A causal role for right frontopolar cortex in directed, but not random, exploration

Wojciech Zajkowski (wzajkowski@st.swps.edu.pl)
University of Social Sciences and Humanities
Warsaw, Poland

Malgorzata Kossut
Departament of Psychology
University of Social Sciences and Humanities
Warsaw, Poland

Robert C. Wilson (bob@arizona.edu)
Department of Psychology and Cognitive Science Program
University of Arizona
Tucson AZ, USA

Abstract

The explore-exploit dilemma occurs anytime we must choose between exploring unknown options for information and exploiting known resources for reward. Previous work suggests that people use two different strategies to solve the explore-exploit dilemma: directed exploration, driven by information seeking, and random exploration, driven by decision noise. Here, we show that these two strategies rely on different neural systems. Using transcranial magnetic stimulation to inhibit the right frontopolar cortex, we were able to selectively inhibit directed exploration while leaving random exploration intact. This suggests a causal role for right frontopolar cortex in directed, but not random, exploration and that directed and random exploration rely on (at least partially) dissociable neural systems.

Keywords: Explore-exploit, decision making, transcranial magnetic stimulation, frontal pole

Introduction

In an uncertain world, adaptive behavior requires us to carefully balance the exploration of new opportunities with the exploitation of known resources. Finding the optimal balance between exploration and exploitation is a hard computational problem and there is considerable interest in how humans and animals strike this balance in practice (Hills et al.,2015). Recent work has suggested that humans use two distinct strategies to solve the explore-exploit dilemma: directed exploration, based on information seeking, and random exploration, based on decision noise (Wilson, Geana, White, Ludvig, & Cohen,2014). Even though both of these strategies serve the same purpose, i.e. balancing exploration and exploitation, it is likely they rely on different cognitive mechanisms. Directed exploration is driven by information and is thought to be computationally complex. On the other hand, random exploration can be implemented in a simpler fashion by using neural or environmental noise to randomize choice.

Of particular interest is the right frontopolar cortex (RFPC) – an area that has been associated with a number of functions, such as tracking alternate options (Boorman, Behrens, Woolrich, & Rushworth,2009), strategies (Domenech & Koechlin,2015) and goals (Pollmann,2016) that may be important for exploration. In addition, a number of studies have implicated the frontal pole in exploration itself (Badre, Doll, Long, & Frank,2012;Daw, O'Doherty, Dayan, Seymour, & Dolan,2006), although importantly, how exploration is defined varies from paper to paper. In one line of work, exploration is defined as information seeking. Understood this way, exploration correlates with FPC activity measured via fMRI (Badre et al.,2012), suggesting a role for FPC in directed exploration. However, in another line of work, exploration is operationalized differently, as choosing the low value option, not the most informative. Such a measure of exploration is more consistent with random exploration where decision noise drives the sampling of low value options by chance. Defined in this way, exploratory choice correlates with FPC activation (Daw et al.,2006) and stimulation and inhibition of RFPC with direct current (tDCS) can increase and decrease the frequency with which such exploratory choices occur (Raja Beharelle, Polania, Hare, & Ruff,2015).

Taken together, these two sets of findings suggest that lateral FPC plays a crucial role in both directed and random exploration. However, we believe that such a conclusion is premature because of a subtle confound that arises between reward and information in most explore-exploit tasks. This confound arises because participants only gain information from the options they choose, yet are incentivized to choose more rewarding options. Thus, over many trials, participants gain more information about more rewarding options such that the two ways of defining exploration, choosing high information or low reward options, become confounded (Wilson et al.,2014). This makes it impossible to tell whether the link between FPC and exploration is specific to directed exploration, random exploration, or whether it is general to both.

To distinguish these interpretations and investigate the causal role of RFPC in directed and random exploration, we used continuous theta-burst TMS (cTBS) (Huang, Edwards, Rounis, Bhatia, & Rothwell,2005) to selectively inhibit RFPC in participants performing the 'Horizon Task', an explore-exploit task specifically designed to separate directed and random exploration (Wilson et al.,2014). Using this task we find that RFPC inhibition selectively inhibits directed exploration while leaving random exploration intact.

Methods

Participants

31 healthy right-handed, adult volunteers (19 female, 12 male; ages 19-32) took part in the study. 6 participants were excluded from the analysis due to chance-level performance or for failure to to return for the second session leaving 25

participants (13 female, 12 male, ages 19-32) for the main analysis. All participants were informed about potential risks connected to TMS and signed a written consent. The study was approved by University of Social Sciences and Humanities ethics committee.

TMS procedure

All TMS was delivered in line with established safety guidelines (Rossi, Hallett, Rossini, Pascual-Leone, & Safety of TMS Consensus Group,2009). There were two experimental TMS sessions (targeting RFPC and vertex, as a control) and a preceding MRI session in which a T1 structural image was acquired in order to target frontal pole. During the TMS sessions, resting motor thresholds were obtained first and then the cTBS procedure took place. This involved 40 second of stimulation at 50Hz at 80% resting motor threshold, a protocol that is thought to decrease cortical excitability for up to 50 minutes (Wischnewski & Schutter,2015). Participants began the main task immediately after stimulation. The two experimental sessions were performed with an intersession interval of at least 5 days. All sessions took place at Nencki Institute of Experimental Biology in Warsaw. Based on previous fMRI work showing a link between FPC and exploration (Daw et al.,2006;Badre et al.,2012), RFPC stimulation was targeted at [x,y,z] = [35,50,15] in MNI (Montreal Neurological Institute) space. Vertex corresponded to the Cz position of the 10-20 EEG system.

Behavioral task

We used our previously published 'Horizon Task' (Figure 1) to measure the effects of TMS stimulation of RFPC on directed and random exploration. In this task, participants play a set of games in which they make choices between two slot machines (one-armed bandits) that pay out rewards from different Gaussian distributions. To maximize their rewards in each game, participants need to exploit the slot machine with the highest mean, but they cannot identify this best option without exploring both options first.

The Horizon Task has two key manipulations that allow us to measure directed and random exploration. The first manipulation is the horizon itself, i.e. the number of decisions remaining in each game. The idea behind this manipulation is that when the horizon is long (6 trials), participants should explore more frequently, because any information they acquire from exploring can be used to make better choices later on. In contrast, when the horizon is short (1 trial), participants should exploit the option they believe to be best. Thus, this task allows us to quantify directed and random exploration as changes in information seeking and behavioral variability that occur with horizon.

The second manipulation is the amount of information participants have about each option *before* making their first choice. This information manipulation is achieved by using four forced-choice trials, in which participants are told which option to pick, at the start of each game. We use these forced-choice trials to setup one of two information conditions: an unequal, or [1 3], condition, in which participants see 1 play from one option and 3 plays from the other option, and an unequal, or [2 2], condition, in which participants see two outcomes from both options. The two information conditions allow us to quantify directed and random exploration in a model-free manner (Figure 1). In particular, directed exploration, which involves information seeking, can be quantified as the probability of choosing the high information option, $p(\text{high info})$ in the [1 3] condition, while random exploration, which involves decision noise, can be quantified as the probability of making a mistake, or choosing the low mean reward option, $p(\text{low mean})$, in the [2 2] condition. Crucially, if $p(\text{high info})$ and $p(\text{low mean})$ increase with horizon, then we infer that participants are using directed and random exploration.

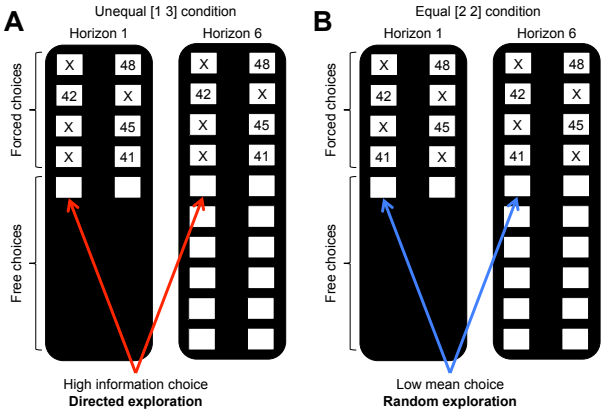

Figure 1: The Horizon Task. Participants make a series of decisions between two one-armed bandits that pay out probabilistic rewards with unknown means. At the start of each game, 'forced-choice' trials give participants partial information about the mean of each option. We use the forced-choice trials to set up one of two information conditions: (A) an unequal (or [1 3]) condition in which participants see 1 play from one option and 3 plays from the other and (B) an equal (or [2 2]) condition in which participants see 2 plays from both options. A model-free measure of directed exploration is then defined as the change in information seeking with horizon in the unequal condition (A). Likewise a model-free measure of random exploration is defined as the change choosing the low mean option in the equal condition (B).

Model-based analysis

While the model-free analyses are intuitive, the model-free statistics, $p(\text{high info})$ and $p(\text{low mean})$, are not pure reflections of information seeking and behavioral variability and could be influenced by other factors such as spatial bias and learning. To account for these possibilities we performed a model-based analysis using a model that extends our earlier work (Wilson et al.,2014;Somerville et al.,2017). In this model, the level of directed and random exploration is captured by two parameters: an information bonus for directed

exploration, and decision noise for random exploration. In addition the model includes terms for the spatial bias and to describe learning. The model naturally decomposes into a learning component and a decision component and we consider each of these components in turn.

Learning component The learning component of the model assumes that participants use a Kalman filter (Kalman,1960) to learn a value for the mean reward of each option. In particular, we assume that participants use a generative model of the task in which the rewards from each bandit, r_t, are generated from Gaussian distribution with a fixed standard deviation, σ_r, and a mean, m_t^i, that is different for each bandit and can vary over time. The time dependence of the mean is determined by a Gaussian random walk with mean 0 and standard deviation σ_d. Note that this generative model, assumed by the Kalman filter, is slightly different to the true generative model used in the Horizon Task, which assumes that the mean of each bandit is constant over time, i.e. $\sigma_d = 0$. This mismatch between the assumed and actual generative models, is quite deliberate and allows us to account for the suboptimal learning of the subjects. In particular, this mismatch introduces the possibility of a recency bias (when $\sigma_d > 0$) whereby more recent rewards are over-weighted in the computation of R_t^i.

The actual equations of the Kalman filter model are straightforward. The model keeps track of an estimate of both the mean reward, R_t^i, of each option, i, and the uncertainty in that estimate, σ_t^i. When option i is played on trial t, these two parameters update according to

$$\begin{aligned} R_{t+1}^i &= R_t^i + \frac{(\sigma_{t+1}^i)^2}{\sigma_r^2}(r_t - R_t^i) \\ \frac{1}{(\sigma_{t+1}^i)^2} &= \frac{1}{(\sigma_t^i)^2 + \sigma_d^2} + \frac{1}{\sigma_r^2} \end{aligned} \tag{1}$$

When option i is not played on trial t we assume that the estimate of the mean stays the same, but that the uncertainty in this estimate grows as the generative model assumes the mean drifts over time. Thus for unchosen option j we have

$$R_{t+1}^j = R_t^j \qquad \text{and} \qquad (\sigma_{t+1}^j)^2 = (\sigma_t^j)^2 + \sigma_d^2 \tag{2}$$

When the option is played, the update equation for R_t^i is essentially just a 'delta rule' (Rescorla, Wagner, et al.,1972), with the estimate of the mean being updated in proportion to the prediction error, $r_t - R_t^i$. This relationship to the reinforcement learning literature is made more clear by rewriting the learning equations in terms of the time varying learning rate, $\alpha_t^i = (\sigma_{t+1}^i)^2/\sigma_r^2$ Written in terms of this learning rate, equations 1 become

$$R_{t+1}^i = R_t^i + \alpha_t^i(r_t - R_t^i) \quad \text{and} \quad \frac{1}{\alpha_t^i} = \frac{1}{\alpha_{t-1}^i + \alpha_d} + 1 \tag{3}$$

where $\alpha_d = \sigma_d^2/\sigma_r^2$. The learning model has four free parameters: the noise variance, σ_r^2, the drift variance, σ_d^2, and the initial values of the estimated reward, R_0, and uncertainty in that variance estimate, σ_0^2. In practice, only three of these parameters are identifiable from behavioral data, and we will find it useful to reparameterize the learning model in terms of R_0 and an initial, α_0, and asymptotic, α_∞, learning rate. In particular, the initial value of the learning rate relates to σ_0 and σ_r as $\alpha_0 = \sigma_0^2/\sigma_r^2$, while the asymptotic value of the learning rate, which corresponds to the steady state value of α_t^i if option i is played forever, relates to α_d (and hence σ_d and σ_r) as

$$\alpha_\infty = \frac{1}{2}\left(-\alpha_d + \sqrt{\alpha_d^2 + 4\alpha_d}\right) \tag{4}$$

Decision component Once the payoffs of each option, R_t^i, have been estimated from the outcomes of the forced-choice trials, the model makes a decision using a simple logistic choice rule:

$$p(\text{choose right}) = \frac{1}{1 + \exp\left(\frac{\Delta R + A\Delta I + B}{\sigma}\right)} \tag{5}$$

where ΔR ($= R_t^{left} - R_t^{right}$) is the difference in expected reward between left and right options and ΔI is the difference in information between left and right options (which we define as +1 when left is more informative, -1 when right is more informative, and 0 when both options convey equal information in the [2 2] condition). The three free parameters of the decision process are: the information bonus, A, the spatial bias, B, and the decision noise σ. We assume that these three decision parameters can take on different values in the different horizon and uncertainty conditions (with the proviso that A is undefined in the [2 2] information condition since $\Delta I = 0$). Thus the decision component of the model has 10 free parameters (A in the two horizon conditions, and B and σ in the 4 horizon x uncertainty conditions). Directed exploration is then quantified as the change in information bonus with horizon, while random exploration is quantified as the change in decision noise with horizon.

Model Fitting

Hierarchical Bayesian Model Between the learning and decision components of the model, each subject's behavior is described by 13 free parameters, all of which are allowed to vary between TMS conditions. These parameters are: the initial mean, R_0, the initial learning rate, α_0, the asymptotic learning rate, α_∞, the information bonus, A, in both horizon conditions, the spatial bias, B, in the four horizon x uncertainty conditions, and the decision noise, σ, in the four horizon x uncertainty conditions (Table 1, Figure 2).

We fit each of the free parameters to the behavior of each subject using a hierarchical Bayesian approach (Lee & Wagenmakers,2014). In this approach to model fitting, each parameter for each subject is assumed to be sampled from a group-level prior distribution whose parameters, the so-called 'hyperparameters', are estimated using a Markov Chain Monte Carlo (MCMC) sampling procedure.

Parameter	Prior	Hyperparameters	Hyperprior
prior mean, $R_0^{\tau s}$	$R_0^{\tau s} \sim \text{Gaussian}(\mu_{R_0}^{\tau}, \sigma_{R_0}^{\tau})$	$\theta_{R_0}^{\tau} = (\mu_{R_0}^{\tau}, \sigma_{R_0}^{\tau})$	$\mu_{R_0}^{\tau} \sim \text{Gaussian}(50, 14)$ $\sigma_{R_0}^{\tau} \sim \text{Gamma}(1, 0.001)$
initial learning rate, $\alpha_0^{\tau s}$	$\alpha_0^{\tau s} \sim \text{Beta}(a_{\alpha_0}^{\tau}, b_{\alpha_0}^{\tau})$	$\theta_{\alpha_0}^{\tau} = (a_{\alpha_0}^{\tau}, b_{\alpha_0}^{\tau})$	$a_{\alpha_0}^{\tau} \sim \text{Uniform}(0.1, 10)$ $b_{\alpha_0}^{\tau} \sim \text{Uniform}(0.5, 10)$
asymptotic learning rate, $\alpha_\infty^{\tau s}$	$\alpha_\infty^{\tau s} \sim \text{Beta}(a_{\alpha_\infty}^{\tau}, b_{\alpha_\infty}^{\tau})$	$\theta_{\alpha_\infty}^{\tau} = (a_{\alpha_\infty}^{\tau}, b_{\alpha_\infty}^{\tau})$	$a_{\alpha_\infty}^{\tau} \sim \text{Uniform}(0.1, 10)$ $b_{\alpha_\infty}^{\tau} \sim \text{Uniform}(0.1, 10)$
information bonus, $A^{\tau shu}$	$A^{\tau shu} \sim \text{Gaussian}(\mu_A^{\tau hu}, \sigma_A^{\tau hu})$	$\theta_A^{\tau hu} = (\mu_A^{\tau hu}, \sigma_A^{\tau hu})$	$\mu_A^{\tau hu} \sim \text{Gaussian}(0, 100)$ $\sigma_A^{\tau hu} \sim \text{Gamma}(1, 0.001)$
spatial bias, $B^{\tau shu}$	$B^{\tau shu} \sim \text{Gaussian}(\mu_B^{\tau hu}, \sigma_B^{\tau hu})$	$\theta_B^{\tau hu} = (\mu_B^{\tau hu}, \sigma_B^{\tau hu})$	$\mu_B^{\tau hu} \sim \text{Gaussian}(0, 100)$ $\sigma_B^{\tau hu} \sim \text{Gamma}(1, 0.001)$
decision noise, $\sigma^{\tau shu}$	$\sigma^{\tau shu} \sim \text{Gamma}(k_\sigma^{\tau hu}, \lambda_\sigma^{\tau hu})$	$\theta_\sigma^{\tau hu} = (k_\sigma^{\tau hu}, \lambda_\sigma^{\tau hu})$	$k_\sigma^{\tau hu} \sim \text{Exp}(0.1)$ $\lambda_\sigma^{\tau hu} \sim \text{Exp}(10)$

Table 1: Model parameters, priors, hyperparameters and hyperpriors.

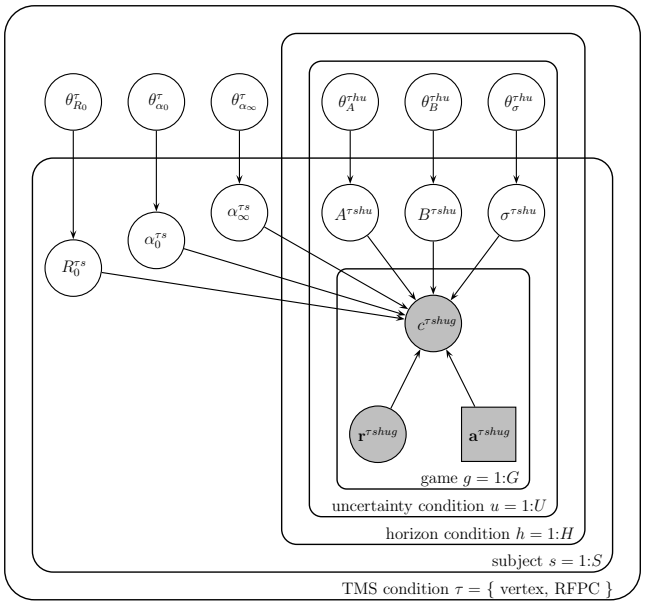

Figure 2: Graphical representation of the model. Each variable is represented by a node, with edges denoting the dependence between variables. Shaded nodes correspond to observed variables, i.e. the free choices $c^{\tau shug}$, forced-trial rewards, $\mathbf{r}^{\tau shug}$ and forced-trial choices $\mathbf{a}^{\tau shug}$. Unshaded nodes correspond to unobserved variables whose values are inferred by the model.

The hyper-parameters themselves are assumed to be sampled from 'hyperprior' distributions whose parameters are defined such that these hyperpriors are broad. For notational convenience, we refer to the hyperparameters that define the prior for variable X as θ^X. In addition we use superscripts to refer to the dependence of both parameters and hyperparameters on TMS stimulation condition, τ, horizon condition, h, uncertainty condition, u, subject, s, and game, g.

The particular priors and hyperpriors for each parameter are shown in Table 1. For example, we assume that the prior mean, $R_0^{\tau s}$, for each stimulation condition τ and horizon condition h, is sampled from a Gaussian prior with mean $\mu_{R_0}^{\tau}$ and standard deviation $\sigma_{R_0}^{\tau}$. These prior parameters are sampled in turn from their respective hyperpriors: $\mu_{R_0}^{\tau}$, from a Gaussian distribution with mean 50 and standard deviation 14, $\sigma_{R_0}^{\tau}$ from a Gamma distribution with shape parameter 1 and rate parameter 0.001.

Model fitting using MCMC The model was fit to the data using a Markov Chain Monte Carlo approach implemented in the JAGS package (Plummer et al.,2003) via the MATJAGS interface (psiexp.ss.uci.edu/research/programs_data/jags/). This package approximates the posterior distribution over model parameters by generating samples from this posterior distribution given the observed behavioral data. In particular we used 4 independent Markov chains to generate 4000 samples from the posterior distribution over parameters (1000 samples per chain). Each chain had a burn in period of 500 samples, which were discarded to reduce the effects of initial conditions, and posterior samples were acquired at a thin rate of 1. Convergence of the Markov chains was confirmed post hoc by eye.

Results

RFPC stimulation selectively inhibits directed exploration on the first free-choice

Model-free analysis Using the measures of directed and random exploration, $p(\text{high info})$ and $p(\text{low mean})$, we found that inhibiting the RFPC had a significant effect on directed exploration but not random exploration (Figure 3A, B). For directed exploration, a repeated measures ANOVA with horizon, TMS condition and order as factors revealed a significant interaction between stimulation condition and horizon ($F(1,24) = 4.96,\ p = 0.036$). Conversely, a similar analysis for random exploration revealed no effects of stimulation condition (main effect of stimulation condition, $F(1,24) = 0.88,\ p = 0.36$; interaction of stimulation condition with horizon, $F(1,24) = 1.24,\ p = 0.28$). Post hoc analyses revealed that the change in directed exploration was driven by changes in information seeking in horizon 6 (one-sided t-test, $t(24) = 2.62,\ p = 0.008$) and not in horizon 1

(two-sided t-test, $t(24) = -0.30$; $p = 0.77$).

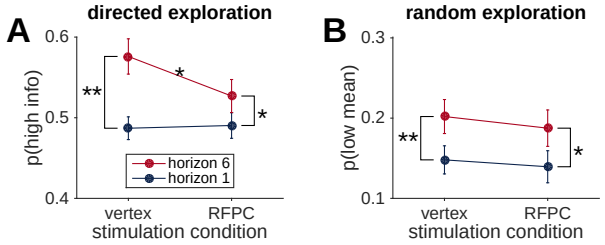

Figure 3: Model-free analysis of the first free-choice trial shows that RPFC stimulation affects directed, but not random, exploration. (A) In the control (vertex) condition, information seeking increases with horizon, consistent with directed exploration. When RFPC is stimulated, directed exploration is reduced, an effect that is entirely driven by changes in horizon 6 (* denotes $p < 0.02$ and ** denotes $p < 0.005$; error bars are $\pm$ s.e.m.). (B) Random exploration increases with horizon but is not affected by RFPC stimulation.

Model-based analysis Posterior distributions over the group-level means of all 13 parameters in the model are shown in Figure 4. The left column of Figure 4 shows the posteriors over each parameter while the right column shows the posteriors over the TMS-related change in each parameter. Both columns suggest a selective effect of RFPC stimulation on the information bonus in horizon 6.

Focussing on the left column first, overall the parameter values seem reasonable. The prior mean is close to the generative mean of 50 used in the actual experiment, and the decision parameters are comparable to those found in our previous work (Wilson et al.,2014). The learning rate parameters, α_0 and α_∞, were not included in our previous models and are worth discussing in more detail. As expected for Bayesian learning (Kalman,1960), the initial learning rate is higher than the asymptotic learning rate (95% of samples in the vertex condition, 94% in the RFPC condition). However, the actual values of the learning rates are quite far from their 'optimal' settings of $\alpha_0 = 1$ and $\alpha_\infty = 0$ that would correspond to perfectly computing the mean reward. This suggests a greater than optimal reliance on the prior ($\alpha_0 < 1$) and a pronounced recency bias ($\alpha_\infty > 0$) such that the most recent rewards are weighted more heavily in the computation of expected reward, R_t^i. Both of these findings are likely due to the fact that the version of the task we employed did not keep the outcomes of the forced trials on screen and instead relied on people's memories to compute the expected value.

Turning to the right hand column of Figure 4, we can see that the model-based analysis yields similar result to the model-free analysis. In particular we see a reduction (of about 4.8 points) in the information bonus in horizon 6 (with 99% of samples showing a reduced information bonus in the RFPC stimulation condition) and no effect on decision noise in either horizon in either the [2 2] or [1 3] uncertainty conditions (with between 40% and 63% of samples below zero).

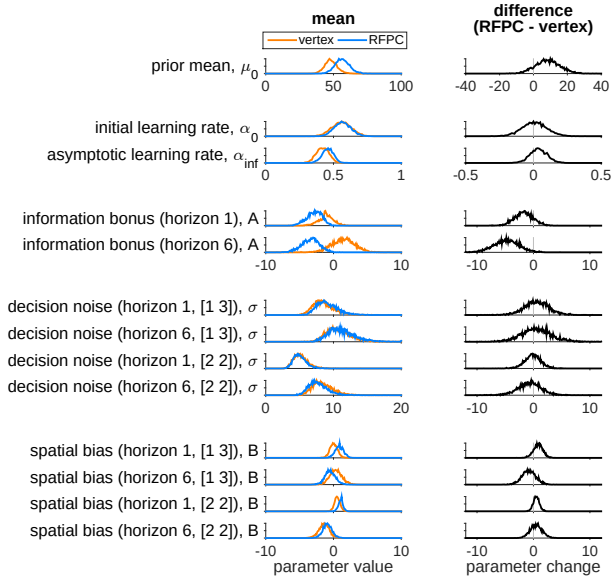

Figure 4: Model-based analysis of the first free-choice trial showing the effect of RFPC stimulation on each of the 13 parameters. Left column: Posterior distributions over each parameter value for RFPC and vertex stimulation condition. Right column: posterior distributions over the change in each parameter between stimulation conditions. Note that, because information bonus, decision noise and spatial bias are all in units of points, we plot them on the same scale to facilitate comparison of effect size.

Discussion

In this work we used continuous theta-burst transcranial magnetic stimulation (cTBS) to investigate whether right frontopolar cortex (RFPC) is causally involved in directed and random exploration. Using a task that is able to behaviorally dissociate these two types of exploration, we found that inhibition of RFPC caused a selective reduction in directed, but not random exploration. To the best of our knowledge, this finding represents the first causal evidence that directed and random exploration rely on dissociable neural systems and is consistent with our recent findings showing that directed and random exploration have different developmental profiles (Somerville et al.,2017). This suggests that, contrary to the assumption underlying many contemporary studies (Daw et al.,2006;Badre et al.,2012), exploration is not a unitary process, but a dual process in which the distinct strategies of information seeking and choice randomization are implemented via distinct neural systems.

Such a dual-process view of exploration is consistent with the classical idea that there are multiple types of exploration (Berlyne,1966). In particular Berlyne's constructs of 'specific exploration', involving a drive for information, and 'diver-

sive exploration', involving a drive for variety, bear a striking resemblance to our definitions of directed and random exploration. Despite the importance of Berlyne's work, more modern views of exploration tend not to make the distinction between different types of exploration, considering instead a single exploratory state or exploratory drive that controls information seeking across a wide range of tasks (Hills et al.,2015;Kidd & Hayden,2015). At face value, such unitary accounts seem at odds with a dual-process view of exploration. However, these two viewpoints can be reconciled if we allow for the possibility that, while directed and random exploration are implemented by different systems, their levels are set by a common exploratory drive. More work will be required to determine whether this is the case.

While the present study does allow us to conclude that directed and random exploration rely on different neural systems, the limited spatial specificity of TMS limits our ability to say exactly what those systems are. In particular, because the spatial extent of TMS is quite large, stimulation aimed at frontal pole may directly affect activity in nearby areas such as ventromedial prefrontal cortex (vmPFC) and orbitofrontal cortex (OFC), both areas that have been implicated in exploratory decision making and that may be contributing to our effect (Daw et al.,2006). In addition to such direct effects of TMS on nearby regions, indirect changes in areas that are connected to the frontal pole could also be driving our effect. For example, cTBS of left frontal pole has been associated with changes in blood perfusion in areas such as amygdala, fusiform gyrus and posterior parietal cortex (Volman, Roelofs, Koch, Verhagen, & Toni,2011). In addition the same study showed that unilateral cTBS of left frontal pole is associated with changes in blood perfusion to the right frontal pole. Indeed, such a bilateral effect of cTBS may explain why our intervention was effective at all given that a number of neuroimaging studies have shown bilateral activation of the frontal pole associated with exploration (Daw et al.,2006;Badre et al.,2012). Future work combining cTBS with neuroimaging will be necessary to shed light on these issues.

References

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012, 9 February). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, *73*(3), 595–607.

Berlyne, D. E. (1966, 1 July). Curiosity and exploration. *Science*, *153*(3731), 25–33.

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009, 11 June). How green is the grass on the other side? frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*(5), 733–743.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006, 15 June). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, *1*, 101–106.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & Cognitive Search Research Group. (2015, January). Exploration versus exploitation in space, mind, and society. *Trends Cogn. Sci.*, *19*(1), 46–54.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005, 20 January). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Int. J. Eng. Trans. A*, *82*(1), 35.

Kidd, C., & Hayden, B. Y. (2015, 4 November). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Plummer, M., et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125).

Pollmann, S. (2016). Frontopolar resource allocation in human and nonhuman primates. *Trends Cogn. Sci.*, *20*(2), 84–86.

Raja Beharelle, A., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial stimulation over frontopolar cortex elucidates the choice attributes and neural mechanisms used to resolve Exploration-Exploitation Trade-Offs. *Journal of Neuroscience*, *35*(43), 14544–14556.

Rescorla, R. A., Wagner, A. R., et al. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64–99.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol.*, *120*(12), 2008–2039.

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., et al. (2017, February). Charting the expansion of strategic exploratory behavior during adolescence. *J. Exp. Psychol. Gen.*, *146*(2), 155–164.

Volman, I., Roelofs, K., Koch, S., Verhagen, L., & Toni, I. (2011, 25 October). Anterior prefrontal cortex inhibition impairs control over social emotional actions. *Curr. Biol.*, *21*(20), 1766–1770.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014, December). Humans use directed and random exploration to solve the explore-exploit dilemma. *J. Exp. Psychol. Gen.*, *143*(6), 2074–2081.

Wischnewski, M., & Schutter, D. J. L. G. (2015, July). Efficacy and time course of theta burst stimulation in healthy humans. *Brain Stimul.*, *8*(4), 685–692.