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

Wojciech Zajkowski[1], Malgorzata Kossut[2,3], & Robert C. Wilson[4,5]

[1]University of Social Sciences and Humanities, Warsaw, Poland
[2]Departament of Psychology, University of Social Sciences and Humanities, Warsaw, Poland
[3]Nencki Institute, Warsaw, Poland
[4]Department of Psychology, University of Arizona, Tucson AZ USA
[5]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 selectively inhibit right frontopolar cortex, we were able to selectively inhibit directed exploration while leaving random exploration intact, suggesting a causal role for right frontopolar cortex in directed, but not random, exploration.

Results and Discussion

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 (Badre et al, 2012; Cavanagh et al, 2011; Cohen et al, 2007; Daw et al, 2006; Frank et al, 2009; Hills et al, 2015; Mehlhorn et al, 2015, Wilson et al, 2014). 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 et al, 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 (Wilson et al, 2014). On the other hand, random exploration can be implemented in a simpler fashion by using neural or environmental noise to randomize choice.

A reasonable, biologically-guided assumption is that these two types of exploration rely on separate brain mechanisms. 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 et al, 2009), strategies (Domenech & Koechlin, 2015) and goals (Pollmann, 2015) that may be important for exploration. In addition, a number of studies have implicated the frontal pole in exploration itself, although importantly, how exploration is defined varies from paper to paper. In one line of work, exploration is defined as information seeking. Defined this way, exploration correlates with FP activity measured via fMRI (Badre et al, 2012) and a frontal theta component in EEG (Cavanagh et al, 2011), suggesting a role for FP in *directed* exploration. However, in another line of work, exploration is defined 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 FP activation (Daw et al, 2006) and stimulation and inhibition of FP with direct current (tDCS) can increase and decrease the frequency with which such exploratory choices occur (Beharelle et al, 2015).

Taken together, these two sets of findings suggest that FP 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 they 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 FP and exploration is specific to either directed or 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 (Huang et al, 2005) to selectively inhibit RFPC in fifteen 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 evidence that RFPC inhibition selectively inhibits directed exploration while leaving random exploration intact.

We used our previously published "Horizon Task" (Figure 1) to measure the effects of TMS stimulation to 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.

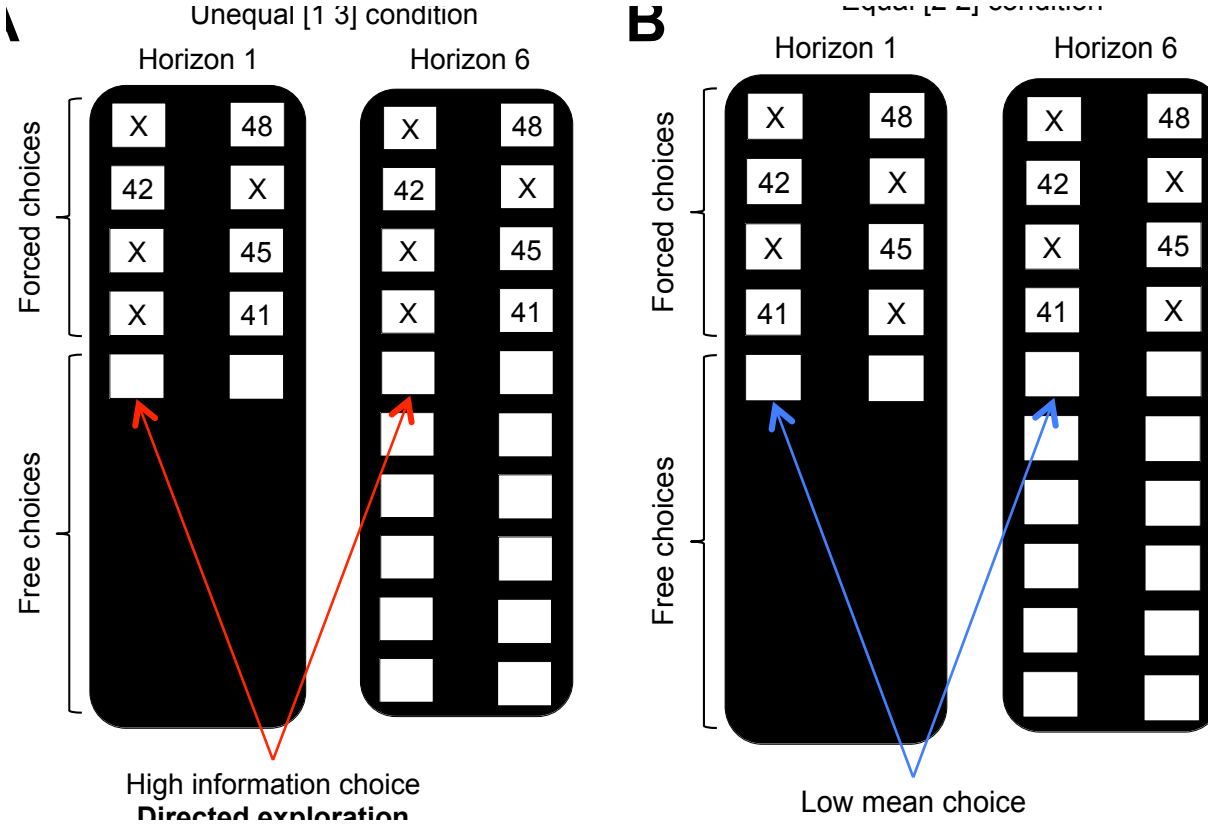

Figure 1 – The Horizon Task. Participants make a series of decisions between two one-armed bandits that pay out probabilistic rewards with unknown means. At the start of each game, forced-choice trials give participants partial information about the mean of each option setting up one of two information 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. Directed exploration is then defined as the change in information seeking with horizon in the unequal condition (**A**). Random exploration is defined as the change choosing the low mean option in the equal condition (**B**).

84 The key manipulation in the Horizon Task is the time horizon, the number of decisions
85 remaining in each game. The idea behind this manipulation is that when the horizon is long
86 (6 trials), participants should explore, because any information they acquire can be used to
87 make better choices later on. In contrast, when the horizon is short (1 trial), participants
88 should exploit the option they believe to be best. Thus, by measuring *changes* in information
89 seeking and behavioral variability that occur with horizon, this task allows us to quantify
90 directed and random exploration.

91 The Horizon Task also allows us to remove the reward-information confound with the use of
92 "forced-choice" trials at the start of each game. These forced-choice trials setup one of two
93 information conditions: an unequal (or [1 3]) condition (Figure 1A), in which one option is
94 played once and the other three times, and an equal (or [2 2]) condition (Figure 1B), in which
95 both options are played twice. Information seeking is quantified as the probability of
96 choosing the more uncertain option in the [1 3] condition (i.e. the option played once in the
97 forced-choice trials), p(high info). Behavioral variability is quantified as the number of
98 mistakes, i.e. choosing the low value option, in the [2 2] condition, p(low mean). To remove
99 the reward-information confound, both of these measures are computed on the *first* free-
100 choice trial in each game, i.e. before participants' choices have a chance to confound
101 information and reward.

102 Using these measures of exploration, we found that inhibiting the RFPC had a significant
103 effect on directed exploration but not random exploration (Figure 2A, B). In particular, for
104 directed exploration, a repeated measures ANOVA revealed a significant interaction between
105 stimulation condition and horizon ($F(1, 14) = 4.77$, $p = 0.047$). Conversely, a similar analysis
106 for random exploration revealed no effects of stimulation condition (main effect of
107 stimulation condition, $F(1, 14) = 0.26$, $p = 0.62$; interaction of stimulation condition with
108 horizon, $F(1, 14) = 0.01$; $p = 0.93$). Post hoc analyses revealed that the change in directed

exploration was driven by changes in information seeking in horizon 6 (one-sided t-test, $t(14) = 2.26$, $p = 0.02$) and not in horizon 1 (two-sided t-test, $t(14) = -0.40$, $p = 0.69$). Finally, a similar analysis using a logistic model of choice behavior yielded similar findings (see Supplementary Material).

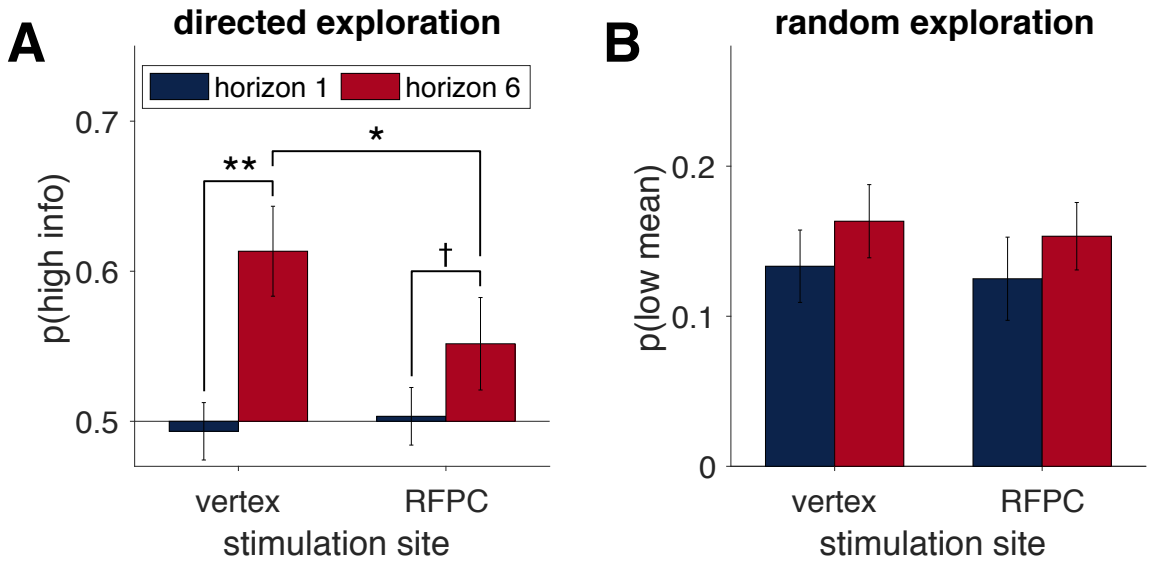

Figure 2 – RPFC stimulation affects directed, but not random, exploration. (**A**) In the control (vertex) condition, information seeking increases with horizon, consistent with directed exploration. When RFPC is stimulated, directed exploration is reduced, an effect that is entirely driven by changes in horizon 6 († denotes $p < 0.1$, * $p < 0.05$, ** $p < 0.01$). (**B**) Random exploration increases with horizon but is not affected by RFPC stimulation.

These results suggest a causal role for RFPC in directed, but not random, exploration. A role in directed exploration is consistent with other findings implicating FP in a number of complex computations. These include tracking the value of the best unchosen option (Boorman et al, 2009), inferring the reliability of alternate strategies (Domenech & Koechlin, 2015), arbitrating between old and new strategies (Donoso et al, 2015; Mansouri et al, 2015),

and reallocating cognitive resources among potential goals in underspecified situations (Pollmann, 2015). Taken together, these findings suggest a role for frontal pole in decisions that involve long-term planning and the consideration of alternative actions, both crucial for directed exploration.

That frontal pole is *not* involved in random exploration suggests that directed and random exploration rely on (at least partially) dissociable neural systems. Exactly what these systems are is currently unknown, but may include other areas of prefrontal cortex (Badre et al, 2012; Cavanagh et al, 2011; Daw et al, 2006) and may involve modulation by noradrenergic (Cohen et al, 2007) and dopaminergic (Costa et al, 2014) inputs. Clearly more work, involving tasks that can dissociate the two types of exploration, will be required to fully understand the neural substrates of exploratory choice.

Methods

Participants

16 healthy right-handed, adult volunteers (9 female; aged 19-32). One participant (female) was excluded from the analysis due to chance-level performance in both experimental sessions. 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.

Procedure

There were two experimental TMS sessions and a preceding MRI session. On the first session T1 structural images were acquired using a 3T Siemens TRIO scanner. The scanning session lasted up to 10 minutes. Before the first two sessions, participants filled in standard safety questionnaires regarding MRI scanning and TMS. During the experimental sessions,

149 prior to the stimulation participants went through 16 training games to get accustomed to the
150 task. Afterwards, resting motor thresholds were obtained and the stimulation took place.
151 Participants began the main task immediately after stimulation. The two experimental
152 sessions were performed with an intersession interval of at least 5 days. The order of
153 stimulation conditions was counterbalanced across subjects. All sessions took place at Nencki
154 Institute of Experimental Biology in Warsaw.

155 **Stimulation site**

156 The RFPC peak was defined as [x,y,z]= [35,50,15] in MNI (Montreal Neurological Institute)
157 space . The coordinates were based on a number of fMRI findings that indicated RFPC
158 involvement in exploration[1,7] and constrained by the plausibility of stimulation (e.g. defining
159 'z' coordinate lower would result in the coil being placed uncomfortably close to the eyes).
160 Vertex corresponded to the Cz position of the 10-20 EEG system. In order to locate the
161 stimulation sites we used a frameless neuronavigation system (Brainsight software, Rogue
162 Research, Montreal, Canada) with a Polaris Vicra infrared camera (Northern Digital,
163 Waterloo, Ontario, Canada).

164 **TMS protocol**

165 We used continuous theta burst stimulation (cTBS)[13]. cTBS requires 50Hz stimulation at
166 80% resting motor threshold. 40 second stimulation is equivalent to 600 pulses and can
167 decrease cortical excitability for up to 50 minutes (Wischnewski & Schutter, 2015).

168 Individual resting motor thresholds were assessed by stimulating the right motor knob and
169 inspecting if the stimulation caused an involuntary hand twitch in 50% of the cases. We used
170 a MagPro X100 stimulator (MagVenture, Hueckelhoven, Germany) with a 70mm figure-
171 eight coil. The TMS was delivered in line with established safety guidelines (Rossi et al,
172 2009).

Limitations

Defining stimulation target by peak coordinates based on findings from previous studies did not allow to account for individual differences in either brain anatomy or the impact of TMS on brain networks (Gratton et al, 2013). However, a study by Volman and collegues (2011) that used the same theta-burst protocol on the left frontopolar cortex has shown biletaral inhibitory effects on blood perfusion in the frontal pole. This suggests that both right and left parts of the frontopolar cortex might have been inhibited in our experiment, which is consistent with imaging results indicating bilateral involvement of the frontal pole in exploratory decisions.

Task

The task was a modified version of the Horizon Task (Wilson et al, 2014). As in the original paper, the distributions of payoffs tied to bandits were independent between games and drawn from a Gaussian distribution with variable means and fixed SD=8. Participants were informed that in every game one of the bandits is objectively 'better' (has a higher payoff mean). Differences between two means were set to either 4, 8, 12 or 20. One of the means was always equal to either 40 or 60 and the second was set accordingly. The order of games was randomized. Mean sizes and order of presentation were counterbalanced. Participants played 160 games and the whole task lasted between 39 and 50 minutes (m=43.4).

Each game consisted of 5 or 10 choices. Every game started with a screen saying "New game" and information about whether it was a long or short horizon, followed by sequentially presented choices. Every choice was presented on a separate screen, so that participants had to keep previous the scores in memory. There was no time limit for decisions. During forced choices participants had to press the prompted key to move to the next choice. During free choices they could press either 'z' or 'm' to indicate their choice of left or right bandit. The decision could not be made in a time shorter than 200ms, preventing participants from

198 accidentally responding too soon. The score feedback was presented for 500ms. A counter at
199 the bottom of the screen indicated the number of choices left in a given game. The task was
200 programmed using PsychoPy software v1.86 (Peirce, 2007).

201 Participants were rewarded based on points scored in two sessions. The payoff bounds were
202 set between 50 and 80 zl (equivalent to approximately 12 and 19 euro). Participants were
203 informed about their score and monetary reward after the second session.

204

205 ## References

206 Badre, D., Doll, B., Long, N., & Frank, M. (2012). Rostrolateral prefrontal cortex and
207 individual differences in uncertainty-driven exploration. Neuron, 73(3), 595–607.
208 http://doi.org/10.1016/j.neuron.2011.12.025

209 Beharelle, A., R., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial Stimulation
210 over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used
211 to Resolve Exploration-Exploitation Trade-Offs. Journal of Neuroscience, 35(43), 14544–
212 14556. http://doi.org/10.1523/JNEUROSCI.2322-15.2015

213 Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How
214 green is the grass on the other side? Frontopolar cortex and the evidence in favor of
215 alternative courses of action. Neuron, 62(5), 733–43.
216 http://doi.org/10.1016/j.neuron.2009.05.014

217 Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects
218 uncertainty and unexpectedness during exploration and exploitation. Cerebral Cortex,
219 22(11), 2575–2586. http://doi.org/10.1093/cercor/bhr332

220 Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the
221 human brain manages the trade-off between exploitation and exploration. Philosophical

Transactions of the Royal Society of London. Series B, Biological Sciences, 362(1481), 933–42. http://doi.org/10.1098/rstb.2007.2098

Costa, V., Tran, V., Turchi, J., & Averbeck, B. (2014). Dopamine modulates novelty seeking behavior during decision making. Behavioral Neuroscience, 128(5), 556–566. http://doi.org/10.1037/a0037128

Daw, N. D., O'Doherty, J. P., Dayan, P., Dolan, R. J., & Seymour, B. (2006). Cortical substrates for exploratory decisions in humans. Nature, 441(7095), 876–9. http://doi.org/10.1038/nature04766

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. Current Opinion in Behavioral Sciences, 1, 101–106. http://doi.org/10.1016/j.cobeha.2014.10.007

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. Science, 344(6191), 1481–1486. http://doi.org/10.1126/science.1252254

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nature Neuroscience, 12(8), 1062–8. http://doi.org/10.1038/nn.2342

Gratton, C., Lee, T. G., Nomura, E. M., & D'Esposito, M. (2013). The effect of theta-burst TMS on cognitive control networks measured with resting state fMRI. Frontiers in Systems Neuroscience, 7(December), 124. http://doi.org/10.3389/fnsys.2013.00124

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., Bateson, M., ... Wolfe, J. W. (2015). Exploration versus exploitation in space, mind, and society. Trends in Cognitive Sciences, 19(1), 46-54. DOI: 10.1016/j.tics.2014.10.004

245 Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst
246 stimulation of the human motor cortex. Neuron, 45(2), 201–206.
247 http://doi.org/10.1016/j.neuron.2004.12.033
248 Mansouri, F. A., Buckley, M. J., Mahboubi, M., & Tanaka, K. (2015). Behavioral
249 consequences of selective damage to frontal pole and posterior cingulate cortices.
250 Proceedings of the National Academy of Sciences of the United States of America,
251 112(29), E3940–3949. http://doi.org/10.1073/pnas.1422629112
252 Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., …
253 Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of
254 human and animal literatures. Decision, 2(3), 191–215.
255 http://doi.org/10.1037/dec0000033
256 Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. Journal of Neuroscience
257 Methods, 162(1-2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017
258 Pollmann, S. (2016). Frontopolar Resource Allocation in Human and Nonhuman Primates.
259 Trends in Cognitive Sciences, 20(2), 84–86. http://doi.org/10.1016/j.tics.2015.11.006
260 Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2012). Safety, ethical
261 considerations, and application guidelines for the use of transcranial magnetic stimulation
262 in clinical practice and research. Clinical Neurophysiology, 120(12), 323–330.
263 http://doi.org/10.1016/j.clinph.2009.08.016.Rossi
264 Volman, I., Roelofs, K., Koch, S., Verhagen, L., & Toni, I. (2011). Anterior prefrontal cortex
265 inhibition impairs control over social emotional actions. Current Biology, 21(20), 1766–
266 1770. http://doi.org/10.1016/j.cub.2011.08.050
267 Wilson, R. C., Geana, A., White, J. M., Ludvig, E. a, & Cohen, J. D. (2014). Humans Use
268 Directed and Random Exploration to Solve the Explore – Exploit Dilemma. Journal of
269 Experimental Psychology: General, 143(6), 2074–2081. http://doi.org/10.1037/a0038199

Wischnewski, M., & Schutter, D. J. L. G. (2015). Efficacy and time course of theta burst stimulation in healthy humans. Brain Stimulation, 8(4), 685–692. http://doi.org/10.1016/j.brs.2015.03.004

Supplementary Material

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

Wojciech Zajkowski[1], Malgorzata Kossut[2,3], & Robert C. Wilson[4,5]

[1]University of Social Sciences and Humanities, Warsaw, Poland
[2]Departament of Psychology, University of Social Sciences and Humanities, Warsaw, Poland
[3]Nencki Institute, Warsaw, Poland
[4]Department of Psychology, University of Arizona, Tucson AZ USA
[5]Cognitive Science Program, University of Arizona, Tucson AZ USA

Model-based analysis

To complement our analysis in the main paper, we used a model-based analysis based on the model described in Wilson et al. (2014). Briefly, this model assumes that participants make their decision on the first free-choice trial based on the difference in the observed mean reward between left and right bandits, $\Delta\mu$, and the difference in information between left and right bandits, ΔI (= +1 when left is more informative in the [1 3] condition, -1 when the right option is more informative in the [1 3] condition and 0 in the [2 2] condition). In particular, we assume that participants choose the left option with probability p_{left} which is given by

$$p_{left} = \frac{1}{1 + \exp\left(-\frac{\Delta\mu + A\Delta I + B}{\sqrt{2}\sigma}\right)}$$

where A is the information bonus that quantifies the value of information and directed exploration, σ is the standard deviation of the decision noise that quantifies random exploration, and B is the spatial bias that accounts for any baseline tendency to favor left or right choices.

As described in Wilson et al. (2014), these three parameters, A, σ and B, were fit separately for each subject in each horizon and information condition using a maximum *a posteriori* approach. In line with our previous work, we assumed a Gaussian prior with mean zero and standard deviation 20 on A, an exponential prior with length scale 20 for σ and no prior on B.

In line with our model-free analysis in the main text, we found that TMS stimulation of RFPC had a significant effect on directed, but not random, exploration (Figure S1). For directed exploration, a repeated measures ANOVA revealed a significant interaction between horizon and stimulation condition ($F(1,14) = 5.11$ $p = 0.04$). For random exploration in the [2 2] condition there was no such interaction ($F(1, 14) = 0.93$, $p = 0.35$). In addition to measuring decision noise in the [2 2] condition, the model-based analysis also allows us to measure random exploration in the [1 3] condition. Again we found no interaction of horizon and condition ($F(1, 14) = 0.93$, $p = 0.35$) further bolstering our claim that RFPC has no effect on random exploration. Post hoc analysis of the directed exploration result was also consistent with the model-free findings in that we found that the interaction effect was entirely driven by changes in horizon 6 (one-sided t-test $t(14) = 2.21$, $p = 0.022$) not horizon 1 (two-side t-test, $t(14) = -0.50$, $p = 0.63$).

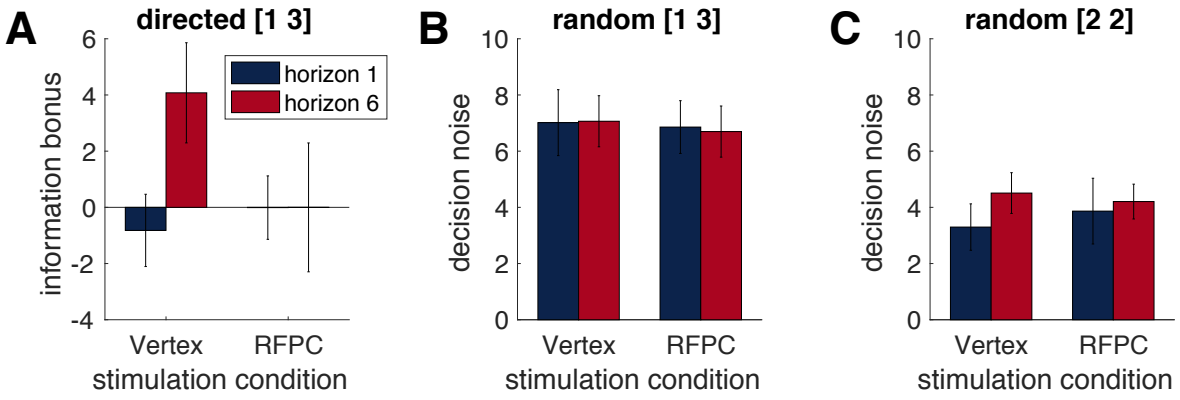

Figure S1 – Mean parameter values for the model-based analysis.

Finally, as another way to visualize the result and to gain a qualitative sense of the quality of the model fit, we computed choice curves for each stimulation condition in the [1 3] condition (Figure S2). These choice curves plot p(high info) as a function of the difference in mean between the high and low information options. In these plots, directed exploration manifests as a left-shift of the horizon 6 curve relative to the horizon 1 curve, as clearly seen in the control condition (Figure S2A). When RFPC is stimulated, this left-shift of the choice curve disappears (Figure S2B) consistent with our other results. These plots also show relatively good agreement between the empirical choice curves computed directly from the data and the fit choice curves computed from the model.

.

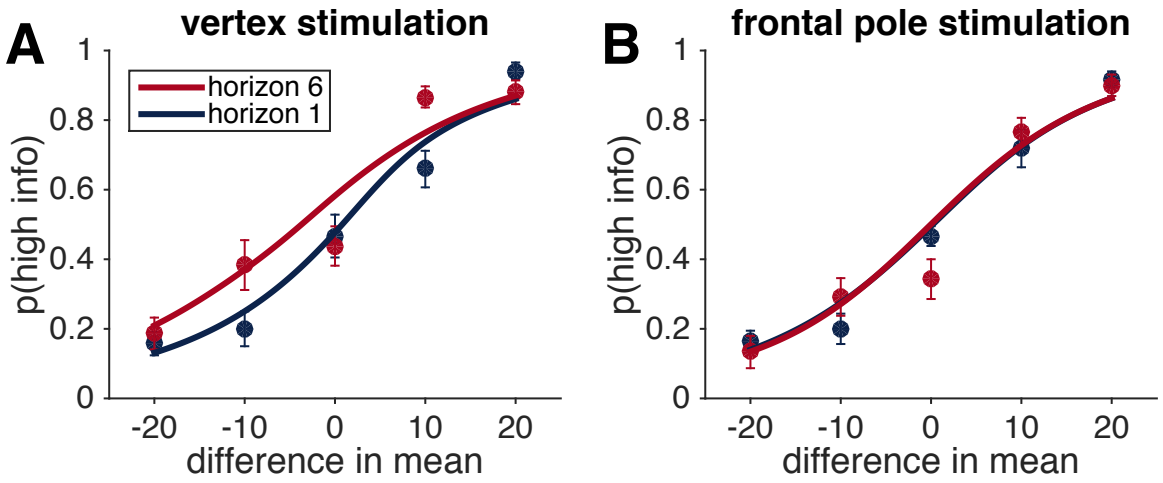

Figure S2 – Choice curves showing p(high info) as a function of the difference in mean between the more and less informative options in the [1 3] condition.