Consciousness and Cognition 62 (2018) 34–41

Contents lists available at ScienceDirect

Consciousness and Cognition

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

On a ‘failed’ attempt to manipulate visual metacognition with transcranial magnetic stimulation to prefrontal cortex

Eugene Ruby[a,*], Brian Maniscalco[b], Megan A.K. Peters[a,c]

[a] Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095-1563, United States
[b] Neuroscience Institute, New York University Langone Medical Center, New York, NY 10016, United States
[c] Department of Bioengineering, University of California Riverside, Riverside, CA 92521, United States

ARTICLE INFO

Keywords:
Consciousness
Visual awareness
Metacognition
Prefrontal cortex
Transcranial magnetic stimulation

ABSTRACT

Rounis, Maniscalco, Rothwell, Passingham, and Lau (2010) reported that stimulation of prefrontal cortex impairs visual metacognition. Bor, Schwartzman, Barrett, and Seth (2017) attempted to replicate this result, but adopted an experimental design that reduced their chance of obtaining positive findings. Despite that, their results appeared initially consistent with those of Rounis et al., but they subsequently claimed it was necessary to discard ~30% of their subjects, after which they reported a null result. Using computer simulations, we found that, contrary to their supposed purpose, excluding subjects by Bor et al.’s criteria does not reduce false positive rates. Including both their positive and negative result in a Bayesian framework, we show the correct interpretation is that PFC stimulation likely impaired visual metacognition, exactly contradicting Bor et al.’s claims. That lesion and inactivation studies demonstrate similar positive effects further suggests that Bor et al.’s reported negative finding isn’t evidence against the role of prefrontal cortex in metacognition.

1. Introduction

Visual metacognition refers to how well one can give subjective judgments to discriminate between correct and incorrect perceptual decisions (Fleming & Lau, 2014). As visual metacognition appears to be closely linked to conscious awareness (Ko & Lau, 2012; Maniscalco & Lau, 2016), it is of interest that the prefrontal cortex has been heavily implicated in mediating both of these faculties (Baird, Smallwood, Gorgolewski, & Margulies, 2013; Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009; Fleming, Ryu, Golfinos, & Blackmon, 2014; Lau & Passingham, 2006; Lumer, Friston, & Rees, 1998; McCurdy et al., 2013; Rounis, Maniscalco, Rothwell, Passingham, & Lau, 2010; Turatto, Sandrini, & Miniussi, 2004). Specifically, one prefrontal area with an empirical link to these abilities is the dorsolateral prefrontal cortex (DLPFC; Lau and Passingham, 2006; Rounis et al., 2010, Turatto et al., 2004).

It has been reported that continuous theta-burst transcranial magnetic stimulation (TMS) to DLPFC can impair visual metacognition (Rounis et al., 2010). Recently, Bor, Schwartzman, Barrett, and Seth (2017) attempted to replicate this finding but reported a null result, which they took to suggest that DLPFC might not be “critical for generating conscious contents” (Bor et al., 2017, p. 16).

However, although the researchers motivated their experiments as direct replications (e.g., Bor et al., 2017, p. 3), they made several changes to the original study design, some of which are known to undermine the chance of finding meaningful results from the outset. In particular, in their main experiment (their Experiment 1) the researchers used a between-subjects design instead of the within-subjects design used by Rounis et al. (2010), which might have limited their statistical power (Greenwald, 1976). Although

---

* Corresponding author at: UCLA Psychology Department, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563, United States.
E-mail address: gene12586@gmail.com (E. Ruby).

https://doi.org/10.1016/j.concog.2018.04.009
Received 20 October 2017; Received in revised form 18 April 2018; Accepted 18 April 2018

they attempted to address this potential issue in a second study (their Experiment 2), we will show below that this study design was unsatisfactory for other reasons.

Importantly, despite these modifications, Bor et al. (2017) in fact found a positive result akin to Rounis et al.'s (2010), with both studies reporting comparable changes in metacognition for subjects who received TMS to DLPFC. However, the researchers proceeded to set stringent exclusion criteria, which they claimed should lower false positive rates (i.e., rate of incorrectly detecting an effect). This caused the removal of a relatively large number of subjects (27 out of 90), and ultimately a null result was found, leading Bor et al. to conclude that the initial significant finding must have been spurious. But their criteria for subject exclusion may have resulted in other important unintended consequences for statistical hypothesis testing and interpretation; therefore, here we formally evaluate the consequences of adopting such criteria in a simulation, and what the interpretation of their results should have been in a Bayesian framework.

2. Methods

Our goal was to assess whether excluding subjects as in Bor et al. (2017) was needed to curb false positive rates as they claim, and also whether doing so led to increased false negative rates and thus lower statistical power (i.e., probability of successfully detecting a true effect). Therefore, we simulated two populations of subjects, one that exhibited the TMS-induced metacognitive impairments, as in Rounis et al. (2010), and one that did not, and included them in two sets of 1000 "experiments" that mirrored Bor et al.'s Experiment 1 (between-subjects design). We then compared statistical power and false positive rates both before and after implementing the exclusion criteria used by Bor et al.

2.1. Building two populations of "subjects"

Each simulated "subject" was characterized by four parameters to produce behavioral outputs with and without TMS. For the first three parameters – objective performance capacity (d'_s), response bias (Type 1 criterion; $c_{s,1}$), and subjective response biases (Type 2 criteria; $c_{s,2}$) – the values were taken from Rounis et al. (2010) to mimic the distributions seen there. These values were then fixed for each simulated subject across task conditions (pre- and post-TMS).

To simulate the effect of degraded metacognitive sensitivity by TMS, we defined a fourth parameter corresponding to Type 2 (i.e. metacognitive) noise ($\sigma_{s,TMS}$), such that in the TMS condition Type 2 criteria ($c_{s,2}$) become unstable and the trial-by trial correspondence between confidence and accuracy is lowered (Maniscalco & Lau, 2012; Maniscalco & Lau, 2016; Peters et al., 2017). Thus, for each simulated subject for these TMS conditions, over trials we added TMS noise ($\sigma_{s,TMS}$) to the subject's internal response, after their discrimination judgment but before their confidence judgment (see task description in Fig. 1), such that across all simulated subjects the average reduction in metacognitive sensitivity mimicked that found in Rounis et al. (2010; see Supplementary Materials for more details).

2.2. Simulating the behavioral task

Our simulated task followed the spatial two-alternative forced-choice (2AFC) task design used by Rounis et al. (2010) and Bor et al. (2017) (see Fig. 1).

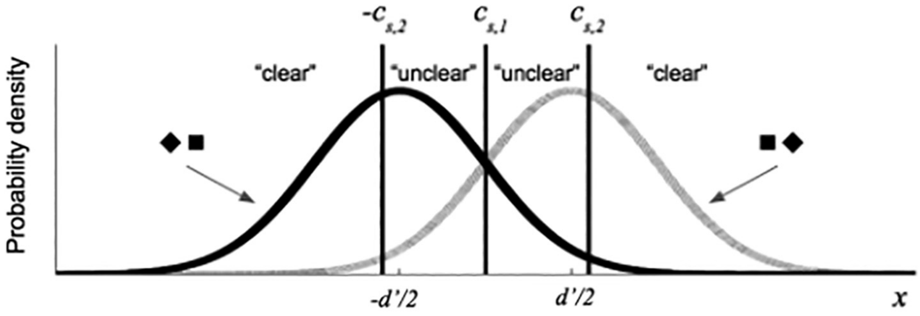

Fig. 1. Signal detection theoretic framework for the simulated spatial 2AFC task. For a given subject s, each stimulus presentation (either □◆ or ◆□) caused an internal response value (x), with $X_{\square\blacklozenge} \sim N(d'/2, 1)$ and $X_{\blacklozenge\square} \sim N(-d'/2, 1)$, and the subject then indicated which of the two shape configurations appeared. If x exceeds the subjects Type 1 criterion ($c_{s,1}$), then the subject responded "□◆"; otherwise the subject responded "◆□". Objective performance capacity (d′) is the normalized distance between the two distributions. The subject then indicated how clearly/confidently they saw the stimulus, based on a comparison between x and the Type 2 criterion ($-c_{s,2}$ and $c_{s,2}$). If $x < c_{s,1}$ or $x > c_{s,2}$ the subject responded "clear"; otherwise the subject responded "unclear." If TMS is present and assumed to degrade metacognitive sensitivity, noise ($\sigma_{s,TMS}$) is added for the Type 2 responses, such that the relevant computations are whether $x < c_{s,1} + \varepsilon_{trial}$ or $x > c_{s,2} + \varepsilon_{trial}$, with $\varepsilon_{trial} \sim N(0,\sigma_{s,TMS})$ and resampled on each trial (see Supplementary Materials for more details).