

Distinguishing the Roles of Dorsolateral and Anterior PFC in Visual Metacognition

Medha Shekhar and Dobromir Rahnev

School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332

Visual metacognition depends on regions within the prefrontal cortex (PFC). Two areas in particular have been implicated repeatedly: the dorsolateral PFC (DLPFC) and the anterior PFC (aPFC). However, it is still unclear what the function of each of these areas is and how they differ from each other. To establish the specific roles of DLPFC and aPFC in metacognition, we used online transcranial magnetic stimulation (TMS) to interfere causally with their functioning during confidence generation. Human subjects from both sexes performed a perceptual decision making and provided confidence ratings. We found a clear dissociation between the two areas: DLPFC TMS lowered confidence ratings, whereas aPFC TMS increased metacognitive ability, but only for the second half of the experimental blocks. These results support a functional architecture in which DLPFC reads out the strength of the sensory evidence and relays it to aPFC, which makes the confidence judgment by potentially incorporating additional, nonperceptual information. Indeed, simulations from a model that incorporates these putative DLPFC and aPFC functions reproduced our behavioral results. These findings establish DLPFC and aPFC as distinct nodes in a metacognitive network and suggest specific contributions from each of these regions to confidence generation.

Key words: confidence; DLPFC; metacognition; perceptual decision making; prefrontal cortex; TMS

Significance Statement

The prefrontal cortex (PFC) is known to be critical for metacognition. Two of its subregions, the dorsolateral PFC (DLPFC) and the anterior PFC (aPFC), have been specifically implicated in confidence generation. However, it is unclear whether these regions have distinct functions related to the underlying metacognitive computation. Using a causal intervention with transcranial magnetic stimulation, we demonstrate that DLPFC and aPFC have dissociable contributions: targeting DLPFC decreased average confidence ratings, whereas targeting aPFC affected metacognitive scores specifically. Based on these results, we postulated specific functions for DLPFC and aPFC in metacognitive computation and corroborated them using a computational model that reproduced our results. Our causal results reveal the existence of a specialized modular organization in PFC for confidence generation.

Introduction

Metacognition, or the ability to assess the quality of our decisions, is crucial for effective decision making (Metcalfe and Shimamura, 1994; Koriat, 2007). However, despite the critical influence of metacognition on our actions and decisions (Nelson and Narens, 1990; Shimamura, 2000a; Koriat, 2007; Fleming et al., 2012a; Yeung and Summerfield, 2012), its neural bases are still not fully elucidated (Shimamura, 2000a; Fleming et al., 2012b).

Early studies pointed to a central role of the prefrontal cortex (PFC) based on findings of impaired metacognition in patients with damage to the frontal lobe (Shimamura and Squire, 1986; Janowsky et al., 1989; Shimamura, 2000a). More recent research has implicated two specific PFC subregions, the dorsolateral PFC (DLPFC) and the anterior PFC (aPFC) (Fleming and Dolan, 2012).

Activity in DLPFC has been linked to the level of reported confidence. Indeed, studies using fMRI have shown consistently that activity within DLPFC tracks confidence levels during metacognitive computations (Henson et al., 2000; Fleck et al., 2006; Lau and Passingham, 2006; Morales et al., 2017). Conversely, aPFC has been linked specifically to subjects' metacognitive ability. For example, structural imaging studies have found that gray matter volume in aPFC correlates with individual metacognitive ability (Fleming et al., 2010; Yokoyama et al., 2010; McCurdy et al., 2013; Allen et al., 2017). Similarly, studies using fMRI show that aPFC activity is modulated by the reliability of confidence judgments (Yokoyama et al., 2010; Fleming et al., 2012b; Morales

Received Dec. 10, 2017; revised April 17, 2018; accepted April 24, 2018.

Author contributions: M.S. wrote the first draft of the paper; D.R. edited the paper. D.R. designed research; M.S. performed research; M.S. analyzed data; M.S. wrote the paper.

This work was funded by a startup grant from the Georgia Institute of Technology (D.R.). We thank Ji Won Bang for help with the experiment setup and Lindley Hudson for assistance with experiment preparation and subject recruitment.

The authors declare no competing financial interests.

Correspondence should be addressed to Medha Shekhar, School of Psychology, Georgia Institute of Technology, 654 Cherry Street NW, Atlanta, GA 30332. E-mail: bsmedha@gmail.com.

DOI:10.1523/JNEUROSCI.3484-17.2018

Copyright © 2018 the authors 0270-6474/18/385078-10\$15.00/0

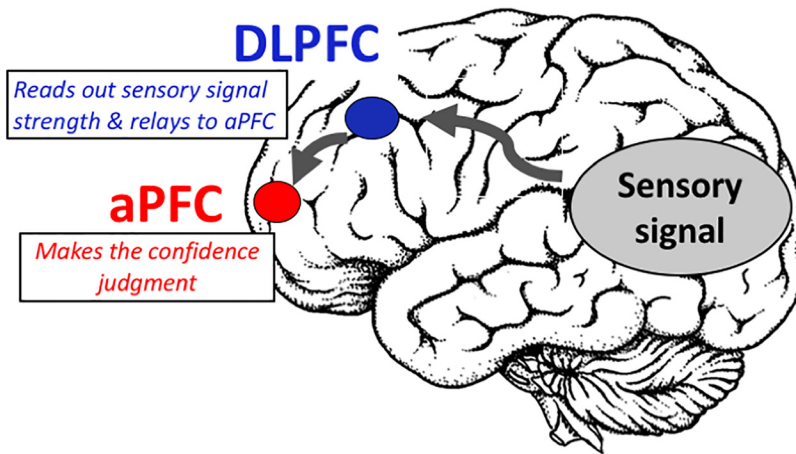


Figure 1. Hypothetical neural mechanism of confidence computation. Based on prior literature, we postulated the following neural mechanism for the roles of DLPFC and aPFC in confidence computation. DLPFC reads out the strength of the sensory signal and relays it to aPFC. Conversely, aPFC translates this readout into a confidence judgment after incorporating additional, nonperceptual factors. The strength of the sensory signal that is read out by DLPFC on a particular trial is related to the level of confidence on that trial. On average, disrupting the readout would convey that less evidence was available for the perceptual decision compared with the evidence that was actually available. Such disrupted readout would be translated by aPFC into a lower confidence rating. Therefore, impaired DLPFC functioning leading to poor-quality readouts would convey that the sensory information was more ambiguous than it really is and would result in lower confidence ratings. In contrast, impaired aPFC functioning would alter how aPFC transforms the sensory readout relayed from DLPFC into a confidence judgment and therefore alter metacognitive performance.

et al., 2017). Finally, metacognitive scores are affected by both lesions (Fleming et al., 2014) and transcranial magnetic stimulation (TMS) (Rahnev et al., 2016; Ryals et al., 2016) to aPFC.

Based on the findings above, we hypothesized specific functions for DLPFC and aPFC in confidence computation. We propose that DLPFC reads out the strength of the sensory signal and relays it to aPFC. The readout of the sensory signal determined by DLPFC conveys how much information was available for the sensory decision. aPFC subsequently integrates this readout with additional, nonperceptual factors and translates all of this information into a confidence judgment (Fig. 1). Disrupted readout of the sensory signal by DLPFC would, on average, convey that less information was available for the sensory decision; in turn, aPFC would translate such disrupted readout into lower confidence ratings. This architecture is consistent with prior findings because reading out the sensory signal strength would link DLPFC activity with confidence level, whereas making the confidence judgment would link aPFC activity with metacognitive ability. In addition, Fleming et al. (2012b) observed that connectivity between aPFC and DLPFC increased during metacognitive reports, suggesting active communication between the two regions during confidence computation.

We tested our hypothesis regarding the putative functions of DLPFC and aPFC in confidence computation by using online TMS. Previous TMS studies on visual metacognition (Rounis et al., 2010; Rahnev et al., 2016; Ryals et al., 2016) used offline approaches that inhibit activity for an extended period of time. These studies showed little or no modulation of overall confidence level presumably because subjects had time to recalibrate their confidence judgments. To address this issue, we applied online TMS in short blocks to avoid behavioral compensation.

Based on our hypothesis about the functions of DLPFC and aPFC, we predicted that TMS to DLPFC would affect subjects' overall confidence level, whereas TMS to aPFC would affect metacognitive ability. The results confirmed these predictions: TMS to DLPFC decreased confidence, whereas TMS to aPFC

increased metacognitive ability, but only for the second half of the blocks. Further, we confirmed that these results could be reproduced by a model in which TMS to DLPFC affected the readout of the sensory information, whereas aPFC TMS affected the noise within the metacognitive computation itself. Our findings demonstrate that DLPFC and aPFC have distinct functions in visual metacognition and suggest a specific mechanistic role for each.

Materials and Methods

Subjects. A total of 21 subjects completed the study (13 females and 8 males, average age 22 years, range 18–32 years). Three subjects were excluded from analyses. For one subject, the sensors registering the subject's brain to their MRI shifted midsession, which likely resulted in imprecise TMS target localization. The other two subjects were excluded due to poor performance or excessive number of interruptions due to discomfort. All subjects were right handed and had normal or corrected-to-normal vision.

Session sequence. We collected data for our experiment over two sessions, which were held on separate days. By dividing data collection into 2 d, we were able to collect more data while

keeping the session short enough to avoid fatigue.

Day 1 started with a short training on the behavioral task, followed by a staircasing procedure used to identify the contrast of the stimulus to be used for the main experiment. After subjects completed the staircasing procedure, we determined the amplitude of TMS stimulation to use and started the main experiment.

The main experiment consisted of four runs of three blocks each. For each of the three blocks within each run, we stimulated one of three regions, DLPFC, aPFC, or the somatosensory cortex (S10, which served as the control site), in a pseudorandom order such that all the three sites were stimulated once within each run. The first run was a practice run and was shorter than the others. It was included to accustom subjects to receiving TMS to the different brain regions and to minimize the chances of the TMS pulse evoking a startle response during the main trials. The blocks from the practice run consisted of 16 trials each and were excluded from further analyses. All other blocks consisted of 40 trials each. Therefore, subjects completed a total of 408 trials during each session.

During day 1, subjects underwent a behavioral training procedure without TMS. The training session started with high stimulus contrast values (40%) and gradually presented lower contrast values (the last block included contrast values of 4%). Subjects were given trial-by-trial feedback on their performance during this training period.

At the end of the training, subjects completed a 3-down-1-up staircasing procedure consisting of trials without feedback. The 3-down-1-up procedure is a variant of the up-down transformed response method used for adaptive estimation of stimulus thresholds (Macmillan and Creelman, 2005). This procedure yielded a contrast value for the stimulus (mean = 6.64% and SD = 0.96%) that was expected to result in an accuracy of 79% (Macmillan and Creelman, 2005; observed mean = 79.6% and SD = 5.8%). We used the contrast value obtained from this procedure for the rest of the experiment.

Day 2 was identical to day 1 except that subjects did not have to undergo behavioral training or staircasing (we used the same stimulus contrast as in day 1).

Task. Each trial began with subjects fixating on a small white dot (size = 0.05°) at the center of the screen for 500 ms, followed by presentation of the stimulus for 100 ms. The stimulus was a Gabor patch (diameter = 3°) oriented either to the right (clockwise, 45°) or to the left (counterclockwise, 135°) of vertical and was superimposed on a noisy

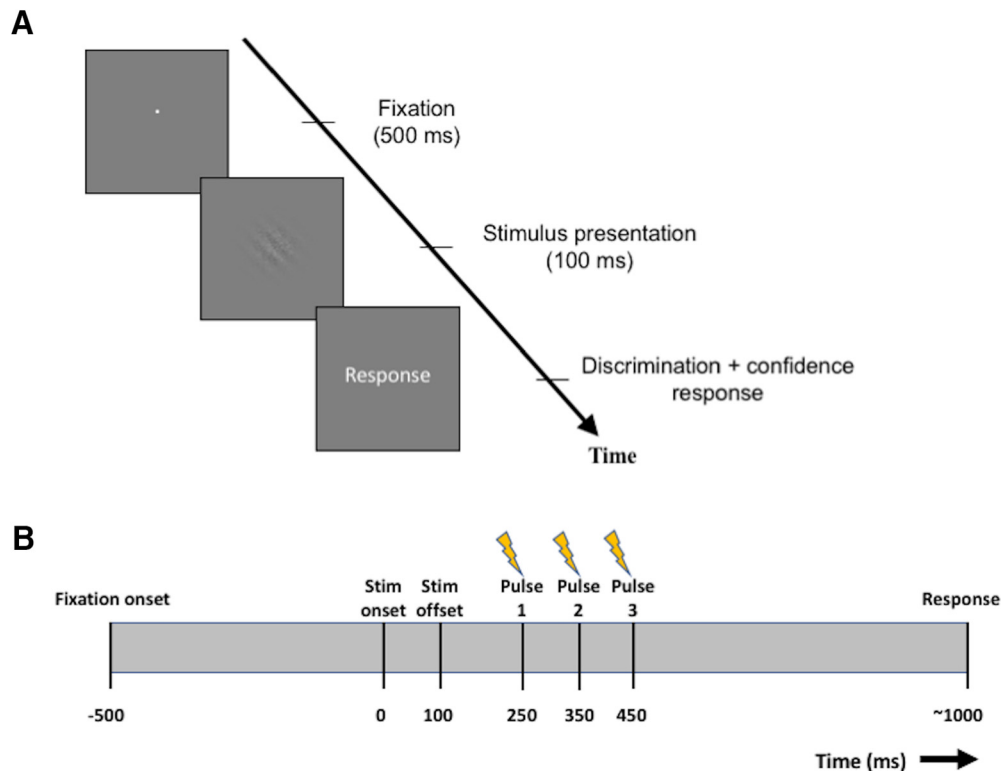


Figure 2. Task. **A**, Trial sequence. Each trial began with short fixation (500 ms), followed by the presentation of an oriented Gabor patch (100 ms). Subjects had to simultaneously indicate the tilt (left/right) of the Gabor patch and their confidence on a 1–4 scale. **B**, Timeline of TMS delivery. TMS was given as a train of three pulses with interpulse interval of 100 ms. The first pulse was delivered 250 ms after onset of the stimulus. Subjects had a mean response time of ~ 1000 ms.

background. Subjects' task was to determine the orientation of the Gabor patch while simultaneously rating their confidence on a 4-point scale (where 1 corresponds to the lowest confidence rating and 4 corresponds to the highest confidence rating) via a single key press (Fig. 2A). Subjects placed their fingers of each hand on a standard keyboard. The four fingers of the left hand were mapped to the four confidence responses for the left-tilted stimulus, whereas the four fingers of the right hand were mapped to the four confidence responses for the right-tilted stimulus. For each hand, the index finger indicated a confidence of 1, whereas the little (pinkie) finger indicated a confidence of 4. The orientation of the stimulus (left/right) was chosen randomly on each trial.

We delivered TMS on each trial as a train of three pulses delivered 250, 350, and 450 ms after stimulus onset (Fig. 2B). We chose this timing so that it coincided with the presumed time window of confidence computation. Although there are no clear data on the precise time window when confidence is computed, neuronal recordings from monkeys suggest that the discrimination response emerges ~ 200 ms after stimulus onset (Siegel et al., 2015), placing confidence computation in human PFC no earlier than 200 ms. To estimate the length of the time window, we collected pilot data from an identical discrimination task in which subjects made two responses: their first response indicated the tilt (left/right) of the Gabor patch and their second response indicated their confidence level on a 4-point scale. Analysis of these data showed that subjects typically take ~ 500 ms to give their confidence response after the discrimination response. After approximately accounting for motor preparation (~ 200 ms), we estimated that the actual duration of confidence computation is ~ 300 ms. Based on this estimation, we timed our TMS pulses so that they targeted a time window that started 250 ms after stimulus onset and spanned the next 200 ms.

Apparatus. Stimuli were generated using Psychophysics Toolbox (RRID: SCR_002881) in MATLAB (The MathWorks, RRID: SCR_001622). During the training and the main experiment, subjects were seated in a dim room and were positioned 60 cm away from the computer screen (21.5 inch display, 1920 \times 1080 pixel resolution, 60 Hz refresh rate).

Defining ROIs for TMS targeting. We defined three sites as targets for TMS: DLPFC, aPFC, and S1 (control site). Based on previous studies (Fleming et al., 2010, 2012b; Yokoyama et al., 2010; Rahnev et al., 2016), aPFC was localized at [28, 56, 26]. We localized DLPFC immediately posterior to aPFC (at a distance of 2.6 cm posterior to aPFC) and used [28, 30, 38] as the target coordinates. For S1, we used [20, -39, 70] as the putative coordinates (Rahnev et al., 2016), but the actual location of stimulation was adjusted based on S1's known anatomical location in the postcentral gyrus. As in previous work (Rahnev et al., 2016), all regions were defined in the right hemisphere because the right hemisphere is dominant for visual processing (Hellige, 1996).

We defined the ROIs on the anatomical MRI scans of each subject. These scans were obtained during previous studies conducted in the laboratory. To determine the subject-specific location for stimulation, we back-normalized the coordinates above to the subject's native space. We created ROIs as 5 mm spheres and their centers were set as targets to guide the placement of the TMS coil. In some cases, the ROIs produced via back-normalization appeared shifted with respect to the expected anatomical location. In such cases, we switched to an alternate method of defining ROI locations. The neural navigator software TMSNavigator (Localite) contains a built-in program for defining a Talairach coordinate system on a subject's MRI that is based on the location of the anterior commissure, the posterior commissure, and the vertex. After these structures are identified manually on an MRI scan, the software generates a Talairach grid, which can be adjusted so that it encloses the whole brain. This grid allows transformation of coordinates between the subject's native coordinate space and the MNI coordinate space.

TMS setup. TMS was delivered with a magnetic stimulator (MagPro R100; MagVenture, RRID: SCR_009601) using a figure-eight coil with a diameter of 75 mm.

We determined the resting motor threshold (RMT) immediately before starting the main experiment. To localize the motor cortex, we marked its putative location and applied suprathreshold single pulses around that location. We determined the location of the right motor

cortex as the region that induced maximal twitches of the fingers in the left hand. Then, using this location as the target, we determined the RMT using an adaptive parameter estimation by sequential testing procedure (Borckardt et al., 2006). For three subjects, we were unable to estimate RMT reliably, even at amplitudes as high as 80. Therefore, for these subjects we chose to determine the active motor threshold instead, which is lower than RMT and could be found reliably. Motor thresholding was done separately for both days (average for day 1 = 59.94, average for day 2 = 58.28), to control for nonspecific factors that can influence the TMS response (Ridding and Ziemann, 2010).

The TMS coil was positioned on the previously defined ROIs using a neural navigation system (TMS Navigator, RRID:SCR_016126). The coil was oriented tangential to the skull and in such a way that the magnetic field induced was orthogonal to the skull. Stimulation was delivered at 90% of the individual RMT. In some cases when the stimulation intensity was uncomfortable to the subject, it was reduced to ~85% (two subjects) or ~80% (three subjects) of RMT depending on the individual's comfort level. No arm or leg movements were elicited by stimulation of any of the three sites.

Analyses. We analyzed the data for two separate measures: average confidence and metacognitive ability. To compute the average confidence, we simply calculated the average of all confidence ratings within each TMS condition. We quantified metacognitive ability using the measure M_{ratio} developed by Maniscalco and Lau (2012). M_{ratio} is derived from signal detection theoretical modeling of the observer's decision and confidence responses. It is the ratio of two measures: the observer's metacognitive sensitivity ($meta-d'$, the ability to discriminate between correct and incorrect responses) and the observer's stimulus sensitivity (d' , the ability to discriminate between the two stimulus classes). The ratio of $meta-d'$ to d' factors out the contribution of stimulus sensitivity toward metacognitive performance and captures the efficiency of the observer's metacognitive processes (Fleming and Lau, 2014).

We compared the effect of TMS on confidence and metacognitive ability between the three TMS conditions (DLPFC, aPFC, and S1) using one-way repeated-measures ANOVAs. In addition, we analyzed the interaction between time (within a block) and TMS location by splitting each block into first (trials 1–20) and second (trials 21–40) halves and performing a two-way repeated-measures ANOVA. Direct comparisons between regions were made using paired t tests.

Splitting blocks in halves and analyzing each half separately may decrease the stability of the M_{ratio} estimates. To confirm that our M_{ratio} estimates were not unacceptably variable, we tested whether splitting blocks in half had a significant influence on the variance of M_{ratio} scores. We verified that all groups of M_{ratio} values (coming from first half, second half, and the whole block) were distributed normally and used the F test of equality of variance to test whether the two distributions came from populations with different variances. First, we compared the population variance of M_{ratio} scores between the first and the second halves (after pooling M_{ratio} scores obtained from all three TMS conditions). The F test showed that the between-subject variance of M_{ratio} was not significantly different between the two halves of the blocks ($F = 0.72$, $p = 0.23$). Next, we pooled the M_{ratio} scores from both halves and compared their variance against M_{ratio} scores obtained from combining trials from both the halves. The F test revealed no significant difference between the variance of these two populations too ($F = 0.71$, $p = 0.16$). In addition, we confirmed that the number of zero-cell counts in the accuracy/confidence matrix (i.e., the number of confidence–accuracy combinations, such as incorrect trials with confidence of 4 that never appeared) were similar between the two halves for all three TMS conditions. Whenever a zero-cell count did occur within any condition, the codes for computing $meta-d'$ provided by Maniscalco and Lau (2012) automatically applied a default correction for all the data cells in that condition (Maniscalco and Lau, 2012). This default correction method was thus the same across all the conditions.

General model architecture. Our results showed that TMS to each prefrontal site affected one specific aspect of confidence ratings, either their average value or their reliability in predicting accuracy. Our neural mechanism implies that the change in average confidence was due to TMS affecting the readout of the sensory signal and the change in metacognitive

ability was caused by TMS affecting the efficiency of the metacognitive evaluation.

To assess our proposed neural mechanism, we performed simulations of a process model of confidence generation that incorporated our hypothesized TMS effects. It should be noted that we could not use previous approaches such as the existing procedure for estimating metacognitive sensitivity ($meta-d'$), which is built on a signal detection theoretical (SDT) framework for modeling perceptual decisions (Maniscalco and Lau, 2012). The reason is that, although this procedure allows for the estimation of metacognitive performance, it does not specify how the confidence data actually come about on a single trial (i.e., it is not a process model of confidence). For the purpose of modeling the process of confidence generation, we sought to build a process model that preserves the assumptions of SDT (Green and Swets, 1966), at the level of the perceptual decisions, but also allowed us to model explicitly the transformations to the sensory signal that are responsible for generating the confidence ratings. The simplest way to model the transformation of the sensory signal at the metacognitive level is to postulate the existence of metacognitive noise that corrupts the sensory signal, as was done previously by the creators of the $meta-d'$ measure (Maniscalco et al., 2014), us (Rahnev et al., 2016; Bang et al., 2017), and others (Mueller and Weidemann, 2008; Jang et al., 2012; De Martino et al., 2013; van den Berg et al., 2017).

Our model assumes that perceptual decisions and confidence ratings are the result of a hierarchical process consisting of two levels: an object level, which generates the discrimination response, and a meta level, which generates the confidence response. At the object level, the presented stimulus produces a sensory response corrupted by Gaussian noise. We modeled the two Gaussian distributions arising from the two stimulus classes (left/right tilted Gabor patches) such that the left-tilted stimuli produce a sensory response $r_{\text{sens}} = N\left(-\frac{\mu}{2}, \sigma_{\text{sens}}^2\right)$ and the right-tilted stimuli produce a sensory response $r_{\text{sens}} = N\left(-\frac{\mu}{2}, \sigma_{\text{sens}}^2\right)$. Note that the distance between these distributions is μ and the stimulus sensitivity can

be expressed as follows: $d' = \frac{\mu}{\sigma_{\text{sens}}}$. A copy of this sensory response, r_{sens} , gets transferred to the meta level as a readout of the sensory signal strength, r_{readout} , where it is further corrupted by metacognitive Gaussian noise such that the metacognitive response is given by the formula $r_{\text{meta}} = N(r_{\text{readout}}, \sigma_{\text{meta}}^2)$.

To simulate how subjects make perceptual and confidence responses on each trial, we specified a decision criterion, c_0 , and confidence criteria, $c_{-n}, c_{-n+1}, \dots, c_{-1}, c_1, \dots, c_{n-1}, c_n$, where n is number of ratings on the confidence scale (in our case, $n = 4$). The criteria c_i were monotonically increasing with $c_{-n} = -\infty$ and $c_n = \infty$.

The object-level decisions were made based on a comparison of r_{sens} with c_0 . For trials in which $r_{\text{sens}} > c_0$, the response was given as “right”; otherwise, the response was given as “left.” Confidence responses were based on r_{meta} such that values falling within the interval $[c_i, c_{i+1}]$ resulted in a confidence of $i + 1$, when $i \geq 0$ and a confidence of $-i$ when $i \leq -1$, where i takes integer values ranging from -4 to 3 . In cases in which r_{sens} and r_{meta} fell on different sides of the decision criterion c_0 , we constrained the confidence response to equal 1.

Finally, our data showed the existence of small (and nonsignificant) decrease in M_{ratio} for the second half of blocks in the S1 and DLPFC TMS conditions. This effect parallels recent findings that metacognitive ability may decrease in second half of blocks due to fatigue (Maniscalco et al., 2017). To model this effect, we allowed σ_{meta} to increase in the second half of all blocks by a value controlled by the parameter $\Delta\sigma_{\text{meta_base}}$.

Our computational model can be related to our hypothesized neural mechanism about the roles of DLPFC and aPFC in confidence computation. According to the neural mechanism that we proposed, the sensory signal strength is read out by DLPFC. Here, we model r_{sens} as the sensory signal produced at the object level. Under normal conditions (no TMS), the readout of this sensory signal by DLPFC, r_{readout} , will equal r_{sens} and will be relayed to aPFC for the confidence judgment. Further, our neural mechanism postulates that the role of aPFC is to integrate the strength of

the sensory signal relayed by DLPFC with non-perceptual cues and make the confidence judgment. Within our model, this process can be seen as the addition of metacognitive noise σ_{meta} at the meta level.

Modeling the TMS effects. According to our proposed neural mechanism, TMS to DLPFC should influence the magnitude of sensory readout that can be used at the meta level. Our data showed that confidence level decreases after DLPFC TMS, suggesting that the sensory readout decreases in magnitude. We formalized this idea in our computational model as DLPFC TMS leading to a decrease in the magnitude of the sensory readout such that the following is true:

$$r_{readout} = \begin{cases} r_{sens} + \Delta r_{sens}, & \text{if } r_{sens} < 0 \\ r_{sens} - \Delta r_{sens}, & \text{if } r_{sens} \geq 0 \end{cases}$$

where Δr_{sens} controls the change in the readout. These conditions satisfy the relation $|r_{readout}| = |r_{sens}| - \Delta r_{sens}$ such that the effect of TMS is to reduce the absolute magnitude of $r_{readout}$ without changing its sign. (As stated above, in cases in which r_{sens} and r_{meta} had a different sign, which occurs when $|r_{sens}| < \Delta r_{sens}$, we constrained the confidence response to equal 1 by setting $r_{readout} = 0$.) Conversely, according to our proposed neural mechanism, TMS to aPFC should affect the level of noise that corrupts the confidence decision. We formalized this idea in our model as aPFC TMS leading to an altered level of metacognitive noise. Because our behavioral results suggested that aPFC TMS increased metacognitive scores only in the second half of blocks, we modeled the effect of aPFC TMS as a decrease in metacognitive noise for the second half of blocks such that $r_{meta} = N(r_{readout}(\sigma_{meta} - \Delta\sigma_{meta})^2)$, where $\Delta\sigma_{meta}$ controls the change of the metacognitive noise.

To simulate actual data, we set the basic parameters of the model such that $\mu = 1.74$, $\sigma_{sens} = 1$, $\sigma_{meta} = 0.6$, $\sigma_{meta_base} = 0.15$, $c_{-3} = -1.45$, $c_{-2} = -0.95$, $c_{-1} = -0.45$, $c_0 = 0$, $c_1 = 0.45$, $c_2 = 0.95$, and $c_3 = 1.45$. We set $\sigma_{sens} = 1$ because choosing other values would simply lead to a multiplicative change in all other parameters. The value of μ was chosen based on the average d' observed across all subjects. The values for the rest of the parameters were chosen to match the overall performance that we observed in the study. However, the effects of TMS do not depend on the specific numbers and the same qualitative results were observed with a wide range of values of the different parameters. Critically, we used different values of Δr_{sens} and $\Delta\sigma_{meta}$ for modeling the different TMS conditions. For S1 TMS, we set $\Delta r_{sens} = 0$ and $\Delta\sigma_{meta} = 0$. For modeling DLPFC TMS, we set $\Delta r_{sens} = 0.072$ and $\Delta\sigma_{meta} = 0$, consistent with the notion that DLPFC TMS should lead to a decrease in the magnitude of the sensory readout for metacognition. Finally, for modeling aPFC TMS, we set $\Delta r_{sens} = 0$ and $\Delta\sigma_{meta} = 0.65$, consistent with the notion that aPFC TMS should change the metacognitive noise.

Data and code. All data analysis and simulation files can be downloaded from https://github.com/Medha66/onlineTMS_DLPFC_aPFC.

Results

We investigated the specific contributions of DLPFC and aPFC to visual metacognition by using an online TMS protocol to disrupt activity within these areas during confidence computation. Subjects indicated the tilt (left/right) of a noisy Gabor patch while simultaneously providing a confidence rating on a four-point scale. On each trial, we delivered a train of three TMS pulses to DLPFC, aPFC, or S1 (which served as a control site).

As in previous studies on the role of PFC in perceptual decision making (Rounis et al., 2010; Rahnev et al., 2016; Ryals et al., 2016), TMS did not influence the overall task performance as

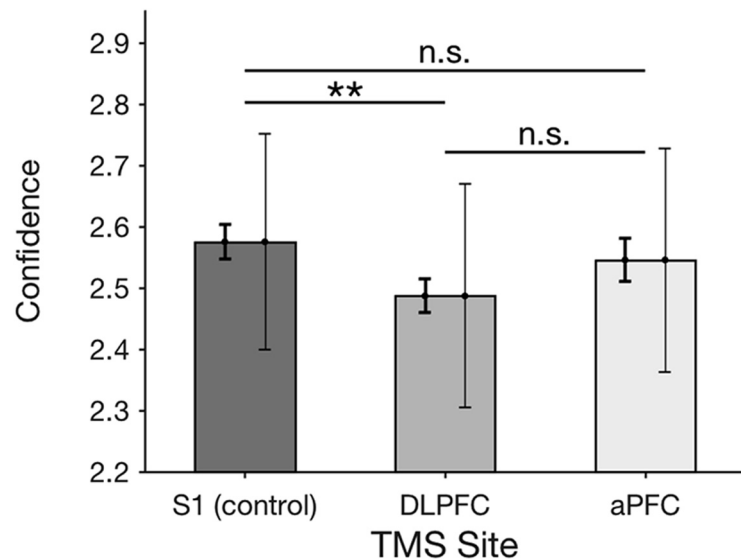


Figure 3. TMS effect on overall confidence level. TMS to DLPFC decreased average confidence, whereas TMS to aPFC did not affect the overall confidence level. The left error bars represent the within-subject SEs for comparisons with S1 (the error bar for S1 is the same as the one for DLPFC) and are indicative of statistical significance. The right error bars represent the between-subject SEs and are not indicative of the statistical significance (instead, they show the overall variability in confidence across subjects). n.s., Not significant. $**p < 0.01$.

measured by accuracy or reaction time ($p > 0.05$ for all pairwise comparisons between the three sites). These results suggest that the PFC is unlikely to be involved in low-level stimulus processing (Rahnev, 2017).

TMS effect on confidence

Based on our hypothesis regarding the functions of DLPFC and aPFC in confidence generation, we predicted that DLPFC TMS, but not aPFC TMS, would affect subjects' overall confidence level. The results were consistent with this prediction. Indeed, a one-way repeated-measures ANOVA with factor TMS site (S1, DLPFC, and aPFC) demonstrated a significant effect of TMS location on confidence ($F_{(2,17)} = 3.68$, $p = 0.04$; Fig. 3). Pairwise comparisons showed a significant decrease in confidence for DLPFC TMS compared with S1 TMS (difference = 0.09, $t_{(17)} = 3.19$, $p = 0.005$). No significant difference was found for comparisons between S1 TMS and aPFC TMS (difference = 0.03, $t_{(17)} = 0.83$, $p = 0.4$), implying that overall confidence level was affected only after DLPFC stimulation. The difference in confidence between DLPFC TMS and aPFC TMS was numerically larger than the difference between S1 TMS and aPFC TMS, but did not reach significance (difference = 0.06, $t_{(17)} = 1.7$, $p = 0.12$).

TMS effect on metacognitive ability

Based on our hypothesis regarding the functions of DLPFC and aPFC in confidence generation, we predicted that aPFC TMS, but not DLPFC TMS, would affect subjects' metacognitive ability. To test this prediction, we used M_{ratio} as a measure of the quality of metacognition (Maniscalco and Lau, 2012). However, a one-way repeated-measures ANOVA with factor TMS site (S1, DLPFC, and aPFC) on M_{ratio} scores showed no main effect of TMS location on metacognitive ability ($F_{(2,17)} = 0.3$, $p = 0.74$).

In contrast to these results, previous studies showed that offline TMS to aPFC increased metacognitive scores (Rahnev et al., 2016; Ryals et al., 2016). Therefore, it is possible that the effects of TMS to aPFC become apparent only after a more sustained period of inhibition. To test this possibility, we investigated whether

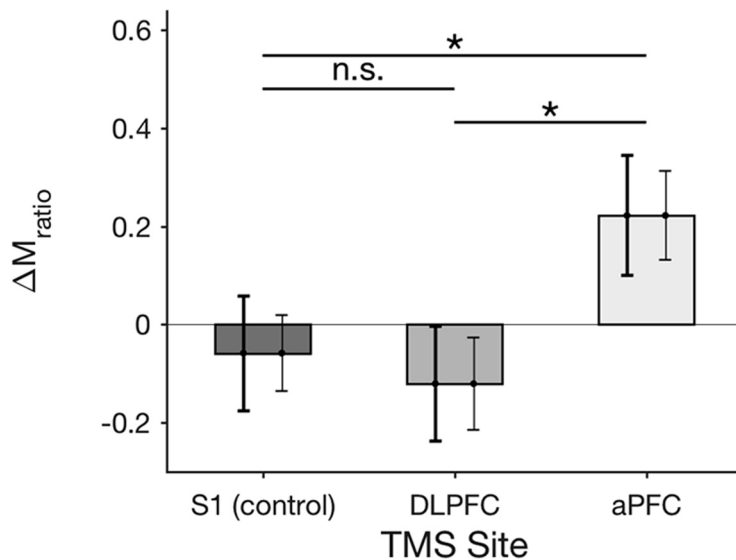


Figure 4. TMS effect on metacognitive ability. TMS to aPFC increased metacognitive ability for the second half compared with the first half of test blocks. No such effect was observed for S1 TMS or DLPFC TMS. Metacognitive ability was operationalized as M_{ratio} (Maniscalco and Lau, 2012). ΔM_{ratio} is the change in M_{ratio} from the first half to the second half of a block. The left error bars represent the within-subject SEs for comparisons with S1 (the error bar for S1 is the same as the one for aPFC) and are indicative of statistical significance. The right error bars represent the within-subject SEs for comparisons between the first half and second half of blocks and are not indicative of the statistical significance for between-site comparisons. n.s., Not significant. * $p < 0.05$.

metacognitive ability differed between the first and second halves of test blocks. We performed a 2 (time: first vs second half of test blocks) \times 3 (TMS site: S1, DLPFC, and aPFC) repeated-measures ANOVA on M_{ratio} and found a significant interaction between time and TMS site ($F_{(2,1)} = 3.9, p = 0.03$; Fig. 4). Further analyses revealed a significant increase in M_{ratio} for the second half (compared with the first half) of test blocks after aPFC TMS (difference = 0.22, $t_{(17)} = -2.44, p = 0.03$) but not after S1 TMS (difference = $-0.06, t_{(17)} = 0.75, p = 0.47$) or DLPFC TMS (difference = $-0.12, t_{(17)} = 1.27, p = 0.22$). Critically, the difference in M_{ratio} between the two halves of test blocks was significantly larger for aPFC TMS compared with both S1 TMS (difference = 0.28, $t_{(17)} = 2.4, p = 0.028$) and DLPFC TMS (difference = 0.34, $t_{(17)} = 2.81, p = 0.012$). Therefore, TMS increased metacognitive ability for the second half of our blocks and this effect was specific to aPFC.

We further examined the interaction between time and TMS location for d' and $meta-d'$ separately. We performed a two-way repeated-measures ANOVA on d' with time and TMS location as factors. The results indicated no overall main effect of time on d' ($F_{(1,17)} = 3.04, p = 0.1$) and no significant interaction between time and TMS location ($F_{(2,1)} = 1.56, p = 0.22$). The data suggested a potential d' decrease in the second half of blocks after aPFC TMS, but this decrease was not significantly different from S1 TMS (difference = 0.22, $t_{(17)} = 1.47, p = 0.16$). Conversely, although the interaction between time and TMS location did not reach significance for $meta-d'$ ($F_{(2,1)} = 2.28, p = 0.12$), a paired t test showed that $meta-d'$ increased significantly for the second half of blocks for aPFC TMS compared with S1 TMS (difference = 0.36, $t_{(17)} = 2.27, p = 0.037$). We note that the weaker effect of aPFC TMS on $meta-d'$ compared with M_{ratio} likely resulted from the small (and nonsignificant) decrease of d' in the second half of blocks after aPFC TMS. Due to their positive correlation, a d' decrease should have led to a corresponding decrease in $meta-d'$. However, despite this d' decrease, aPFC TMS still caused a significant (compared with S1 TMS) increase in

$meta-d'$ for the second half of blocks. In other words, the lack of significant interaction between time and TMS location for $meta-d'$ is likely due to the fact that the $meta-d'$ increase for the second half of blocks after aPFC TMS was masked by its lower baseline. We further note that the M_{ratio} analyses above automatically correct for fluctuations in d' values and are thus more diagnostic of putative changes in metacognitive ability.

Finally, we also examined the interaction between time and TMS location for confidence. A two-way ANOVA revealed a main effect of time on confidence ($F_{(1,17)} = 14.8, p = 0.001$) driven by a confidence decrease across all three TMS sites. Critically, the interaction between time and TMS location was nonsignificant ($F_{(2,1)} = 0.42, p = 0.66$). A direct paired t test confirmed that the decrease in confidence from first to second half of blocks was not significantly different between aPFC and the control site S1 ($t_{(17)} = 0.49, p = 0.63$). Therefore, TMS location did not have a significant effect on confidence.

Simulating the effects of TMS with a hierarchical confidence generation model

The results above confirmed our prediction that disrupting DLPFC would affect average confidence, whereas disrupting aPFC would affect metacognitive ability. This prediction was based on the hypothesis that DLPFC reads out the strength of the sensory signal and relays it to aPFC, which translates it into a confidence judgment by also incorporating nonperceptual factors. To test whether these mechanistic effects can indeed reproduce our results, we implemented them in simulations of a computational model of confidence generation.

The model that we developed is based on the common assumption of the existence of independent sensory and metacognitive noise (Mueller and Weidemann, 2008; De Martino et al., 2013; Rahnev et al., 2016; Bang et al., 2017; van den Berg et al., 2017). The two noise stages lead to separate representations for object-level and meta-level judgments (Fig. 5A). At the object level, the stimulus is corrupted by sensory noise and the resulting signal is used to make a perceptual decision. To make the confidence judgment, the signal strength from the object level is read out at the meta level. The final confidence decision is based on the sensory readout, as well as other factors such as the history of confidence responses (Rahnev et al., 2015), perceived attentional state, etc. We modeled all of these influences collectively as the addition of metacognitive noise.

Within this architecture, our proposed effects of inhibiting DLPFC and aPFC can be operationalized as DLPFC TMS affecting the strength of the sensory readout and aPFC TMS affecting the level of metacognitive noise (Fig. 5A, boxed equations). Quantitatively, we modeled the effect of TMS to DLPFC as a loss of the sensory readout at the meta level and the effect of TMS on aPFC as a decrease in metacognitive noise (see Materials and Methods).

Simulations of our computational model faithfully reproduced the TMS effects for both overall confidence level (Fig. 5B) and metacognitive ability (Fig. 5C). Therefore, within this estab-

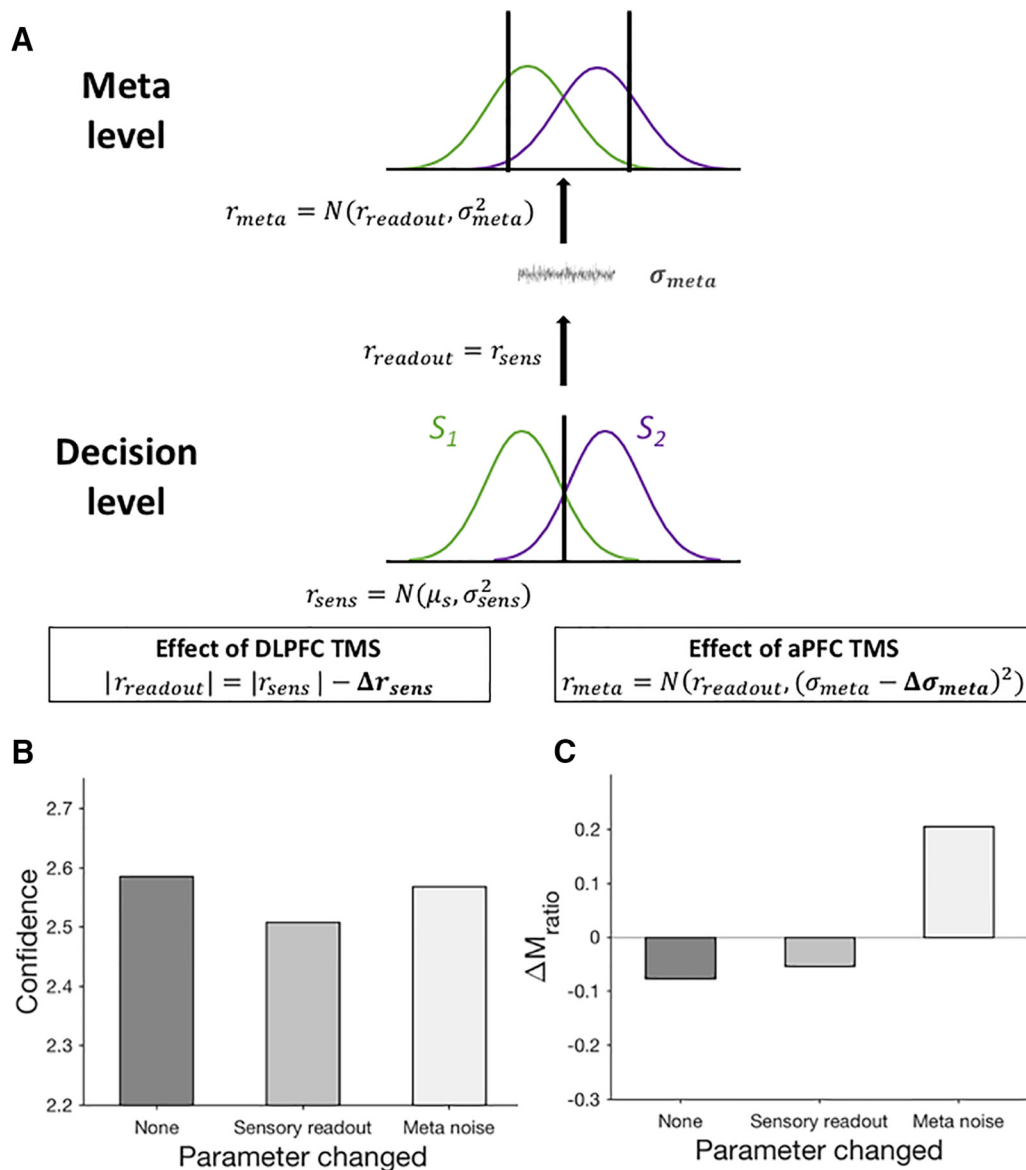


Figure 5. A computational model of confidence generation. **A**, The sensory signal (r_{sens}) available at the decision level is read out ($r_{readout}$) to the metacognitive level and additional noise (σ_{meta}) is added before obtaining the confidence signal (r_{meta}). The perceptual decision is based on the sensory signal r_{sens} , whereas the confidence judgment is based on the confidence signal r_{meta} . Consistent with the hypothesized roles of DLPFC and aPFC in confidence computation, we modeled the effect of DLPFC TMS as a signal loss from the readout (quantified as Δr_{sens} ; boxed equation on the left) and the effect of aPFC TMS as lowered metacognitive noise (quantified as $\Delta\sigma_{meta}$; boxed equation on the right). **B**, **C**, Model simulations show that decreasing the magnitude of the readout decreases the overall confidence level (**B**), but does not influence metacognitive ability (**C**). Conversely, decreasing the amount of metacognitive noise in the second half of test blocks has a small effect on average confidence (**B**), but a large effect on increasing the difference in metacognitive ability between the first and second half of blocks (**C**). These results mirror the effects of TMS to DLPFC and aPFC in our data (see Figs. 3 and 4).

lished architecture of hierarchical confidence generation, our TMS results can be recreated by assuming a role for DLPFC in the reading out the sensory signal strength at the meta level and a role for aPFC in making the final confidence judgment based on a combination of perceptual and nonperceptual factors.

Discussion

We sought to determine the distinct roles of subregions of the PFC in visual metacognition. Previous research identified the DLPFC and aPFC as being critical to metacognitive computations, but a mechanistic understanding of their functions in confidence judgments is still lacking (Shimamura, 2000a; Fleming and Dolan, 2012). We proposed a neural mechanism for confidence computation where DLPFC reads out the sensory signal strength and relays it to aPFC, whereas aPFC makes the confi-

dence judgment by integrating this readout with nonperceptual factors. Based on this architecture, we predicted that disrupting DLPFC would affect average confidence (without affecting metacognitive ability), whereas disrupting aPFC would affect metacognitive ability (without affecting confidence). A causal intervention with online TMS confirmed these predictions. Further, we simulated a confidence generation model that incorporated our hypothesized neural mechanism and successfully reproduced the observed behavioral results. These findings establish the existence of independent causal contributions of DLPFC and aPFC to confidence generation and suggest specific mechanistic roles for these prefrontal sites. Further, they indicate that a significant portion of confidence computation in PFC takes place 250–450 ms after stimulus onset.

Role of DLPFC in confidence computation

Our experiment tested the hypothesis that the role of DLPFC in confidence computation is to read out the strength of the sensory signal and relay it to aPFC. We derived this hypothesis from previous studies showing that DLPFC activity is related to the level of confidence but not to metacognitive ability (Henson et al., 2000; Fleck et al., 2006; Lau and Passingham, 2006). This proposed function of DLPFC in reading out the sensory signal strength is consistent with the view that DLPFC maintains, re-routes, and facilitates manipulations of sensory information (Shimamura, 2000b; Fleming and Dolan, 2012).

The correlation between DLPFC activity and confidence level has received different interpretations. Henson et al. (2000) hypothesized that DLPFC activity reflects retrieval monitoring in a memory task. Fleck et al. (2005) suggested a general role for DLPFC in information monitoring during decision making. Finally, Lau and Passingham (2006) theorized that DLPFC plays a role in conscious perception. Our proposal, that the role of DLPFC in confidence computations is to read out the strength of the sensory signal, is not necessarily at odds with these previous theories. Instead, here, we specify a precise computational role for DLPFC in the domain of confidence generation.

There has been some controversy about whether DLPFC is involved more directly in confidence computation. Rounis et al. (2010) delivered bilateral TBS to DLPFC and reported a decrease in mean visibility as well as metacognitive performance. These findings have been controversial: Bor et al. (2017) argued that they could not replicate them, whereas Ruby et al. (2018) disputed Bor et al.'s exclusion criteria and argued that the original effects replicate under different exclusion criteria. Although our study certainly has implications about the role of DLPFC in metacognition, it is not clear whether it can be used to inform the above debate. Indeed, both studies above (Rounis et al., 2010; Bor et al., 2017) targeted a relatively posterior portion of DLPFC, whereas we targeted a relatively anterior DLPFC region. DLPFC is anatomically large and it is likely that its different subregions have different functions. Other important differences between ours and the two studies above include the use of an online versus offline TMS protocol, unilateral versus bilateral stimulation, and confidence versus visibility ratings, respectively, with each of these factors making direct comparisons difficult.

Role of aPFC in confidence computation

Our experiment tested the hypothesis that the role of aPFC in confidence computation is to decide the exact value of the confidence rating based on both the sensory readout relayed by DLPFC and other, nonperceptual factors. Consistent with this hypothesis, many previous studies have found a link between aPFC and metacognitive ability (Fleming et al., 2010, 2012b, 2014; Yokoyama et al., 2010; McCurdy et al., 2013; Rahnev et al., 2016; Ryals et al., 2016; Allen et al., 2017). Our proposal that aPFC is the seat of metacognitive computation is also consistent with the view that aPFC is at the highest level in the cognitive and perceptual decision-making hierarchy (Badre and D'Esposito, 2009; Fleming and Dolan, 2012; Rahnev, 2017).

A wide range of higher-order functions in the domains of memory, cognition, and perceptual decision making have been attributed to aPFC. These functions include top-down manipulations of working memory representations, switching between task sets, attentional allocation to subgoals, and relational integration (Koechlin et al., 1999; Kaas et al., 2007; Domenech and Koechlin, 2015; Lara and Wallis, 2015; Parkin et al., 2015). Ramnani and Owen (2004) integrate these theories into a common

framework proposing that aPFC recruitment facilitates the coordination of information processing from separate mental processes toward a higher goal. This view is fully consistent with our theory's implication of aPFC in generating metacognitive computations. Indeed, assessing the confidence in one's own perceptual decisions requires the integration of both perceptual and nonperceptual factors (Fleming and Dolan, 2012).

We found that TMS influenced metacognitive ability only for the second half of blocks. It appears that a sustained period of inhibition may be required to influence metacognitive ability. Indeed, previous studies that successfully manipulated metacognitive ability (Rahnev et al., 2016; Ryals et al., 2016) used offline TMS, which involves a sustained period of stimulation. More research is needed to determine whether TMS may interact differently with the unique cytoarchitectonic characteristics of aPFC (Semendeferi et al., 2001).

Disrupting the activity of aPFC during confidence computation improved metacognitive performance. Although such an improvement appears surprising at first, it is consistent with previous studies finding increases in metacognitive ability after offline TMS to aPFC (Rahnev et al., 2016; Ryals et al., 2016). One possible explanation for this increased ability is that aPFC TMS increased the attentional resources for the confidence decision. However, increased attentional resources could be expected to also lead to increases in d' and confidence, but aPFC TMS had no effect on either of these measures. Another possibility is that TMS might have inhibited the influence of certain factors that are detrimental to metacognition. For example, people consider their confidence history while making a confidence judgment, a phenomenon called confidence leak (Rahnev et al., 2015). Confidence ratings may also be contaminated by other factors such as arousal (Allen et al., 2016) and action fluency (Fleming et al., 2015). The use of these extra factors generally decreases metacognitive ability in laboratory settings (Rahnev et al., 2015). Therefore, the improvement of metacognitive ability with aPFC TMS in our study may stem from the reduced use of some of these nonperceptual factors in confidence generation.

Computational model

We built a computational model that instantiates the hypothesized neural mechanism regarding the roles of DLPFC and aPFC. It is important to note that, whereas the TMS data provide support for the proposed neural mechanism, our experiment was not designed to corroborate the computational model directly. Instead, the role of the computational model was to verify that the substantive claims made by our neural mechanism could indeed lead to the pattern of behavioral results that we observed. We have explored the plausibility of our computational model elsewhere (Bang et al., 2017).

We modeled the effect of TMS on aPFC and DLPFC as a decrease in metacognitive noise and a decrease of signal in the sensory readout (r_{readout}), respectively. The modeling choice for aPFC TMS is natural given that, within our model, metacognitive ability is controlled by the metacognitive noise parameter. However, the effects of DLPFC TMS on decreased confidence can also be explained as a shift in the confidence criteria. The reason that we do not favor this explanation is because it is unclear why TMS would shift the criteria in one direction and not the other. Specifically, we are not aware of any mechanism that predicts that TMS would increase the confidence criteria (to decrease confidence). Instead, our explanation, that TMS causes a loss of signal that leads to a confidence decrease, relates more naturally to the expected effect of TMS, which is to disrupt neural activity.

Conclusion

Our results show that TMS produced distinct effects on confidence measures depending on which prefrontal site was stimulated: TMS to DLPFC decreased confidence, whereas TMS to aPFC increased metacognitive ability for the second half of the experimental blocks. This dissociation confirms our hypothesis that DLPFC and aPFC have distinct roles in visual metacognition. Further, it supports our hypothesized neural mechanism, according to which DLPFC reads out the sensory signal strength and relays it to aPFC for the confidence computation. Simulations of a confidence generation model based on our neural mechanism reproduced the observed TMS effects and thus corroborated this mechanism. Together, our results uncover the functional organization of PFC for confidence computations.

References

- Allen M, Frank D, Samuel Schwarzkopf DS, Fardo F, Winston JS, Hauser TU, Rees G (2016) Unexpected arousal modulates the influence of sensory noise on confidence. *Elife* 5:e18103. [CrossRef Medline](#)
- Allen M, Glen JC, Müllensiefen D, Schwarzkopf DS, Fardo F, Frank D, Callaghan MF, Rees G (2017) Metacognitive ability correlates with hippocampal and prefrontal microstructure. *Neuroimage* 149:415–423. [CrossRef Medline](#)
- Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci* 10:659–669. [CrossRef Medline](#)
- Bang JW, Shekhar M, Rahnev D (2017) Sensory noise increases metacognitive efficiency. Available at: <https://www.biorxiv.org/content/early/2017/09/15/189399>. Accessed February 24, 2018.
- Bor D, Schwartzman DJ, Barrett AB, Seth AK (2017) Theta-burst transcranial magnetic stimulation to the prefrontal or parietal cortex does not impair metacognitive visual awareness. *PLoS One* 12:e0171793. [CrossRef Medline](#)
- Borckardt JJ, Nahas Z, Koola J, George MS (2006) Estimating resting motor thresholds in transcranial magnetic stimulation research and practice. *J ECT* 22:169–175. [CrossRef Medline](#)
- De Martino B, Fleming SM, Garrett N, Dolan RJ (2013) Confidence in value-based choice. *Nat Neurosci* 16:105–110. [CrossRef Medline](#)
- Domenech P, Koechlin E (2015) Executive control and decision-making in the prefrontal cortex. *Curr Opin Behav Sci* 1:101–106. [CrossRef](#)
- Fleck MS, Daselaar SM, Dobbins IG, Cabeza R (2006) Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and non-memory tasks. *Cereb Cortex* 16:1623–1630. [CrossRef Medline](#)
- Fleming SM, Dolan RJ (2012) The neural basis of metacognitive ability. *Philos Trans R Soc Lond B Biol Sci* 367:1338–1349. [CrossRef Medline](#)
- Fleming SM, Lau HC (2014) How to measure metacognition. *Front Hum Neurosci* 8:443. [CrossRef Medline](#)
- Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329:1541–1543. [CrossRef Medline](#)
- Fleming SM, Dolan RJ, Frith CD (2012a) Metacognition: computation, biology and function. *Philos Trans R Soc Lond B Biol Sci* 367:1280–1286. [CrossRef Medline](#)
- Fleming SM, Huijgen J, Dolan RJ (2012b) Prefrontal contributions to metacognition in perceptual decision making. *J Neurosci* 32:6117–6125. [CrossRef Medline](#)
- Fleming SM, Ryu J, Golfinos JG, Blackmon KE (2014) Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* 137:2811–2822. [CrossRef Medline](#)
- Fleming SM, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H (2015) Action-specific disruption of perceptual confidence. *Psychol Sci* 26:89–98. [CrossRef Medline](#)
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York, NY: John Wiley & Sons Ltd.
- Hellige JB (1996) Hemispheric asymmetry for visual information processing. *Acta Neurobiol Exp (Wars)* 56:485–497. [Medline](#)
- Henson RN, Rugg MD, Shallice T, Dolan RJ (2000) Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. *J Cogn Neurosci* 12:913–923. [CrossRef Medline](#)
- Jang Y, Wallsten TS, Huber DE (2012) A stochastic detection and retrieval model for the study of metacognition. *Psychol Rev* 119:186–200. [CrossRef Medline](#)
- Janowsky JS, Shimamura AP, Kritchevsky M, Squire LR (1989) Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behav Neurosci* 103:548–560. [CrossRef Medline](#)
- Kaas AL, van Mier H, Goebel R (2007) The neural correlates of human working memory for haptically explored object orientations. *Cereb Cortex* 17:1637–1649. [CrossRef Medline](#)
- Koechlin E, Basso G, Pietrini P, Panzer S, Grafman J (1999) The role of the anterior prefrontal cortex in human cognition. *Nature* 399:148–151. [CrossRef Medline](#)
- Koriat A (2007) Metacognition and consciousness. In: *The Cambridge handbook of consciousness* (Zelazo PD, Moscovitch M, Thompson E, eds) pp 289–325. New York, NY: Cambridge University Press. [CrossRef](#)
- Lara AH, Wallis JD (2015) The role of prefrontal cortex in working memory: a mini review. *Front Syst Neurosci* 9:173. [CrossRef Medline](#)
- Lau HC, Passingham RE (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc Natl Acad Sci U S A* 103:18763–18768. [CrossRef Medline](#)
- Macmillan N A, Creelman CD (2005) Detection theory: a user's guide. Available at: https://beluga.sub.uni-hamburg.de/vufind/Record/01845402X?institution=GBV_ILN_22&rank=1. Accessed March 26, 2018.
- Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 21:422–430. [CrossRef Medline](#)
- Maniscalco B, Lau H, Maniscalco B, Lau ÁH, Lau H, Fleming SM, Frith CD (2014) Signal detection theory analysis of type 1 and type 2 data: Meta-d, response-specific Meta-d, and the unequal variance SDT Model. In: *The cognitive neuroscience of metacognition*, pp 25–65. Available at: http://www.columbia.edu/~bsm2105/type2sdt/metad_rs.pdf. Accessed February 5, 2018.
- Maniscalco B, McCurdy LY, Odegaard B, Lau H (2017) Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *J Neurosci* 37:1213–1224. [CrossRef Medline](#)
- McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, de Lange FP, Lau H (2013) Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J Neurosci* 33:1897–1906. [CrossRef Medline](#)
- Metcalfe J, Shimamura AP (1994) *Metacognition: knowing about knowing*. Cambridge, MA: MIT.
- Morales J, Lau H, Fleming SM (2017) Domain-specific patterns of activity support metacognition in human prefrontal cortex. Available at: <https://www.biorxiv.org/content/early/2017/08/04/172445>. Accessed October 14, 2017.
- Mueller ST, Weidemann CT (2008) Decision noise: An explanation for observed violations of signal detection theory. *Psychon Bull Rev* 15:465–494. [CrossRef Medline](#)
- Nelson TO, Narens L (1990) Metamemory: a theoretical framework and some new findings. *Psychol Learn Motiv* 6:125–173. [CrossRef](#)
- Parkin BL, Hellyer PJ, Leech R, Hampshire A (2015) Dynamic Network Mechanisms of Relational Integration. *J Neurosci* 35:7660–7673. [CrossRef Medline](#)
- Rahnev D (2017) Top-down control of perceptual decision making by the prefrontal cortex. *Curr Dir Psychol Sci* 26:464–469. [CrossRef](#)
- Rahnev D, Koizumi A, McCurdy LY, D'Esposito M, Lau H (2015) Confidence leak in perceptual decision making. *Psychol Sci* 26:1664–1680. [CrossRef Medline](#)
- Rahnev D, Nee DE, Riddle J, Larson AS, D'Esposito M (2016) Causal evidence for frontal cortex organization for perceptual decision making. *Proc Natl Acad Sci U S A* 113:6059–6064. [CrossRef Medline](#)
- Ramnani N, Owen AM (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci* 5:184–194. [CrossRef Medline](#)
- Ridding MC, Ziemann U (2010) Determinants of the induction of cortical plasticity by non-invasive brain stimulation in healthy subjects. *J Physiol* 588:2291–2304. [CrossRef Medline](#)
- Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn Neurosci* 1:165–175. [CrossRef Medline](#)
- Ruby E, Maniscalco B, Peters MAK (2018) On a 'failed' attempt to manipulate visual metacognition with transcranial magnetic stimulation to prefrontal cortex. *Conscious Cogn* 62:34–41. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1053810017305226>. Accessed May 5, 2018.

- Ryals AJ, Rogers LM, Gross EZ, Polnaszek KL, Voss JL (2016) Associative recognition memory awareness improved by theta-burst stimulation of frontopolar cortex. *Cereb Cortex* 26:1200–1210. [CrossRef Medline](#)
- Semendeferi K, Armstrong E, Schleicher A, Zilles K, Van Hoesen GW (2001) Prefrontal cortex in humans and apes: A comparative study of area 10. *Am J Phys Anthropol* 114:224–241. [CrossRef Medline](#)
- Shimamura AP (2000a) Toward a cognitive neuroscience of metacognition. *Conscious Cogn* 9:313–323; discussion 324–326. [CrossRef Medline](#)
- Shimamura AP (2000b) The role of the prefrontal cortex in dynamic filtering. *Psychobiology* 28:207–218.
- Shimamura AP, Squire LR (1986) Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. *J Exp Psychol Learn Mem Cogn* 12:452–460. [Medline](#)
- Siegel M, Buschman TJ, Miller EK (2015) Cortical information flow during flexible sensorimotor decisions. *Science* 348:1352–1355. [CrossRef Medline](#)
- van den Berg R, Yoo AH, Ma WJ (2017) Fechner's law in metacognition: a quantitative model of visual working memory confidence. *Psychol Rev* 124:197–214. [CrossRef Medline](#)
- Yeung N, Summerfield C (2012) Metacognition in human decision-making: confidence and error monitoring. *Philos Trans R Soc Lond B Biol Sci* 367:1310–1321. [CrossRef Medline](#)
- Yokoyama O, Miura N, Watanabe J, Takemoto A, Uchida S, Sugiura M, Horie K, Sato S, Kawashima R, Nakamura K (2010) Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neurosci Res* 68:199–206. [CrossRef Medline](#)