh.ch.

DOI:10.1523/JNEUROSCI.2322-15.2015

 0270-6474/15/3514544-13$15.00/0

fronted with choices requiring a trade-off between exploration and exploitation, organisms ranging from plants to primates behave according to the approximate best solution (Krebs et al., 1978; Stephens and Krebs, 1986; McNickle and Cahill, 2009; Bendesky et al., 2011; Hayden et al., 2011b). However, in humans, the neural mechanisms underlying this behavior are not fully understood, particularly because the decision environments where they are encountered are often dynamic, uncertain, and involve no explicit cues for behavior, even when objectives are clearly formulated (Cohen et al., 2007).

Neuroimaging studies have demonstrated a network of regions that are activated during exploration but have not yet determined whether this brain activity actually drives exploratory choices or only reflects correlated cognitive processes. Specifically, these studies have shown that frontopolar cortex (FPC) and the mid-intraparietal sulcus are more active when human participants engage in exploratory modes of behavior (Daw et al., 2006; Boorman et al., 2009, 2011; Laureiro-Martínez et al., 2013). Furthermore, the FPC has been shown to track the reward probability of unselected choice options (Boorman et al., 2009, 2011; Donoso et al., 2014), and activity in an extended section of prefrontal cortex that includes the FPC is associated with uncertainty in individuals who rely on this metric for exploration (Badre et al., 2012). These studies suggest that FPC activity may reflect a neural mechanism that triggers exploration, but whether or not neural computations in FPC indeed directly control exploration or exploitation is an open question.

In the present study, we addressed this question by applying anodal and cathodal transcranial direct current stimulation (tDCS) (Nitsche and Paulus, 2000) over right FPC (rFPC) to enhance or decrease neural excitability, respectively, while participants played a three-armed bandit task with continuously drifting payoff magnitudes. Our central hypothesis was that enhancing rFPC excitability would increase deliberative exploration (to collect information on bandit attributes), whereas decreasing rFPC excitability would lead to a focus on exploiting the bandit expected to give the maximum reward.

We also used choice variables derived from the computational modeling analyses to test two more mechanistic hypotheses clarifying the role of rFPC in controlling exploration. First, we examined the hypothesis that the tDCS-induced changes in exploration are mediated by altered sensitivity to the payoff magnitude of the bandits. This hypothesis was derived from findings that rFPC activity is increased during exploratory choices but is decreased during exploitative choices that focus on the immediate payoffs (e.g., Daw et al., 2006). Second, we investigated the more novel hypothesis that tDCS-mediated increases or decreases in exploration are related to higher or lower sensitivity to previous unexpected outcomes in payoff magnitudes (i.e., prediction errors), respectively. This hypothesis was motivated by proposals that the rFPC is involved in integrating memories of recent events to guide behavior (Tsujimoto et al., 2011). Our results were consistent with all three hypotheses: Anodal and cathodal rFPC-targeted tDCS indeed caused increased and decreased exploration, respectively. The increased exploitation during cathodal stimulation was strongest following higher payoff magnitudes, whereas the increased exploration under anodal stimulation was driven by increased sensitivity to previous negative prediction errors from unexpectedly low payoffs. These findings establish the rFPC as a neural *sine qua non* for integrating information about past, present, and future payoffs to arbitrate between exploration and exploitation in human decision making.

Materials and Methods

Participants

Seventy-nine students at the University of Zurich (23 female, mean ± SEM age 23 ± 0.34 years, range = 19–38 years) participated in our study. Only healthy participants who were not taking any medication for neurological or psychiatric illnesses or other psychotropic drugs were included. Smokers were not excluded from the study. Participants were randomly assigned to one of three groups that differed with respect to the type of tDCS they received: anodal ($N = 26$, 8 females), cathodal ($N = 27$, 7 females), or sham ($N = 26$, 8 females). Participants were well matched with respect to socioeconomic and personality variables across the three groups (Table 1). All participants gave informed consent before the study, and all experimental procedures were approved by the Zurich Cantonal Ethics Committee.

The experiment was conducted in a custom-designed, multiparticipant tDCS testing room at the Laboratory for Social and Neural Systems research. Participants were tested in groups of 6–12 people that were evenly distributed across the three stimulation conditions by random assignment in each testing session. Stimulation was initiated simultaneously for participants across all conditions in each testing session. Assignment to one of the three tDCS groups was performed in a double-blind fashion, where the persons who conducted the experiment did not know which seats received active or sham stimulation, and the participants did not know which type of stimulation they received. The group testing of participants thus controlled for unspecific effects, such as order and time of day effects that could potentially confound serial testing regimens.

Experimental paradigm and measures

Before the behavioral experiment, participants completed a questionnaire on basic demographic information and several personality scales measuring traits that may be confounded with exploration, such as impulsivity (BIS-11), sensation seeking (SSS-V), and anxiety (STAI). In addition, three questions were asked to get a gross assessment of IQ (Frederick, 2005) as well as the subjects' optimism about their answers to these questions, which has been shown to relate to exploratory behavior (Herz et al., 2014). In addition, three questions were asked to obtain a gross indication of risk aversion (Dohmen et al., 2011), impulsivity, and temporal discounting. Finally, subjects played incentivized lotteries measuring risk and ambiguity aversion preferences (Ellsberg, 1961) and answered questions related to their current mood [Multidimensional Mood State Questionnaire (MDMQ), which is the English version of the Mehrdimensionale Befindlichkeitsfragebogen] (Steyer et al., 1997) before and after tDCS. Importantly, none of these measures was differentially affected by tDCS (for details, see Results).

Participants then played a computerized virtual slot machine game in which they had to choose repeatedly among three bandits with real financial consequences, during a baseline control period and again during stimulation over rFPC, to quantify stimulation effects relative to baseline exploratory decision-making. The bandit task is well suited for assessing exploration and exploitation because the payout values of the slot machines drift randomly and independently over trials. The variation over time prompts participants to explore occasionally the payout values of other slot machines for comparison with the slot machine that they currently believe to be the highest-paying option (Fig. 1*B*).

The bandit task we used was similar to tasks used in previous fMRI studies (Daw et al., 2006). The payout values for each bandit were generated with a decaying Gaussian random walk. Specifically, the reward for choosing the ith slot machine on trial t was between 1 and 100 points (rounded to the nearest integer), drawing from a Gaussian distribution (SD, $\sigma_o = 4$) centered on a mean $\mu_{i,t}$. On each trial t, the means diffused in a decaying Gaussian random walk, with the following:

$$\mu_{i,t+1} = \lambda \mu_{i,t} + (1 - \lambda)\theta + \nu \quad (1)$$

for each slot machine i. The decay parameter λ was set at 0.9836, the decay center θ was 50, and the diffusion noise ν was zero-mean Gaussian (SD, $\sigma_d = 2.8$). Eighteen such reward payout sequences were generated to have at least four instances where the bandit with the highest reward switched and at least four instances where the bandit with the highest

Table 1. Demographic, socioeconomic, and personality variables[a]

Demographic, socioeconomic, or personality variable	Anodal ($N = 26$)	Cathodal ($N = 27$)	Sham ($N = 26$)	$F_{(2,78)}$	p
Age (years)	23.35 ± 4.15	22.88 ± 2.16	22.96 ± 2.49	0.170	0.844
Gender (2 = female)	1.31 ± 0.47	1.26 ± 0.45	1.31 ± 0.47	0.097	0.907
Nationality	1.46 ± 0.85	1.85 ± 1.03	1.69 ± 0.88	1.183	0.312
First language	2.38 ± 2.12	2.52 ± 2.21	1.69 ± 1.69	1.264	0.288
Marital status	1.38 ± 0.64	1.30 ± 0.47	1.23 ± 0.43	0.577	0.564
Religious affiliation	3.69 ± 2.46	4.00 ± 2.15	4.15 ± 2.39	0.263	0.769
Direction of studies	4.15 ± 1.35	4.15 ± 1.23	4.04 ± 1.46	0.061	0.941
Community size (scale 0–5)	3.27 ± 1.48	3.30 ± 1.20	3.12 ± 1.53	0.125	0.882
Relative affluence (scale 0–6)	3.73 ± 1.37	4.04 ± 0.81	3.46 ± 1.10	1.767	0.178
Risk aversion (scale 0–10)	5.81 ± 2.61	5.56 ± 2.31	6.04 ± 2.47	0.255	0.776
Impulsivity (scale 0–10)	4.54 ± 2.10	4.96 ± 2.39	4.00 ± 2.14	1.254	0.291
Temporal discounting (scale 0–10)	7.50 ± 2.04	6.78 ± 2.08	7.31 ± 1.38	1.069	0.348
IQ no. correct (of 3)	2.58 ± 0.76	2.52 ± 0.89	2.42 ± 0.76	0.241	0.786
Optimism about IQ (scale 0–3)	3.00 ± 0.69	2.81 ± 0.74	2.88 ± 0.71	0.453	0.637
BIS-11	60.54 ± 7.92	65.00 ± 9.24	62.27 ± 8.56	1.819	0.169
SSS-V	21.96 ± 26.40	18.11 ± 23.51	15.54 ± 25.45	0.430	0.652
STAI	60.00 ± 6.82	62.85 ± 6.23*	62.04 ± 5.68	0.151	0.860
MDMQ mood ^{B}	23.58 ± 2.61	23.48 ± 3.86	24.12 ± 2.85	0.307	0.737
MDMQ mood $^{\Delta}$	−3.07 ± 4.61	−4.63 ± 4.11	−2.54 ± 3.59	1.842	0.165
MDMQ alertness ^{B}	17.15 ± 2.26	18.00 ± 3.06	17.50 ± 3.46	0.545	0.582
MDMQ alertness $^{\Delta}$	2.92 ± 5.23	1.51 ± 6.05	2.08 ± 6.47	0.374	0.689
MDMQ calmness ^{B}	17.27 ± 3.75	16.04 ± 4.53	18.08 ± 3.07	1.902	0.156
MDMQ calmness $^{\Delta}$	−1.77 ± 4.16	−0.26 ± 4.30	−3.08 ± 4.02	3.036	0.054
Risk aversion lottery ^{B}	2.96 ± 1.04	2.85 ± 1.06	3.19 ± 0.85	0.813	0.447
Risk aversion lottery $^{\Delta}$	0.12 ± 0.52	0.33 ± 0.48	0.04 ± 0.60	2.189	0.119
Ambiguity aversion lottery ^{B}	3.00 ± 1.10	3.48 ± 1.09	3.54 ± 1.24	1.751	0.181
Ambiguity aversion lottery $^{\Delta}$	0.46 ± 0.76	0.04 ± 0.81	0.42 ± 1.21	1.647	0.199
Math Task performance ^{B}	0.93 ± 0.08	0.88 ± 0.21	0.93 ± 0.15	0.712	0.494
Math Task performance $^{\Delta}$	0.03 ± 0.08	0.03 ± 0.11	0.03 ± 0.15	0.002	0.998

[a]The first three columns of this table display the descriptive statistics (means ± standard deviations) for the various demographic, socioeconomic, and personality variables collected on individuals in each treatment group. The last two columns contain *F*-statistics and corresponding *p* values for the main effect of tDCS group derived from one-way ANOVAs with a single three-level factor: tDCS group assignment (anodal, cathodal, and sham). The details of the demographic, socioeconomic, and personality measures are given in the section "Material and Methods." The tDCS groups did not differ on any of these variables, suggesting that any differences in exploration were unlikely to be caused by interactions of the stimulation with underlying personality, cognitive, demographic, or socioeconomic characteristics. The superscript letter B refers to a measure collected during the pre-stimulation baseline, while a superscript symbol Δ indicates a difference score calculated as tDCS minus baseline. Measures without any superscript symbol were collected only once prior to the start of the experiment. *One subject was missing the STAI score.

reward was at least 30 points greater than the bandit with the lowest reward. The payout sequences between bandits were nearly orthogonal (pairwise correlations <0.1). A subset of four sequences was selected for the experiment based on separation among the individual payout sequences; each participant saw one randomly drawn sequence during baseline and another one during stimulation.

The task consisted of 284 "decision" trials and 15 "belief" trials. On each trial, participants were presented with three boxes representing the virtual slot machines (Fig. 1*A*). In the center of the screen, a diamond or square symbol indicated whether the trial was a decision or a belief trial (3–7.5 s intertrial interval). Decision trials started with presentation of a bonus [randomly drawn from a β distribution ($\alpha = 2$, $\beta = 5$) ranging from 0 to 20 points] in each of the slot machines. Modifying the original task (Daw et al., 2006) by adding these bonuses did not affect task performance, except for the fact that the bonuses were also factored into choices (see Bonus values alone on do not differentially affect choice behavior as a function of stimulation type). Participants then selected a slot machine, using three arrow keys on the keyboard, and the slot machine's summed reward (bonus plus payout value) was displayed on the screen (2.5 s). To compute the payout value of a bandit, participants simply had to subtract the bonus assigned to that bandit from the total payoff displayed. This design ensured that participants could not plan their next response immediately after receiving the current feedback and therefore had to take their choice at the start of each trial, when the bonuses appeared onscreen. On belief trials, subjects were instructed to rate the slot machines in terms of estimated payouts from 1 (highest payout value) to 3 (lowest payout value).

Participants completed a short practice session before proceeding with the actual task. After the participants completed the bandit task during both baseline and stimulation runs, they performed a Math Task (consisting of 50 trials) designed to assess whether stimulation over rFPC affected their abilities to perform the subtraction necessary to update a bandit's payout value. During this task, a single box appeared that was visually identical to the boxes representing the virtual slot machines. The box contained a randomly drawn bonus ranging from 0 to 20 points, which was followed by the sum of the bonus and a random integer ranging from 1 to 100 representing a summed reward. Participants were instructed to input the underlying payout value of the slot machine (i.e., the difference of the displayed total amount and the bonus).

tDCS

During the experiment, we applied tDCS over the participants' rFPC using a commercially available multichannel stimulator (neuroConn; http://www.neuroconn.de/dc-stimulatormc/) that allows for simultaneous stimulation of up to 16 participants with individually customized stimulation protocols. tDCS can alter cortical excitability via application of direct currents, with anodal tDCS increasing and cathodal tDCS decreasing excitability of the area under the target electrode (Nitsche and Paulus, 2000). In the present study, we applied anodal, cathodal, or sham tDCS over the right FPC region (MNI peak: $x = 27$, $y = 57$, $z = 6$) that had shown significantly enhanced BOLD signal for exploratory relative to exploitative choices at the group level (Daw et al., 2006). We chose the right over left FPC because the peak of the BOLD signal was reported on the right side in previous studies (Daw et al., 2006; Boorman et al., 2009; Laureiro-Martínez et al., 2013). This standardized coordinate was transformed to each individual's native headspace by aligning it to the T1-weighted MR scan of the participant's neuroanatomy (T1-weighted 3D turbo field echo, 181 sagittal slices, matrix size 256×256, voxel size = $1 \times 1 \times 1$ mm). The point on the scalp overlying this brain area was marked and used as the center point for the active electrode. The reference electrode was placed over the vertex, defined for each participant as

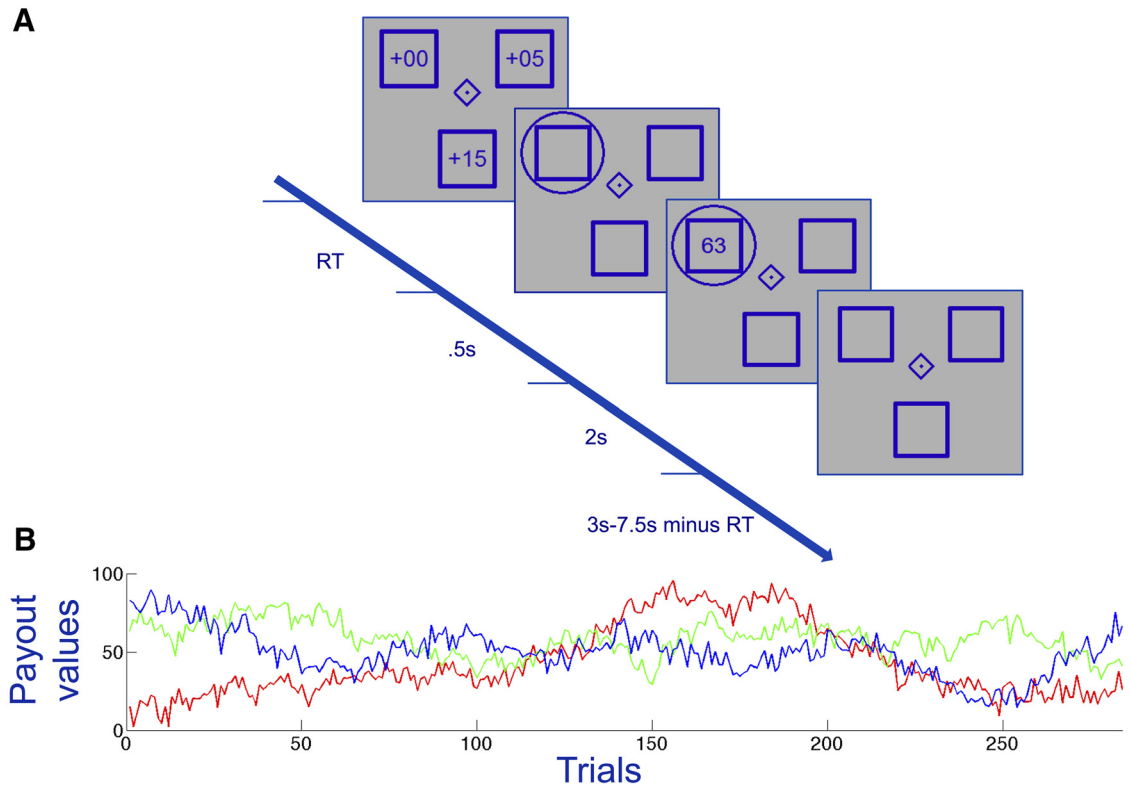

Figure 1. The bandit task paradigm. ***A***, Participants selected among three virtual slot machines (shown as blue squares) whose payout values drifted independently and randomly across trials. The time-varying monetary rewards required participants to learn continuously about the slot machines to maximize their monetary payoffs. At the start of each trial, participants saw three bonuses (numbers displayed within the squares in first screenshot) that had to be added to the slot machine's underlying payout value to determine the total reward. After participants made their choice (circled option), the total reward was displayed. ***B***, One example payoff sequence for all three slot machines over the course of 284 trials. Each colored line indicates a different slot machine.

the point corresponding to the confluence of the left and right central sulcus. For each participant, both the stimulation and reference points were located using Brainsight 2.2 frameless stereotaxy (Rogue Research; https://www.rogue-research.com/).

tDCS was applied using a 5×5 cm^2 electrode over the site of stimulation and a 10×10 cm^2 electrode over the reference site. Both electrodes were fixed by rubber straps and gauze. We chose a more focal electrode to maximize the current density over the area of interest and a large electrode for the reference to minimize current density at the vertex area. Using such a tDCS montage makes the large reference electrode functionally ineffective (Nitsche et al., 2007); we could therefore be certain that the effects of tDCS on exploratory behavior would not be affected by neuromodulatory influences on neural activity under the reference electrode. Therefore, the only difference between the two active stimulation groups was whether the electrode placed over the rFPC was the anode or the cathode, respectively.

We stimulated with a current intensity of 1 mA for both the anodal and cathodal groups. It has been shown in basic neurophysiological studies in humans (Nitsche and Paulus, 2000) and mouse slice preparations (Fritsch et al., 2010) that the impact of tDCS on brain excitability becomes more robust and long-lasting after several minutes, possibly reflecting delayed short-term neuroplastic processes occurring in addition to the immediate changes in membrane electric potential. To account for this possible delay in the onset of stable tDCS effects, we waited for 3 min of stimulation to pass before subjects started the bandit task. At the start of stimulation, the current was slowly ramped up for 20 s to minimize the itching or tingling sensations caused by abrupt onsets of tDCS. Normally, the participants adjust to the sensation of the stimulation after a short period of time, and tDCS is usually not noticeable throughout the rest of the duration of stimulation (Gandiga et al., 2006). At the end of anodal and cathodal stimulation, the current was slowly ramped down for 20 s. In the sham group, the tDCS was ramped down after 30 s of stimulation. This allowed the participants to feel an initial sensation identical to that experienced by the participants in the anodal and cathodal groups, but there was no subsequent effect on neural excitability. Indeed, when participants were asked at the end of the experiment to indicate ("yes" [1] or "no" [0]) whether they felt the stimulation throughout the whole experiment, the sham group's responses to this question were not significantly different from the anodal and cathodal group's responses ($F_{(2,78)} = 0.838$, $p = 0.432$; mean ± SD: Anodal = 0.23 ± 0.43; Cathodal = 0.37 ± 0.49; Sham = 0.23 ± 0.43). Likewise, there were no significant differences among groups when participants were asked to rate (on a scale of 0–6) whether the tDCS affected their behavior during the slot machine game ($F_{(2,78)} = 0.087$, $p = 0.917$; mean ± SD: Anodal = 1.58 ± 1.77; Cathodal = 1.67 ± 1.73; Sham = 1.46 ± 1.88). However, the sham group did rate (on a scale of 0–6) the tDCS to be less annoying compared with the anodal group as shown by Tukey *post hoc* tests ($F_{(2,78)} = 6.072$, $p = 0.004$; Sham vs Anodal $t_{(50)} = -3.4267$, $p = 0.001$, Sham vs Cathodal $t_{(51)} = -1.9583$, $p = 0.056$, Anodal vs Cathodal $t_{(51)} = 1.6014$, $p = 0.115$; mean ± SD: Anodal = 3.27 ± 2.16; Cathodal = 2.37 ± 1.92; Sham = 1.35 ± 1.87). There were no significant differences in the responses to this question between anodal and cathodal groups or cathodal and sham groups. Furthermore, a more rigorous assessment of participants' emotional state using a standardized questionnaire (MDMQ) indicated that the tDCS manipulations did not result in changes in subscales indexing mood, alertness, and calmness from the baseline period to after stimulation. There were also no initial baseline differences among groups in mood, alertness, and calmness (Table 1). Together, these control analyses show that tDCS did not have any unspecific non-neural effects on beliefs about stimulation or emotional state. Thus, changes in exploratory behavior due to stimulation over rFPC are unlikely to have been caused by these unrelated factors.

Bayesian reinforcement learning model

To assess participants' exploratory behavior, we used the Bayesian mean-tracking rule described by Daw et al. (2006) that modeled participants' estimates of the mean and variance of the reward values for each bandit across trials. We assumed the participant believed the process of generating bandit rewards is governed by parameters $\hat{\sigma}_o$, $\hat{\sigma}_d$, $\hat{\lambda}$, and $\hat{\theta}$, corresponding to σ_o, σ_d, λ, and θ from the process described above. Assume on trial t a prior distribution over the true mean reward values, $\mu_{i,t}$, as independent Gaussian distributions Normal $(\hat{\mu}_{i,t}^{pre}, \hat{\sigma}_{i,t}^{2\,pre})$. If option c_t is chosen and reward value r_t is received, then the posterior mean for that option is as follows:

$$\hat{\mu}_{c_t,t}^{post} = \hat{\mu}_{c_t,t}^{pre} + \kappa_t \delta_t, \tag{2}$$

with the prediction error $\delta_t = r_t - \hat{\sigma}_{c_t,t}^{pre}$ and learning rate $\kappa_t = \hat{\sigma}_{c_t,t}^{2\,pre}/(\hat{\sigma}_{c_t,t}^{2\,pre} + \hat{\sigma}_o^2)$. The posterior variance for the chosen option is then:

$$\hat{\sigma}_{c_t,t}^{post} = (1 - \kappa_t)\, \hat{\sigma}_{c_t,t}^{pre}. \tag{3}$$

The posterior means and variances for the unchosen bandits are not changed by the result of the choice; however, the prior distributions for all bandits, i, are updated on the subsequent trial with the following diffusion process:

$$\hat{\mu}_{i,t+1}^{pre} = \hat{\lambda}\hat{\mu}_{i,t}^{post} + (1 - \hat{\lambda})\hat{\theta} \tag{4}$$

and

$$\hat{\sigma}_{i,t+1}^{2\,pre} = \hat{\lambda}^2\, \hat{\sigma}_{i,t}^{2\,post} + \hat{\sigma}_d^2. \tag{5}$$

Together with this tracking rule, we used a softmax choice rule that included an uncertainty bonus to determine the probability $P_{i,t}$ of choosing slot machine i on trial t as a function of the estimated means and variances of the rewards as follows:

$$P_{i,t} = \frac{\exp[\beta(\hat{\mu}_{i,t}^{pre} + bonus + \varphi\hat{\sigma}_{i,t}^{pre})]}{\sum_j \exp[\beta\,(\hat{\mu}_{j,t}^{pre} + bonus + \varphi\hat{\sigma}_{j,t}^{pre})]}, \tag{6}$$

where β is the gain parameter that determines how tightly the decisions are constrained by the relative mean rewards among the bandits. $\hat{\mu}_{i,t}^{pre} + bonus$ refers to the participant's estimates of the mean reward (payout value plus bonus) of bandit i at trial t. $\hat{\sigma}_{i,t}^{pre}$ refers to the participant's estimates of variance around the estimated mean reward of bandit i at trial t. If $\varphi > 0$, a participant's choices are also influenced by their estimate of the relative variance (or uncertainty) among the bandit reward values.

Model fits

The computational model was fit to the participants' choices using a hierarchical Bayesian framework. We adopted this approach to increase the sensitivity of our modeling framework, as simulations have shown that hierarchical Bayesian modeling techniques outperform standard individual- or group-level maximum likelihood estimation in recovering true parameters (Ahn et al., 2011). However, our model remains a minor adaption of the model originally proposed by Daw et al. (2006). At the participant level, we treated the interindividual differences as random effects for β and φ.

$$\beta \sim \text{Beta}(\breve{\mu}_\beta \times \breve{k}_\beta, (1 - \breve{\mu}_\beta) \times \breve{k}_\beta)$$

$$\varphi \sim \text{Beta}(\breve{\mu}_\varphi \times \breve{k}_\varphi, (1 - \breve{\mu}_\varphi) \times \breve{k}_\varphi) \times 10^3$$

where $\breve{\mu}$ and $\breve{k}$ correspond to the expected value and the precision of the Beta distribution, respectively. The Beta distribution used for φ is scaled by 10^3 to allow this parameter to have a wide range exceeding 1 (the upper boundary of a Beta distribution), thus allowing it to act as a bonus (i.e., multiplier) of the participant's estimate of variance $(\hat{\sigma}_{i,t}^{pre})$ within the choice rule. For both $\breve{\mu}$ and $\breve{k}$, we assumed uninformed flat priors such that $\breve{\mu}$ was initialized with a Beta distribution with $\alpha = 1$ and $\beta = 1$ and $\breve{k}$ was initialized with a Gamma distribution with shape = 1 and rate = 0.1. At the population level, a single instance was used to fit $\hat{\lambda}$, $\hat{\theta}$, and $\hat{\sigma}_d$, assuming uninformed flat priors via a Beta distribution with $\alpha = 1$ and $\beta = 1$. $\hat{\sigma}_o$ was held constant at 4.00 due to model degeneracy. To maintain consistency with the modeling approach taken by Daw et al. (2006), we fit $\hat{\lambda}$, $\hat{\theta}$, and $\hat{\sigma}_d$ at the population level while allowing the softmax choice rule parameters to vary individually. We also allowed the choice rule parameters to be fit individually as these parameters were a focus of our modeling hypotheses. At the trial level, the categorical choices y at each trial t were assumed to be drawn from a multinomial distribution as follows:

$$y_{\text{choices}} \sim \chi(P_{1:3,t}, 1)$$

Posterior inference of the parameters in the hierarchical Bayesian models was performed via the Gibbs sampler using the Markov chain Monte-carlo sampling scheme implemented in JAGS (Plummer, 2003) (http://mcmc-jags.sourceforge.net/). Three chains were derived based on a different random number generator engine, each starting with a different seed. A total of 5000 samples were drawn during an initial burn-in step to allow the model parameters to reach a stable range. Subsequently, a total of 5000 new samples were drawn for each of the chains. We applied a thinning (i.e., downsampling) of five steps to this final sample, thus resulting in a final set of 1000 samples for each parameter and chain. This thinning assured that there were no auto-correlations in the final samples for all of the parameters of interest investigated in this study. We conducted Gelman-Rubin tests (Gelman et al., 2013) for each parameter to confirm convergence of the chains. All estimated parameters in our Bayesian models had <1.05, which suggests that all three chains converged to a target posterior distribution (Gelman et al., 2013).

When comparing the fit for both the above choice rule with the uncertainty bonus and the standard softmax choice rule used by Daw et al. (2006), we found that the former fit the data better [average Deviance Information Criterion (DIC)] (Gelman et al., 2013) across groups = 2.19×10^4 compared with an average DIC = 2.22×10^4 for the standard softmax choice rule; the smaller the DIC, the better the fit. We therefore applied this rule to generate the model-based estimates of reward values. However, the model-based estimates of the reward values for both choice rules were strongly correlated in any case ($r = 0.7972$, $p < 0.00001$). The fitted parameters for each participant group and condition are presented in Table 2.

Table 2. Bayesian learning model parameters[a]

	Baseline			tDCS − Baseline		
Model parameter	Anodal	Cathodal	Sham	Anodal	Cathodal	Sham
β	0.0846 ± 0.0318	0.0707 ± 0.0358	0.0808 ± 0.0403	−0.0031 ± 0.0332	0.0115 ± 0.0267	0.0107 ± 0.0480
φ	0.0304 ± 0.0349	0.0133 ± 0.0162	0.0700 ± 0.0770	−0.0285 ± 0.0313	−0.0018 ± 0.0114	−0.0638 ± 0.0699
$\hat{\lambda}$	0.9489	0.9537	0.9406	−0.0073	−0.0083	−0.0127
$\hat{\theta}$	31.189	19.279	12.326	−2.3406	1.9519	11.465
$\hat{\sigma}_d$	0.0010	0.0028	0.0010	0.0025	−0.0001	0.0053

[a]The first three columns show the parameter estimates across all participants and trials within each stimulation group (anodal, cathodal, or sham) in the baseline condition. Columns 4 to 6 show the change in the parameter estimates under tDCS relative to baseline for each stimulation condition. Parameters β and φ were allowed to vary individually across participants and were fit in a hierarchical manner (see Materials and Methods). For these parameters, the table reports the means (± standard deviations) of the median values of each subject's posterior distribution. The remaining parameters were fit across the population; for these parameters, the table reports the medians of the population posterior distribution.

Next, we sampled the posterior distributions obtained from the Bayesian model 100 times for each parameter and generated subject-specific estimates of the means and variances of the reward values for each bandit. The final estimates for each subject were taken as the average of these 100 samples. To test how well the model estimates of reward values approximated the actual reward values, we conducted a linear mixed effects regression using the lme4 package (Barr et al., 2013) in R (R Core Team, 2014, Version 3.0.2) as follows:

$$Y_{st} = \beta_0 + \beta_1 \hat{\mu}_{st} + (S_{0s} + S_{1s}\hat{\mu}_t) + e_{st}, \quad (7)$$

where Y_{st} refers to the real reward value for a given subject (s) and trial (t) on each slot machine, and $\hat{\mu}_{st}$ refers to the estimated reward value of the same slot machine. The coefficients β_0 and β_1 are fixed effects fit to the entire sample, whereas $S_{0s} + S_{1s}$ are subject-specific random effects.

Analyses of bandit task choices

We conducted several analyses to test our three hypotheses (H1, H2, and H3) about how rFPC-targeted stimulation would affect choice behavior on the three-armed bandit task. To clearly convey the logic of our analysis strategy, we first introduce our hypotheses and their underlying rationale in a numbered list before describing in correspondingly labeled subsections the analyses conducted to test each of them.

H1: The right FPC is causally involved in biasing choices toward exploration or exploitation. Based on prior neuroimaging work implicating the rFPC in exploratory behavior (Daw et al., 2006; Boorman et al., 2009, 2011; Laureiro-Martínez et al., 2013), we predicted that enhanced FPC neural excitability will increase deliberative exploratory choices whereas decreasing excitability will result in more exploitative decisions. We tested this hypothesis in two steps. First, we compared the frequency and degree of exploration following anodal, cathodal, and sham stimulation. Second, we compared reaction times between the stimulation groups to determine whether any increase in exploratory behavior is (1) the result of a neural mechanism promoting deliberate exploration, which should slow choices because the increased consideration among several alternative options increases choice difficulty (Bogacz et al., 2010; Krajbich et al., 2010, 2015; Shenhav et al., 2014) or (2) instead reflects unspecific effects of tDCS on response inhibition or neural signal-to-noise ratio, which would result in faster choices as predicted by modeling (Bestmann et al., 2014; Bonaiuto and Arbib, 2014) and neurophysiological studies (Terzuolo and Bullock, 1956; Bindman et al., 1962, 1963; Creutzfeldt et al., 1962; Nitsche and Paulus, 2000, 2001; Fritsch and Hitzig, 2009) (for further theoretical rationale behind this hypothesis, see RFPC-targeted stimulation affects both exploration and exploitation).

H2: The tDCS-elicited biases toward exploration or exploitation reflect changes in behavioral sensitivity to the estimated reward magnitudes of the highest-paying and/or alternative options. FPC activity is thought to override tendencies to choose based on reward signals generated in striatal and ventromedial prefrontal areas of the brain (Daw et al., 2006). We thus predicted that cathodal stimulation inhibiting activity in rFPC would render choices more sensitive to the immediate payoff magnitude of the highest-paying option, whereas anodal stimulation would cause participants' choices to be less influenced by the highest-paying option and perhaps more sensitive to the payoffs of the second- or third-highest-paying bandits (for further theoretical motivation of this hypothesis, see tDCS-induced exploitation relates to increased sensitivity to predicted payoffs).

H3: The bias toward exploration elicited by anodal tDCS reflects a change in the sensitivity to unexpected choice outcomes. Because the FPC is thought to integrate memory of recent experiences to directly inform choice (e.g., Tsujimoto et al., 2011), we hypothesized that anodal rFPC stimulation will increase participants' likelihood to explore either following negative prediction errors (i.e., unexpectedly low payoffs) for exploitative choices in the recent past and/or after positive prediction errors (i.e., unexpectedly high payoffs) for recent exploratory choices. By contrast, cathodal stimulation will result in reduced sensitivity to these prediction errors (for further information motivating this hypothesis, see tDCS-induced exploration relates to increased sensitivity to negative prediction errors).

Tests of H1: analyses quantifying exploration and exploitation

Using the model-generated estimates of mean reward values, we classified choices as exploitative if the participant chose the option with the highest estimated payoff (mean plus bonus) and as exploratory if the participant chose the options with the estimated second- or third-highest payoff. Once we had classified choices as exploitative or exploratory, we conducted three analyses to test the (1) frequency, (2) degree, and (3) deliberative nature of exploratory choices. The frequency of exploratory choices was simply the fraction of trials on which participants chose to explore. We defined the degree (or strength) of exploration as the amount of monetary reward the participant was willing to give up by not selecting the highest-paying option and instead explore. This was quantified as the difference between what a participant estimated as the highest reward value (mean + bonus) that they could receive on any given trial and the actual estimated reward value that they chose. We tested the directional hypotheses that the frequency and strength of exploration in the task during stimulation over rFPC would follow the pattern: anodal ≥ sham ≥ cathodal, using a one-sided Jonckheere-Terpstra trend (JT) test. Finally, we analyzed reaction times to assess whether tDCS-elicited changes in exploration reflected faster, more random or slower, more deliberative responses. We therefore tested the directional hypothesis that relative reaction times would be affected in the following way: anodal ≥ sham ≥ cathodal stimulation with a one-sided JT test.

Tests of H2: analyses of how monetary reward magnitudes influence choice

To examine how stimulation over rFPC affected the participants' sensitivity to the monetary reward magnitudes of the slot machines, we first standardized the estimated payoffs (underlying mean + bonus) using a z-transformation for each participant and bandit (ranked from highest to lowest value in a trial-wise manner). We then conducted a logistic regression individually for each participant (using the glmfit function in MATLAB, Release R2014a, version 8.3.0.532; The MathWorks, 2014) to assess how much each bandit's estimated reward value (and interactions of these estimated values) predicted the participant's choice to explore or exploit as follows:

$$y = \beta_0 + \beta_1\hat{\mu}_1 + \beta_2\hat{\mu}_2 + \beta_3\hat{\mu}_3 + \beta_4(\hat{\mu}_1 * \hat{\mu}_2) + \beta_5(\hat{\mu}_1 * \hat{\mu}_3) + \beta_6(\hat{\mu}_2 * \hat{\mu}_3) + \beta_7(\hat{\mu}_1 * \hat{\mu}_2 * \hat{\mu}_3) + e \quad (8)$$

where y is the choice to explore (1) or exploit (0) and $\hat{\mu}_i$ is the estimated monetary reward value for bandit Rank $i = 1, 2,$ or 3. Bandit rank refers to a trial-wise ranking of the bandits estimated to yield the highest (1) to lowest (3) payoff. These subject-wise regressions yielded a set of coefficients quantifying the relationship between estimated payout values and exploratory choices in each participant. The β_1, β_2, and β_3 parameters of this regression were then submitted to repeated-measures ANOVAs to examine the effects of Condition (baseline, tDCS), stimulation Polarity (anodal, cathodal, and sham), and payoff-based bandit Rank (1–3) on exploration. The interaction terms were added to quantify any potential interaction effects among the estimated bandit reward values when estimating the coefficients for each individual bandit. However, none of these interaction terms was significantly different from zero. Nevertheless, for the sake of completeness, we tested whether any of these interaction terms was significantly affected by tDCS, but found no significant effects [i.e., there was no interaction between the factors Condition and stimulation Polarity (ANOVA stimulation Polarity × Condition, $F_{(2,76)}$ for $\beta_{4-7} = 0.228, 0.930, 0.883, 0.984$, $p = 0.413, 0.293, 0.702, 0.651$)]. Moreover, the results for our analyses are similar if we omit these interaction terms from the individual regressions for each participant.

Tests of H3: analyses of how current choices are guided by previous prediction errors

To examine how tDCS alters the degree to which recent outcomes influence the choice to explore or exploit, we calculated the prediction error (i.e., the difference between the observed and expected outcome) separately for (1) the highest-paying bandit and (2) either the second- or

third-highest-paying bandits on the preceding trials. To approximate a short-term memory effect for this signal (e.g., Cowan, 2011), the prediction error signals were modeled as persisting over two subsequent choices and returned to zero if a bandit was not sampled for two consecutive trials.

We then conducted a logistic regression (using the glmfit function in MATLAB) to assess how these two prediction error signals related to exploratory or exploitative decisions when the relative mean estimates of the bandit reward values were controlled for the following:

$$y = \beta_0 + \beta_1 PE_{explore} + \beta_2 PE_{exploit} + \beta_3(\hat{\mu}_1 - \hat{\mu}_2) + \beta_4(\hat{\mu}_1 - \hat{\mu}_3) + e \quad (9)$$

where y is the choice to explore (1) or exploit (0), PE is the prediction error on the previous trials of the highest-paying option (exploit) or the second or third best options (explore), and $\hat{\mu}_i$ is the estimated reward value for bandit i. Next, we examined the effects of Condition (baseline, tDCS), stimulation Polarity (anodal, cathodal, and sham) on the β_1 β_2 parameters of these logistic regressions with a repeated-measures ANOVA.

The reported results are robust with respect to the temporal influence of the prediction errors across trials. We chose a time-limited influence to approximate a short-term memory-effect (e.g., Cowan, 2011) and because model comparisons favored our model in which prediction errors influenced future choices only over a short time window (lag = 2 trials) over a model with a prediction error that was continuously held in mind until the bandit was sampled again (quantified by lower Akaike information criteria; Akaike, 1974) (paired-samples t test: $t_{(78)} = -2.917$, $p = 0.005$). Moreover, the effects of Condition (baseline, tDCS), stimulation Polarity (anodal, cathodal, and sham), and Choice type (exploratory vs exploitative) remained consistent when examining the β parameters of logistic regressions with prediction errors that persisted for limited lags greater or less than 2, indicating that, although short lags provide a superior fit compared with continuously persisting PEs, the precise duration of the lag is not crucial (repeated-measures ANOVA for prediction errors that persisted for one subsequent trial: three-way interaction, $F_{(2,76)} = 3.766$, $p = 0.028$; repeated-measures ANOVA for prediction errors that persisted for three consecutive trials: three-way interaction, $F_{(2,76)} = 3.600$, $p = 0.032$).

Additional supporting analyses

The effect of bonus values on choice behavior. Because we modified the bandit task used by Daw et al. (2006) to incorporate trial-wise bonuses that prevented participants from making their choices before the onset of a trial, we also tested whether stimulation altered the impact of the bonuses themselves on choice for completeness. This was done with a logistic regression similar to the one used to examine estimated payoff magnitudes (i.e., underlying mean + bonus), but which now used only the three bonuses and not the estimated underlying means. Specifically, we conducted a logistic regression individually for each participant (using the glmfit function in MATLAB, as above) to assess how much each bandit's bonus value (and interactions of these estimated values) predicted the participant's choice to explore or exploit as follows:

$$y = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 B_3 + \beta_4(B_1 * B_2) + \beta_5(B_1 * B_3) + \beta_6(B_2 * B_3) + \beta_7(B_1 * B_2 * B_3) + e \quad (10)$$

where y is the choice to explore (1) or exploit (0) and B_i is the bonus value on each trial for bandit with Rank $i = 1$, 2, or 3. Bandit rank refers to a trial-wise ranking of the bandits expected to yield the highest (1) to lowest (3) payoff.

The parameters of this regression were then submitted to repeated-measures ANOVAs to examine the effects of Condition (baseline, tDCS), stimulation Polarity (anodal, cathodal, and sham), and reward-based bandit Rank (1–3; ranked by bonus + estimated mean) on exploration.

Average earnings. To examine the effects of stimulation on task performance, we tested the effects of Condition (baseline, tDCS) and stimulation Polarity (anodal, cathodal, and sham) on the average amount of monetary reward earned across all trials, using repeated-measures ANOVA.

Results

Participants could track the bandit payoffs

Before examining any tDCS effects, we ascertained that participants understood the underlying structure and time-varying nature of the payoffs in the bandit task during the prestimulation baseline. To this end, we compared the payoff estimates derived from our Bayesian reinforcement learning model with the true underlying parameters of the task. During the baseline period, model-generated estimates of the participants' beliefs about the payoff magnitudes were strongly correlated with the actual magnitudes seen during the experiment (linear mixed effects regression across all slot machines and participants during the baseline period: intercept coefficient = 7.253 ± 0.767, $t = 9.46$, $p < 0.001$; slope coefficient = 0.966 ± 0.01, $t = 98.10$, $p < 0.001$), confirming that the participants' beliefs closely tracked the actual payoffs.

Effects of rFPC stimulation are not confounded by preexisting group differences or modulation of general cognitive ability

Before testing our core hypotheses about the effects of rFPC stimulation on exploration, we also established that the three stimulation groups did not show personality or cognitive differences at baseline. ANOVAs (Table 1) showed that the three groups did not differ in terms of basic sociodemographic and personality variables. There were also no baseline differences in decisions over incentivized lotteries measuring aversion to risk and ambiguity or in the ability to perform the subtractions necessary to calculate the bandit payout values (Math Task percentage correct). We also ascertained that the three rFPC stimulation protocols applied after the baseline session did not differentially affect basic aspects of cognition that might contribute to exploratory and exploitative choices in our task. Importantly, tDCS did not significantly change preferences for risk, ambiguity, or Math Task performance relative to baseline behavior (Table 1). Thus, any tDCS effects on exploratory or exploitative choices are not confounded by changes in these basic cognitive functions that may potentially relate to exploration.

RFPC-targeted stimulation affects both exploration and exploitation

We then tested our three main hypotheses, namely, that rFPC-targeted stimulation would alter deliberate exploratory behavior (H1) and that these effects would be mediated by changes in the influence of current payoff estimates (H2) and feedback from preceding outcomes (H3). We derived and tested a series of predictions from these three interrelated hypotheses, as already described in the Introduction and Materials and Methods. These predictions are presented in greater detail alongside the corresponding results in the paragraphs below.

Previous neuroimaging work has shown that the rFPC is more active during exploratory decisions (Daw et al., 2006; Boorman et al., 2009, 2011; Laureiro-Martínez et al., 2013). Therefore, to examine Hypothesis 1, we first tested the prediction that enhancing rFPC neural excitability through anodal stimulation will increase exploration, whereas decreasing excitability with cathodal stimulation will result in more exploitative decisions. In line with our hypothesis, anodal stimulation increased the number of exploratory choices, whereas cathodal stimulation decreased the number of exploratory choices, relative to the sham group (JT test; $z = -2.59$, $p = 0.005$; Fig. 2*A*). Both types of tDCS also had

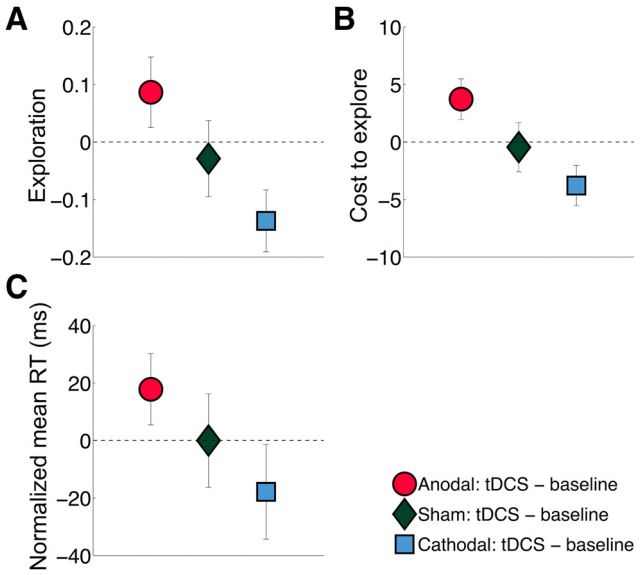

Figure 2. Brain stimulation effects on behavior (Tests of H1). Anodal and cathodal stimulation over rFPC had opposing effects on exploration and reaction times. ***A***, Anodal stimulation led to an increase in exploration, whereas cathodal stimulation decreased exploration, relative to inactive sham stimulation that left exploration unaffected. Exploration was measured as the fraction of exploratory choices made, and all displayed data points indicate changes in this index under tDCS relative to the preceding baseline. ***B***, Anodal stimulation made participants willing to incur a significantly higher monetary cost to explore, whereas cathodal stimulation rendered participants less willing to pay for exploration, compared with sham stimulation. Cost to explore was calculated as the best amount possible for a given trial minus the monetary units (i.e., the payout magnitude consisting of reward + bonus) received on the trial. All displayed data points indicate changes in this index under tDCS relative to the preceding baseline. ***C***, All groups showed faster reaction times from baseline to stimulation due to training effects. However, anodal stimulation over rFPC led to relatively slower choices than sham stimulation. Conversely, cathodal stimulation resulted in relatively faster choices. Reaction times (RT) are plotted as relative to the mean of the sham group (which was defined as zero). Error bars indicate SEM.

the predicted opposite effects on the degree of exploration, with anodal increasing and cathodal decreasing the monetary cost participants were willing to incur to explore (JT test; $z = -3.02$, $p = 0.001$; Fig. 2*B*). Thus, stimulation over rFPC causally regulated both the frequency and degree to which participants engaged in exploratory versus exploitative decision behavior.

In addition, we found broadly consistent results for the softmax β parameter, the inverse of which has been suggested as a proxy for exploration (e.g., Cohen et al., 2007). This parameter was indeed decreased more in the anodal group and increased more in the cathodal group relative to baseline (Table 2). However, the magnitude of the inverse β parameter is also proportional to the noise or randomness across choices, which may potentially be altered by unspecific tDCS effects on response inhibition and neural signal-to-noise ratios (Terzuolo and Bullock, 1956; Bindman et al., 1962, 1963; Creutzfeldt et al., 1962; Nitsche and Paulus, 2000, 2001; Fritsch and Hitzig, 2009; Bestmann et al., 2014; Bonaiuto and Arbib, 2014). Such random responses may therefore be mistaken for enhanced exploration but would be inconsistent with our hypothesis that anodal stimulation over rFPC will lead to more deliberate decisions to explore.

To test whether the increase in exploratory choices under rFPC stimulation reflected enhancement of a neural process for deliberate consideration of choice properties or rather an unspecific random process, we directly compared reaction times for choices across stimulation groups. If the observed choice modulations were caused by unspecific tDCS effects (such as generally reduced response inhibition or altered signal-to-noise characteristics), then anodal stimulation should lead to faster reaction times relative to sham stimulation for all types of decisions, whereas cathodal stimulation should result in slower reaction times, as predicted by neurophysiological studies (Terzuolo and Bullock, 1956; Bindman et al., 1962, 1963; Creutzfeldt et al., 1962; Nitsche and Paulus, 2000, 2001; Nitsche et al., 2003; Fritsch and Hitzig, 2009) and model simulations (Bestmann et al., 2014; Bonaiuto and Arbib, 2014). However, if tDCS specifically affects neural computations in rFPC that result in deliberate choices to explore or exploit, the opposite pattern should emerge: Reaction times should be longer for anodal stimulation (due to increased consideration of exploratory options and thus greater choice difficulty) but shorter during cathodal stimulation (that results in a selective focus on the current best option). This prediction is based on repeated findings that choices take longer if they involve more alternatives and/or entail options that are more similar in value (Bogacz et al., 2010; Krajbich et al., 2010, 2015; Shenhav et al., 2014). Evidence that exploratory choices in our task indeed entailed increased deliberation comes from our participants' baseline behavior before tDCS: Exploratory choices indeed took longer than exploitative choices (ANOVA, main effect of trial type: $F_{(1,76)} = 6.397$, $p = 0.014$; Exploratory trial mean ± SD = 3032.81 ± 121.45 ms; Exploitative trial mean ± SD = 2937.86 ± 145.94 ms). Crucially, once stimulation was applied, participants in the anodal group responded more slowly while participants in the cathodal group responded more quickly, relative to the sham stimulation group response times (JT test; $z = -1.71$, $p = 0.043$; Fig. 2*C*). This pattern of tDCS-induced response time alterations was only evident for the bandit task and was not observed for the control task involving mental calculation ability (JT test; $z = 0.468$, $p = 0.320$). Thus, these directional and selective tDCS effects on reaction times argue against an explanation of our results in terms of altered neural noise levels or response inhibition (Terzuolo and Bullock, 1956; Bindman et al., 1962, 1963; Creutzfeldt et al., 1962; Nitsche and Paulus, 2000, 2001; Fritsch and Hitzig, 2009; Bestmann et al., 2014; Bonaiuto and Arbib, 2014). Instead, our results show that rFPC-targeted stimulation caused either slower, more deliberative exploratory decisions or faster, more targeted exploitative choices, depending on whether the stimulation was set up to increase or decrease neural excitability, respectively. Further tests of how tDCS affected the nature of the choices in the bandit task are included in the analyses addressing Hypotheses 2 and 3 below.

tDCS-induced exploitation relates to increased sensitivity to predicted payoffs

After establishing that anodal and cathodal rFPC-targeted tDCS increased and decreased the frequency of exploratory choices, respectively, we sought to determine which attributes of the bandit task were associated with each decision type as a function of stimulation polarity. By definition, value-based choices are thought to be driven by the level of reward that is predicted to result from each possible action in the choice set (i.e., the expected value) (Rangel et al., 2008). Frontopolar activity has been postulated to promote exploration by overriding exploitative response tendencies that are based on expected value signals computed in striatal and ventromedial prefrontal areas (Daw et al., 2006). Therefore, we hypothesized that the modulation of exploratory behavior through rFPC-targeted stimulation may be mediated by changes in the degree to which choices are driven by the immediate payout values of each slot machine. To test this

Table 3. The influence of estimated payoffs on choice as a function of stimulation Polarity, Condition, and reward-based bandit Rank[a]

	Baseline			tDCS		
	Bandit rank 1	Bandit rank 2	Bandit rank 3	Bandit rank 1	Bandit rank 2	Bandit rank 3
Anodal	−0.44 ± 0.55	0.21 ± 0.40	0.04 ± 0.25	0.03 ± 0.91	−0.02 ± 0.56	0.02 ± 0.37
Cathodal	−0.11 ± 0.69	0.23 ± 0.40	−0.00 ± 0.27	−0.49 ± 0.58	0.26 ± 0.40	0.02 ± 0.30
Sham	−0.31 ± 0.68	0.24 ± 0.36	−0.03 ± 0.35	−0.45 ± 0.77	0.21 ± 0.39	0.01 ± 0.38

[a]The values in this table represent the mean ± SD coefficients for each bandit rank (1–3) from Regression 8. In the baseline condition, participants were more likely to exploit the higher the estimated payoffs of the highest ranked bandit and more likely to explore the higher the estimated payoffs of the second and third highest-ranked bandits. In the stimulation condition, anodal participants' choices were no longer driven by the estimated rewards of the highest ranked bandit, whereas cathodal participants' choices were more strongly influenced by the estimated payoffs of the highest ranked bandit. This table corresponds to information presented in Figures 3 and 4 and tests of Hypothesis 2.

hypothesis, we specifically examined the relationship between participants' current estimations of the bandit payout values and their exploratory choices (see Tests of H2: analyses of how monetary reward magnitudes influence choice). Compared with baseline choice patterns, anodal and cathodal stimulation differentially affected the way the estimated slot machine payoffs influenced exploratory choices [repeated-measures ANOVA, three-way interaction of Condition (baseline, stimulation), stimulation Polarity (anodal, cathodal, sham), and reward-based Rank of the slot machines (1–3), $F_{(2.5,96.1)} = 1.907, p = 0.034$; Fig. 4*B*; Table 3]. In the baseline condition, all participants were more likely to exploit as the estimated monetary rewards of the highest-paying slot machine increased but became more likely to explore as the estimated magnitudes of the second- or third-highest-paying (exploratory) slot machine increased (*post hoc* one-sample *t* tests: highest-paying slot machine, $t_{(78)} = -3.874$, $p = 0.0002$; highest-paying slot machine, $t_{(78)} = 5.287$, $p = 0.000001$; third-highest-paying slot machine, $t_{(78)} = 0.035$, $p = 0.972$; Figure 3). The strength of this relationship was significantly altered during tDCS compared with the preceding baseline period (interaction of stimulation Polarity and reward-based Rank of slot machines, $F_{(2.5,95.9)} = 3.134$, $p = 0.037$; Fig. 4*A*). The estimated rewards of the highest-paying slot machine became less influential in driving the more exploratory anodal group's choices ($t_{(26)} = -2.151, p = 0.041$) but had a stronger effect on the choices of the more exploitative cathodal group ($t_{(26)} = 2.107, p = 0.045$). Importantly, these changes were not due to general time or learning effects, as sham tDCS did not significantly change the impact of estimated monetary rewards on choice compared with baseline ($t_{(25)} = -0.711, p = 0.484$). Moreover, these tDCS effects were specific to the slot machine with the highest estimated monetary reward, as the influence of the second- and third-highest payouts on choice did not change with tDCS (second-highest baseline vs stimulation: anodal paired-sample, $t_{(25)} = -1.479, p = 0.152$; cathodal paired-sample $t_{(26)} = 0.320, p = 0.751$, sham paired-sample, $t_{(25)} = -0.199, p = 0.844$; third-highest baseline vs stimulation: anodal paired-sample, $t_{(25)} = -0.177, p = 0.861$; cathodal paired-sample, $t_{(26)} = 0.285, p = 0.778$; sham paired-sample, $t_{(25)} = 0.351, p = 0.723$).

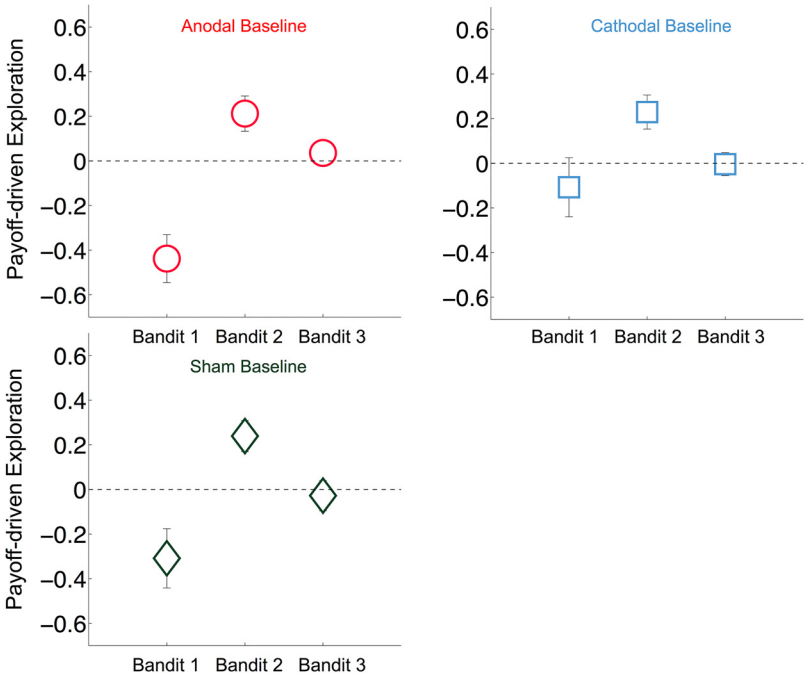

Figure 3. Payoff-driven exploration during baseline. During the baseline before stimulation, the rewards of the slot machines influenced participants' decisions to explore or exploit similarly across all three groups. The *y*-axis represents standardized β values from a logistic regression of decisions (explore vs exploit on estimated slot machine rewards). Positive β values indicate that the participant was more likely to explore with higher estimated rewards of the slot machine. Negative β values indicate that the participant was more likely to exploit with higher estimated rewards of the slot machine. There was a significant main effect of the reward-based rank of the slot machines on the probability to explore or exploit. Participants in all three groups were more likely to exploit if the value of the highest-paying slot machine was greater and more likely to explore if the value of the second highest-paying slot machine was greater. Bandit 1, Bandit 2, and Bandit 3 refer to a trial-wise ranking of the bandits estimated to yield the highest (1) to lowest (3) payoff. Error bars indicate SEM.

Bonus values alone do not differentially affect choice behavior as a function of stimulation type

Because we modified the bandit task used by Daw et al. (2006) to incorporate trial-wise bonuses that prevented participants from making their choices before the onset of a trial, we tested whether stimulation altered the impact of the bonuses themselves, rather than the predicted payoffs, on choice. Comparisons of the relevant regression weights (see Additional supporting analyses) showed a nonsignificant interaction between Condition, stimulation Polarity, and bandit Rank [repeated-measures ANOVA, three-way interaction of Condition (baseline, stimulation), stimulation Polarity (anodal, cathodal, sham), and reward-based Rank of the slot machines (1–3), $F_{(2.9,109.71)} = 2.091, p = 0.108$; Polarity × Rank interaction: $F_{(3.135,119.121)} = 1.863, p = 0.137$]. Thus, the tDCS effects on exploitative choices were specifically related to the more choice-relevant total payoff values (underlying mean + bonus) rather than to the bonuses in isolation.

Together, our results concerning Hypothesis 2 suggest that the more frequent exploitation during cathodal stimulation reflected

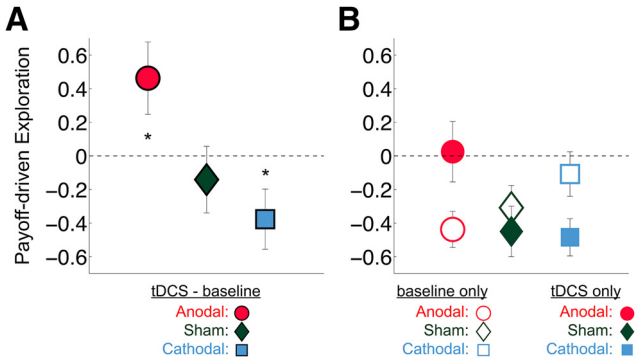

Figure 4. Brain stimulation effects on the influence of estimated rewards on choice (Test of H2). Right, FPC stimulation changed how the estimated rewards of the slot machines influenced participants' decisions to explore or exploit. The *y*-axis represents standardized β values from a logistic regression of decisions (explore vs exploit) on estimated slot machine rewards. Positive (negative) β values indicate that the participant was more (less) likely to explore as the estimated monetary rewards of a slot machine increased. Significant effects were found only for the highest-paying slot machine; therefore, only those coefficients are shown here. ***A***, Generated from the difference between tDCS and baseline values shown in ***B***. ***A***, tDCS led to a significant change in the influence of estimated monetary rewards from the highest-paying slot machine on participants' choices relative to baseline. ***B***, Participants who received anodal stimulation over rFPC became significantly less influenced by the estimated rewards relative to baseline when deciding to explore or exploit, whereas participants who received cathodal stimulation over rFPC became significantly more influenced by the estimated rewards relative to baseline. Error bars indicate SEM. *$p < 0.05$.

Table 4. Earnings as a function of stimulation polarity and condition[a]

	Baseline earnings	Stimulation earnings
Anodal	65.25 ± 6.28	61.70 ± 7.64
Cathodal	61.84 ± 8.20	65.01 ± 6.08
Sham	62.61 ± 6.87	64.03 ± 7.39

[a]The earnings from task payoffs are reported in experimental monetary units (MUs). The conversion rate from experimental MUs to CHF was 1 MU = 0.42 CHF. The anodal group earned significantly less than average during stimulation compared with baseline. The cathodal group earned more than they did during baseline, whereas there was no change in earnings in the sham group (repeated-measures ANOVA: interaction between Condition and stimulation Polarity, $F_{(2,76)} = 3.864$, $p = 0.025$; *post hoc* paired-sample *t* test for stimulation vs baseline: Anodal $t_{(25)} = 2.110$, $p = 0.045$, Cathodal $t_{(26)} = -1.856$, $p = 0.075$, Sham $t_{(25)} = -0.736$, $p = 0.469$). Data are mean ± SD.

an increased focus on monetary reward magnitudes expected for the highest-paying option, whereas the increased exploration during anodal stimulation related to a lower sensitivity to monetary reward magnitudes. Interestingly, the stimulation-induced exploration came at a financial cost: The anodal group earned significantly less monetary units (MUs) on average during stimulation compared with baseline, the cathodal group earned more than they did during baseline, whereas there was no change in earnings in the sham group (repeated-measures ANOVA: interaction between Condition and stimulation Polarity, $F_{(2,76)} = 3.864$, $p = 0.025$; *post hoc* paired-sample *t* tests for stimulation vs baseline: anodal, $t_{(25)} = 2.110$, $p = 0.045$; cathodal, $t_{(26)} = -1.856$, $p = 0.075$; sham, $t_{(25)} = -0.736$, $p = 0.469$; Table 4). The conversion was 1 MU = 0.42 CHF. This difference in monetary earnings concurs with the results above concerning Hypothesis 1, by showing again that the anodal group was willing to sacrifice more money to explore.

tDCS-induced exploration relates to increased sensitivity to negative prediction errors

Our third hypothesis was that the bias toward exploration elicited by anodal tDCS reflects a change in the sensitivity to unexpected choice outcomes. This hypothesis was derived from previous proposals that the FPC integrates and extrapolates recent experiences in short-term memory (Ramnani and Owen, 2004; Kovach et al., 2012). In the context of our task, the saliency of an outcome may be determined by the difference between the actually received and the expected payoff (i.e., the prediction error) (Schultz, 1998; Tobler et al., 2006; Hayden et al., 2011a; Hauser et al., 2014). If the rFPC promotes exploration based on salient recent outcomes, then this tendency should be particularly strong following large negative prediction errors (unexpectedly low payoffs) for the highest-paying option and/or for large positive prediction errors (unexpectedly high payoffs) for the two lower-paying options. Therefore, we examined whether anodal stimulation caused participants to explore more either because they were more attracted by positive prediction errors on the alternative options (Boorman et al., 2009) or because they shifted away from the exploited option after negative prediction errors. We also examined the corresponding prediction that cathodal stimulation would make participants' choices more robust to recent deviations from expected outcomes while maintaining the already established focus on exploitation of maximum estimated payoffs. To conduct these tests, we estimated an additional set of GLMs at the individual level and then compared the effects of previous prediction errors on current choices across stimulation conditions (see Tests of H3: analyses of how current choices are guided by previous prediction errors).

We found that rFPC-tDCS indeed differentially affected how previous prediction errors influenced current choice behavior, depending on the types of stimulation and slot machine chosen [repeated-measures ANOVA: three-way interaction Condition, stimulation Polarity, and Choice type (exploratory vs exploitative), $F_{(2,76)} = 3.789$, $p = 0.027$; Fig. 5*B*; Table 5]. Compared with the preceding baseline period, stimulation polarity had opposite effects on how current choices were influenced by prediction errors associated with past exploitative choices (ANOVA, $F_{(2,76)} = 4.204$, $p = 0.019$; Fig. 5*A*). In the exploratory anodal group, participants became more likely to explore when the recent outcome of the highest-paying slot machine was less than expected and more likely to exploit if the recent outcome of the

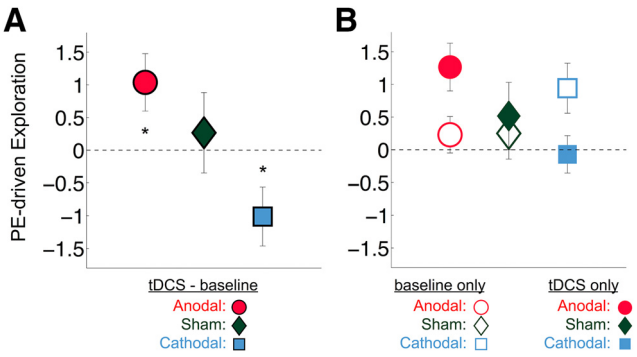

Figure 5. Brain stimulation effects on the influence of prediction errors on choice (Test of H3). Right FPC stimulation changed how prediction errors (i.e., the difference between true and expected outcome; PE, prediction errors) of the highest-paying slot machine influenced subsequent decisions to explore or exploit. Because of recoding of prediction errors (multiplication with −1), positive β values indicate that the participant was more likely to explore following negative prediction errors. ***A***, Generated from the difference between tDCS and baseline values shown in ***B***. ***A***, tDCS changed how participants' choices were driven by the recent outcomes of the highest-paying slot machine relative to baseline. ***B***, Participants who received anodal stimulation over rFPC became significantly more likely to explore when the outcome of the highest-paying slot machine was lower than expected and significantly more likely to exploit when the outcome was higher than expected. Participants who received cathodal stimulation over rFPC were less influenced by the recent outcomes of the highest-paying slot machines. Error bars indicate SEM. *$p < 0.05$.

Table 5. The influence of previous prediction errors on choice as a function of stimulation Polarity and Condition[a]

	Baseline		tDCS	
	PE explore	PE exploit	PE explore	PE exploit
Anodal	0.68 ± 0.42	0.60 ± 1.91	0.68 ± 0.64	1.97 ± 2.56
Cathodal	0.67 ± 0.55	1.60 ± 2.53	1.03 ± 0.46	0.49 ± 1.78
Sham	0.84 ± 0.54	0.89 ± 2.23	0.95 ± 0.70	1.19 ± 3.13

[a]The values in this table represent the mean (± SD) coefficients of interest in Regression 9. Because of recoding of prediction errors (multiplication with −1), positive beta values indicate that the participant was more likely to explore following negative prediction errors. During stimulation, the anodal group's choice became significantly more influenced by the past negative prediction errors (PE) on exploitative choices (β_2 in Regression 9) such that the more stronger the recent negative prediction error, the more likely they were to explore. In contrast, the influence of PEs for the exploratory options (β_1 in Regression 9) on current choice did not vary significantly based on stimulation Polarity or Condition. This table corresponds to information presented in Figure 5 and tests of Hypothesis 3.

highest slot machine was more than expected (*post hoc* paired-sample *t* test: $t_{(25)} = 2.362$, $p = 0.026$). By contrast, the exploitative cathodal group's choices became less strongly influenced by the recent outcomes of the highest-paying option under stimulation (paired-sample $t_{(26)} = -2.243$, $p = 0.034$). Consistent with the analysis of payoff values, these prediction error effects were only observed for recent outcomes of the potentially exploited highest-paying option, but not for the recent outcomes of the exploratory choices on the second- and third-highest-paying bandits. Again, these changes were not related to time or learning effects, as sham tDCS did not lead to a significant change in the effect of recent outcomes on choice (paired-sample $t_{(25)} = 0.432$, $p = 0.670$). Moreover, the stronger influence of previous negative prediction errors on choices under anodal stimulation goes beyond the updating of bandit payoff estimates because those effects are already accounted for within the reinforcement-learning model and accounted for in the individual regressions by including the relative payoff estimates as additional regressors. Together, these results suggest that the increased exploration in the anodal stimulation group reflected an increased responsiveness to previous lower-than-expected outcomes of exploitative choices, whereas the increased exploitation in the cathodal group related to a weaker influence of recent prediction errors and a stronger focus on the current monetary reward of the highest-paying option.

Discussion

Our results establish a causal role for the rFPC in regulating both exploration and exploitation, and they underscore that this region is critical for participants to look beyond the current benefits at hand to search for potentially greater rewards (Wilson et al., 2014). Together, findings from the tests of our three hypotheses support that the activation observed in FPC when participants switch to exploratory choices (e.g., Daw et al., 2006; Boorman et al., 2009) indeed relates to behavioral control in those situations. Previous research has characterized distributed neuroanatomical systems that underlie exploration and exploitation, with frontopolar cortex and intraparietal sulcus preferentially active during exploratory choices and striatal and ventromedial prefrontal regions more activated during value-driven exploitative choices (Daw et al., 2006). Our findings demonstrate that a parsimonious causal neurobiological mechanism underlies both exploratory and exploitative behavior, as deciding both to explore or exploit could be manipulated along a continuum with tDCS over the same brain region. While these influences of stimulation may be mediated by interactions of the rFPC with a network of interconnected cortical and subcortical brain areas (Boorman et al., 2009, 2011), they nevertheless support the view that a unified neural architecture flexibly resolves the range of the exploration and exploitation tradeoffs (Gittins and Jones, 1974; Kaelbling, 1993). This view is consistent with the ubiquity of this capacity across organisms of varying complexity.

Our findings go beyond simply demonstrating that tDCS over FPC changes exploratory behavior, by showing that tDCS-increased exploration or exploitation is associated with altered sensitivity to particular attributes of the current decision problem. As expected, we found that reducing rFPC excitability by means of cathodal tDCS resulted in an increased focus on the payoffs available from the bandit currently estimated to yield the most money. Furthermore, enhancing rFPC neuronal excitability via anodal tDCS increased participants' willingness to pay to explore alternative, currently less profitable options. These findings significantly extend the correlative results of previous neuroimaging studies, by suggesting that the rFPC causally resolves the exploration/exploitation dilemma by balancing behavioral sensitivity to expected payoffs.

In addition to the findings that could be predicted based on previous neuroimaging work, we identified a novel computational factor driving exploration within FPC circuitry. The increased exploration due to enhancement of frontopolar excitability was related to a stronger focus on recent negative prediction errors from the highest-paying slot machine, rather than to outcomes from, or beliefs about, the forgone slot machines, as might have been predicted based on previous work (Boorman et al., 2009). That is, the tDCS-enhanced exploration reflected a stronger influence of recent disappointments (i.e., unexpectedly low payoffs) on the highest-paying option, thus suggesting that rFPC-induced exploration may in part be motivated by thinking that "the grass is becoming less green on this side." This is in line with findings that FPC lesions impair the ability to extrapolate trends (Kovach et al., 2012) and is broadly congruent with suggestions that FPC may integrate outcomes of multiple cognitive events (Ramnani and Owen, 2004), compare sequential outcomes within changing context (Pollmann et al., 2007), and use memory of recent events to generate responses (Wagner et al., 1998; Fletcher and Henson, 2001; Badre and Wagner, 2005; Tsujimoto et al., 2011). Previous neuroimaging findings have shown that the FPC more generally tracks multiple probabilistic events that are temporally linked to forecast future outcomes (Koechlin and Hyafil, 2007; Koechlin, 2008; Boorman et al., 2009). However, tracking and forecasting is not enough to explain why increasing excitability over rFPC triggers a signal to more readily abandon a "sinking" option. Instead, anodal tDCS may have lowered decision inertia and made participants less willing to accept outcomes that were lower than expected, in line with findings that the FPC may play a role in self-generated actions (Christoff et al., 2003; Tsujimoto et al., 2010) and that prefrontal-basal ganglia circuits are important for overcoming status quo bias (Fleming et al., 2010) and apathy (Alexander and Stuss, 2000; Levy, 2012). Thus, our results generate the interesting hypothesis for future neuroimaging work that the rFPC may work in concert with other interconnected regions to initiate changes in behavioral strategies after outcomes that are worse than expected (see also Nevo and Erev, 2012).

Importantly, our results cannot be explained by unspecific effects of tDCS on neural activity or by changes in decision noise. If tDCS had generally lowered response inhibition during the task, then we should have observed faster reaction times for anodal stimulation and slower reaction times for cathodal stimulation (Terzuolo and Bullock, 1956; Bindman et al., 1962, 1963; Creutzfeldt et al., 1962; Nitsche and Paulus, 2000, 2001; Fritsch

and Hitzig, 2009; Bestmann et al., 2014; Bonaiuto and Arbib, 2014). Instead, stimulation over rFPC appeared to affect specific neural computations related to choice, resulting in longer deliberation for the exploratory anodal group and shorter, more targeted decisions in the exploitative cathodal group. The notion of such a strategic exploration mechanism is congruent with previous demonstrations that exploratory choices are generally associated with longer reaction times than exploitative choices (Hassall et al., 2013) as well as with the finding that decisions to leave a patch in a foraging task have longer average response times (Shenhav et al., 2014). Critically, anodal stimulation over rFPC did not generally make participants process information more slowly, as we did not find stimulation-related changes in reaction times during numerical processing control tasks. Finally, our results cannot be explained simply by tDCS-induced changes in decision noise. This notion is contradicted by the findings that choices were systematically related to recent outcomes of the highest-paying option and that the beliefs about the slot machines' reward values remained similarly accurate after stimulation. Thus, our results suggest that anodal stimulation indeed modulated choice-related computations in the rFPC that result in slower, more deliberative decisions to shift toward exploratory options.

Our findings identify a locus of transcranial stimulation that may be used to help steer decision making in cases of pathologies involving deficits in task switching or stereotyped behavior (Ehring and Watkins, 2008; Stuss, 2011; Kleinman et al., 2013). The rFPC is thought to be embedded within a larger neural network regulating the exploration–exploitation tradeoff (Boorman et al., 2011) and is most likely not the only site where transcranial stimulation can have such effects. Stimulation of regions located within reward-valuation regions that are situated too deep within the cortex to be strongly affected by currently established tDCS techniques may have opposite effects on exploration and exploitation compared with those demonstrated in the present study (Daw et al., 2006; Boorman et al., 2013). Finally, the fact that excitatory anodal stimulation over rFPC causes decisions to be less dependent on immediate payoffs may serve to facilitate learning about the environment, by increasing the selection of the more uncertain second- or third-best options (Gittins and Jones, 1974; Badre et al., 2012; Payzan-Lenestour and Bossaerts, 2012). Foregoing immediate reward in favor of information about options that are potentially more rewarding in the long term may have beneficial implications for understanding and mitigating a wide range of behaviors associated with extreme forms of exploration (impulsivity, attention deficits, failure to integrate social information) or exploitation (compulsivity, lack of innovation, addiction) in individuals or social organizations.

References

Ahn WY, Krawitz A, Kim W, Busmeyer JR, Brown JW (2011) A model-based fMRI analysis with hierarchical Bayesian parameter estimation. J Neurosci Psychol Econ 4:95–110. CrossRef Medline

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723. CrossRef

Alexander MP, Stuss DT (2000) Disorders of frontal lobe functioning. Semin Neurol 20:427–437. CrossRef Medline

Ame JM, Rivault C, Deneubourg JL (2004) Cockroach aggregation based on strain odour recognition. Anim Behav 68:793–801. CrossRef

Badre D, Wagner AD (2005) Frontal lobe mechanisms that resolve proactive interference. Cereb Cortex 15:2003–2012. CrossRef Medline

Badre D, Doll BB, Long NM, Frank MJ (2012) Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. Neuron 73:595–607. CrossRef Medline

Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 68:3. CrossRef Medline

Ben Jacob E, Becker I, Shapira Y, Levine H (2004) Bacterial linguistic communication and social intelligence. Trends Microbiol 12:366–372. CrossRef Medline

Bendesky A, Tsunozaki M, Rockman MV, Kruglyak L, Bargmann CI (2011) Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. Nature 472:313–318. CrossRef Medline

Bestmann S, de Berker AO, Bonaiuto J (2014) Understanding the behavioural consequences of noninvasive brain stimulation. Trends Cogn Sci: 1–8. CrossRef Medline

Bindman LJ, Lippold OC, Redfearn JW (1962) Long-lasting changes in the level of the electrical activity of the cerebral cortex produced by polarizing currents. Nature 196:584–585. CrossRef Medline

Bindman LJ, Lippold OC, Redfearn JW (1963) Comparison of the effects on electrocortical activity of general body cooling and local cooling of the surface of the brain. Electroencephalogr Clin Neurophysiol 15:238–245. CrossRef Medline

Bogacz R, Wagenmakers EJ, Forstmann BU, Nieuwenhuis S (2010) The neural basis of the speed-accuracy tradeoff. Trends Neurosci 33:10–16. CrossRef Medline

Bonaiuto J, Arbib MA (2014) Modeling the BOLD correlates of competitive neural dynamics. Neural Netw 49:1–10. CrossRef Medline

Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62:733–743. CrossRef Medline

Boorman ED, Behrens TE, Rushworth MF (2011) Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. PLoS Biol 9:e1001093. CrossRef Medline

Boorman ED, Rushworth MF, Behrens TE (2013) Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. J Neurosci 33:2242–2253. CrossRef Medline

Christoff K, Ream JM, Geddes LP, Gabrieli JD (2003) Evaluating self-generated information: anterior prefrontal contributions to human cognition. Behav Neurosci 117:1161–1168. CrossRef Medline

Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos Trans R Soc Lond B Biol Sci 362:933–942. CrossRef Medline

Cowan N (2011) The focus of attention as observed in visual working memory tasks: making sense of competing claims. Neuropsychologia 49:1401–1406. CrossRef Medline

Creutzfeldt OD, Fromm GH, Kapp H (1962) Influence of transcortical d-c currents on cortical neuronal activity. Exp Neurol 5:436–452. CrossRef Medline

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879. CrossRef Medline

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG (2011) Individual risk attitudes: measurement, determinants, and behavioral consequences. J Eur Econ Assoc 9:522–550. CrossRef

Donoso M, Collins AG, Koechlin E (2014) Human cognition: foundations of human reasoning in the prefrontal cortex. Science 344:1481–1486. CrossRef Medline

Ehring T, Watkins ER (2008) Repetitive negative thinking as a transdiagnostic process. Int J Cogn Ther 2:192–205.

Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. Q J Econ 75: 643–669. CrossRef

Fleming SM, Thomas CL, Dolan RJ (2010) Overcoming status quo bias in the human brain. Proc Natl Acad Sci U S A 107:6005–6009. CrossRef Medline

Fletcher PC, Henson RN (2001) Frontal lobes and human memory: insights from functional neuroimaging. Brain 124:849–881. CrossRef Medline

Frederick S (2005) Cognitive reflection and decision making. J Econ Perspect 19:25–42. CrossRef Medline

Fritsch B, Reis J, Martinowich K, Schambra HM, Ji Y, Cohen LG, Lu B (2010) Direct current stimulation promotes BDNF-dependent synaptic plasticity: potential implications for motor learning. Neuron 66:198–204. CrossRef Medline

Fritsch G, Hitzig E (2009) Electric excitability of the cerebrum (Uber die

elektrische Erregbarkeit des Grosshirns). Epilepsy Behav 15:123–130. CrossRef Medline

Gandiga PC, Hummel FC, Cohen LG (2006) Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. Clin Neurophysiol 117:845–850. CrossRef Medline

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis, Ed 3. Boca Raton, FL: CRC.

Gittins JC, Jones DM (1974) A dynamic allocation index for the sequential design of experiments. In: Progress in statistics, pp 241–266. Amsterdam: North-Holland.

Hassall CD, Holland K, Krigolson OE (2013) What do I do now? An electroencephalographic investigation of the explore/exploit dilemma. Neuroscience 228:361–370. CrossRef Medline

Hauser TU, Iannaccone R, Stämpfli P, Drechsler R, Brandeis D, Walitza S, Brem S (2014) The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. Neuroimage 84:159–168. CrossRef Medline

Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011a) Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci 31:4178–4187. CrossRef Medline

Hayden BY, Pearson JM, Platt ML (2011b) Neuronal basis of sequential foraging decisions in a patchy environment. Nat Neurosci 14:933–939. CrossRef Medline

Herz H, Schunk D, Zehnder C (2014) How do judgmental overconfidence and overoptimism shape innovative activity? Games Econ Behav 83:1–23. CrossRef

Kaelbling LP (1993) Learning in embedded systems. Cambridge, MA: Massachusetts Institute of Technology.

Keasar T (2002) Bees in two-armed bandit situations: foraging choices and possible decision mechanisms. Behav Ecol 13:757–765. CrossRef

Kleinman JT, DuBois JC, Newhart M, Hillis AE (2013) Disentangling the neuroanatomical correlates of perseveration from unilateral spatial neglect. Behav Neurol 26:131–138. CrossRef Medline

Koechlin E (2008) The cognitive architecture of human lateral prefrontal cortex. In: Sensorimotor foundations of higher cognition (Haggard P, Rosetti Y, Kawato M, eds), pp 483–509. Oxford: Oxford UP.

Koechlin E, Hyafil A (2007) Anterior prefrontal function and the limits of human decision-making. Science 318:594–598. CrossRef Medline

Kovach CK, Daw ND, Rudrauf D, Tranel D, O'Doherty JP, Adolphs R (2012) Anterior prefrontal cortex contributes to action selection through tracking of recent reward trends. J Neurosci 32:8434–8442. CrossRef Medline

Krajbich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci 13:1292–1298. CrossRef Medline

Krajbich I, Bartling B, Hare T, Fehr E (2015) Rethinking fast and slow based on a critique of reaction-time reverse inference. Nat Commun 6:7455. CrossRef Medline

Krebs JR, Kacelnik A, Taylor P (1978) Test of optimal sampling by foraging great tits. Nature 275:27–31. CrossRef

Laureiro-Martínez D, Canessa N, Brusoni S, Zollo M, Hare T, Alemanno F, Cappa SF (2013) Frontopolar cortex and decision-making efficiency: comparing brain activity of experts with different professional background during an exploration–exploitation task. Front Hum Neurosci 7:927. CrossRef Medline

Levy R (2012) Apathy: a pathology of goal-directed behaviour: a new concept of the clinic and pathophysiology of apathy. Rev Neurol (Paris) 168:585–597. CrossRef Medline

McNickle GG, Cahill JF Jr (2009) Plant root growth and the marginal value theorem. Proc Natl Acad Sci U S A 106:4747–4751. CrossRef Medline

Nevo I, Erev I (2012) On surprise, change, and the effect of recent outcomes. Front Psychol 3:24. CrossRef Medline

Nitsche MA, Paulus W (2000) Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 527:633–639. CrossRef Medline

Nitsche MA, Paulus W (2001) Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology 57:1899–1901. CrossRef Medline

Nitsche MA, Schauenburg A, Lang N, Liebetanz D, Exner C, Paulus W, Tergau F (2003) Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. J Cogn Neurosci 15:619–626. CrossRef Medline

Nitsche MA, Doemkes S, Karaköse T, Antal A, Liebetanz D, Lang N, Tergau F, Paulus W (2007) Shaping the effects of transcranial direct current stimulation of the human motor cortex. J Neurophysiol 97:3109–3117. CrossRef Medline

Payzan-Lenestour E, Bossaerts P (2012) Do not bet on the unknown versus try to find out more: estimation uncertainty and "unexpected uncertainty" both modulate exploration. Front Neurosci 6:150. CrossRef Medline

Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria.

Pollmann S, Mahn K, Reimann B, Weidner R, Tittgemeyer M, Preul C, Müller HJ, von Cramon DY (2007) Selective visual dimension weighting deficit after left lateral frontopolar lesions. J Cogn Neurosci 19:365–375. CrossRef Medline

Pratt SC, Sumpter DJ (2006) A tunable algorithm for collective decision-making. Proc Natl Acad Sci U S A 103:15906–15910. CrossRef Medline

R Core Team (2014) R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Ramnani N, Owen AM (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat Rev Neurosci 5:184–194. CrossRef Medline

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545–556. CrossRef Medline

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27. Medline

Shenhav A, Straccia MA, Cohen JD, Botvinick MM (2014) Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. Nat Neurosci 17:1249–1254. CrossRef Medline

Stephens DW, Krebs JR (1986) Foraging theory. Princeton, NJ: Princeton UP.

Steyer R, Schwenkmezger P, Notz P, Eid M (1997) Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF). Göttingen: Hogrefe-Huber.

Stuss DT (2011) Traumatic brain injury: relation to executive dysfunction and the frontal lobes. Curr Opin Neurol 24:584–589. CrossRef Medline

Terzuolo CA, Bullock TH (1956) Measurement of imposed gradient adequate to modulate neuronal firing. Proc Natl Acad Sci U S A 42:687–694. CrossRef Medline

Tobler PN, O'Doherty JP, Dolan RJ, Schultz W (2006) Human neural learning depends on reward prediction errors in the blocking paradigm. J Neurophysiol 95:301–310. CrossRef Medline

Tsujimoto S, Genovesio A, Wise SP (2010) Evaluating self-generated decisions in frontal pole cortex of monkeys. Nat Neurosci 13:120–126. CrossRef Medline

Tsujimoto S, Genovesio A, Wise SP (2011) Frontal pole cortex: encoding ends at the end of the endbrain. Trends Cogn Sci 15:169–176. CrossRef Medline

Wagner AD, Desmond JE, Glover GH, Gabrieli JD (1998) Prefrontal cortex and recognition memory: functional-MRI evidence for context-dependent retrieval processes. Brain 121:1985–2002. CrossRef Medline

Watkinson S, Boddy L, Burton K (2005) New approaches to investigating the function of mycelial networks. Mycologist 19:11–17. CrossRef

Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD (2014) Humans use directed and random exploration to solve the explore-exploit dilemma. J Exp Psychol Gen 143:2074–2081. CrossRef Medline