Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness

Elisabeth Rounis [a] , Brian Maniscalco [b] , John C. Rothwell [a] , Richard E. Passingham [a c] & Hakwan Lau [a b c]

[a] University College London, London, UK

[b] Columbia University in the City of New York, New York, USA

[c] University of Oxford, Oxford, UK

Available online: 23 Mar 2010

PLEASE SCROLL DOWN FOR ARTICLE

Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness

**Elisabeth Rounis[1], Brian Maniscalco[2], John C. Rothwell[1],
Richard E. Passingham[1,3], and Hakwan Lau[1,2,3]**

[1]University College London, London, UK
[2]Columbia University in the City of New York, New York, USA
[3]University of Oxford, Oxford, UK

We used a recently developed protocol of transcranial magnetic stimulation (TMS), theta-burst stimulation, to bilaterally depress activity in the dorsolateral prefrontal cortex as subjects performed a visual discrimination task. We found that TMS impaired subjects' ability to discriminate between correct and incorrect stimulus judgments. Specifically, after TMS subjects reported lower visibility levels for correctly identified stimuli, as if they were less fully aware of the quality of their visual information processing. A signal detection theory analysis confirmed that the results reflect a change in metacognitive sensitivity, not just response bias. The effect was specific to metacognition; TMS did not change stimulus discrimination performance, ruling out alternative explanations such as TMS impairing visual attention. Together these results suggest that activations in the prefrontal cortex in brain imaging experiments on visual awareness are not epiphenomena, but rather may reflect a critical metacognitive process.

INTRODUCTION

Current theories suggest that the prefrontal cortex plays an important role in visual awareness (e.g. Crick & Koch, 1995; Dehaene, Sergent, & Changeux, 2003).

This hypothesis has been supported by a number of brain imaging studies (Marois, Yi, & Chun, 2004; Rees, Kreiman, & Koch, 2003; Sahraie et al., 1997). However, critics have claimed that there has been a lack of neuropsychological demonstration of the

Correspondence should be addressed to: Brian Maniscalco, Columbia University, Department of Psychology, 1190 Amsterdam Avenue, MC 5501, New York, NY 10027, USA.

E-mail: brian@psych.columbia.edu

essential role of the prefrontal cortex in visual awareness: Damage to the prefrontal cortex does not seem to lead to cortical blindness (Pollen, 1995). Here we attempt to clarify this issue by showing that bilateral TMS to the prefrontal cortex does have an effect on visual awareness, in particular the metacognitive sensitivity with which it discriminates between effective and ineffective stimulus processing. While we agree with critics that disruption of prefrontal activity may have little or no effect on primary visual awareness, i.e., the ability to represent visual targets, higher monitoring aspects of awareness may critically depend on prefrontal activity.

Recent studies have shown that visual awareness can be assessed by metacognitive procedures (Kolb & Braun, 1995; Persaud, McLeod, & Cowey, 2007). Typically, when awareness is lacking, one tends to place low or inappropriate subjective ratings (Weiskrantz, 1997). Such metacognitive approaches have also been used in other studies of visual awareness (Galvin, Podd, Drga, & Whitmore, 2003; Kolb & Braun, 1995; Lau & Passingham, 2006; Szczepanowski & Pessoa, 2007) and implicit learning (Dienes & Perner, 1999; Persaud et al., 2007). In general, one can assess metacognitive sensitivity by measuring how well subjective ratings (e.g., of confidence or visibility) distinguish

between correct and incorrect judgments (e.g., about the identity of a presented stimulus). High levels of metacognitive sensitivity imply that subjects are introspectively aware of the effectiveness of their internal information processing (Kolb & Braun, 1995; Galvin et al., 2003). As there is not yet widespread agreement on the ideal measure of metacognitive sensitivity, in the present study we use two separate approaches—a correlation approach and a signal detection theory approach—and demonstrate converging interpretations of the data.

We required volunteers to perform a two-alternative forced-choice visual task, identifying the spatial arrangement of two visual stimuli (a square and a diamond, Figure 1A). At the same time, they also rated the subjective visibility of the stimuli ("clear" or "unclear"). Subjects performed these tasks before and after transcranial magnetic stimulation (TMS), which was aimed at the dorsolateral prefrontal cortex (DLPFC, Figure 1B). We applied to this region thetaburst stimulation (TBS), a recently developed protocol that is known to effectively depress cortical excitability by mimicking the action of long-term potentiation and long-term depression in cortical tissues (Huang et al., 2005). One advantage of this technique is that the effect of 20 s of stimulation is known to last for up to
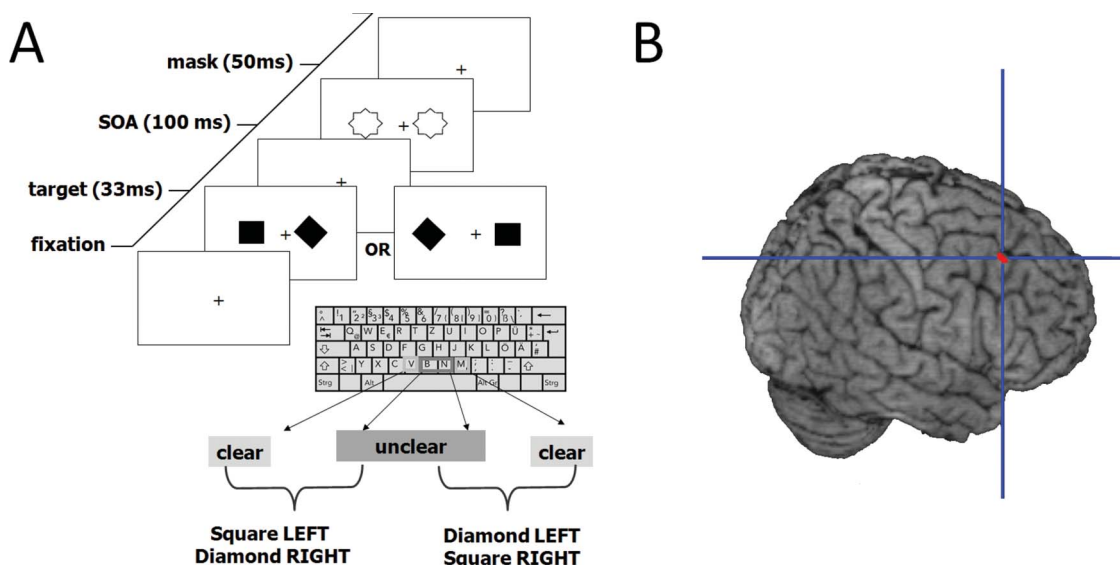
Figure 1. Experimental design. (A) Visual task and stimuli. Volunteers were required to perform a two-alternative forced-choice visual task, identifying the spatial arrangement of two visual stimuli (square on the left and diamond on the right, or the other way round). They rated the subjective visibility ("clear" or "unclear") at the same time. So in every trial subjects had four options as to which key to press in order to respond. (B) Site of stimulation. The dorsolateral prefrontal cortex (DLPFC) was the targeted site of stimulation, and was chosen because neural activity from this area has been shown to reflect a difference in the subjective ratings of visibility even when performance in a forced-choice visual task was matched (Lau & Passingham, 2006). The image showing the site of stimulation is based on magnetic resonance brain scans of 6 of the 20 subjects in this study. The scans were collected after completion of the TMS experiments. Right and left DLPFC coordinates were [37 26 50] and [–41 18 52], with standard deviations [4.6 5.6 5.3] and [4.3 5.1 3.8] respectively.

20 min, which means we had the opportunity to depress both sides of the DLPFC by stimulating them sequentially. We opted for bilateral stimulation as this has been suggested to be critical: Sahraie et al. (1997) have suggested that one reason visual defects do not seem to frequently follow prefrontal lesions may be that such lesions have to be large and bilateral. Using this sequential method to depress the DLPFC bilaterally, we found that the metacognitive sensitivity of reported visual awareness was reduced after TMS.

METHOD

Subjects

Twenty healthy volunteers (eight women, mean age 25.6 ± 6.1), with normal or corrected-to-normal vision and no history of neurological disorders or head injury were recruited from the database of volunteers at the Functional Imaging Laboratory, Institute of Neurology, University College London. Written informed consent was obtained from all participants. The study was approved by the joint ethics committee for the National Hospital for Neurology and Neurosurgery (UCLH NHS Trust) and the Institute of Neurology (UCL).

Experimental design

Subjects were asked to perform a two-alternative forced-choice task (Figure 1A). Testing was performed in a darkened room. Stimuli were presented against the white background of a CRT monitor refreshing at 120 Hz. The monitor was placed 40 cm away from the subjects' eyes. On each trial, a diamond and a square were presented on either side of a central crosshair for 33 ms. The stimuli had sides measuring 0.8° of visual angle and were centered 1° to the left and right of the central crosshair. 100 ms after stimulus onset, a metacontrast mask was displayed for 50 ms in order to enhance task difficulty. The two possibilities for the sequence of stimuli (square on the left and diamond on the right, and vice versa) were presented with equal probability in a pseudorandom order. The subjects' task was to identify which stimulus sequence had just been presented, square left/diamond right or vice versa. At the same time, subjects gave subjective ratings of stimulus visibility ("clear" or "unclear"). Subjects were instructed to make the visibility judgment in a relative manner, to distinguish between stimuli that were relatively

more or less visible. Since stimulus contrast was adjusted so as to yield threshold performance on the stimulus classification task, stimuli used in this experiment were somewhat difficult to see. Nonetheless, subjects were instructed to judge stimulus visibility on each trial relative to the context of stimuli used in this experiment. For instance, a subject might judge that the stimulus on a certain trial was more readily visible than the majority of stimuli seen in the experiment up to that point, even if its visibility was poor by everyday standards. Subjects were encouraged to judge such stimuli as exhibiting "high clarity," i.e., having relatively high clarity compared to other stimuli observed in the experimental context.

Subjects attended two separate testing sessions, both preceded by a demonstration and a practice phase of 100 trials intended to familiarize the subjects with the task and to allow them to reach a stable level of performance. Performance level was controlled to be at approximately 75% correct throughout the experiment by titrating the contrast of the stimuli, using a standard up–down transformed-response staircasing procedure (Macmillan & Creelman, 2005). Each trial was randomly designated as belonging to staircase A or staircase B. For staircase A, contrast on the current trial was increased if the subject responded incorrectly on the previous "A" trial, whereas contrast on the current trial was decreased if the subject responded correctly on the two previous "A" trials. Staircase B worked in a similar manner, except it required three consecutive correct responses on "B" trials in order to reduce contrast. After practice, subjects underwent an initial ("pre") block of 300 trials to measure forced-choice task performance and subjective ratings of visibility. On average this took 10.9 min, excluding brief breaks after every 100 trials. After completing this block, two real (or sham) continuous TBS (cTBS) conditioning stimulations, one to the left and one to the right, were delivered to the dorsolateral prefrontal area. The two stimulations were separated by a 1-min intertrain interval. Following real (or sham) stimulation, subjects did another ("post") block of 300 trials. On average this took 10.4 min, excluding brief breaks after every 100 trials. Session order by type of cTBS (real vs. sham) was counterbalanced across subjects.

Theta-burst stimulation

In each TBS session, 600 biphasic stimuli, at a stimulation intensity of 80% of active motor threshold (AMT) for the right first dorsal interosseous (FDI) hand muscle, were given over the left and right

DLPFC area using a Magstim Super Rapid stimulator (Whitland, UK) connected to four booster modules. The conditioning cTBS stimuli were delivered in two separate 20-s trains of 300 cTBS pulses, one for the left and one for the right, separated by an intertrain interval of 1 min. A similar bilateral procedure has been used in a recent clinical study (Artfeller, Vonthein, Plontke, & Plewnia, 2009).

A standard figure-of-eight-shaped coil (Double 70 mm Coil Type P/N 9925; Magstim) was used for both real and sham cTBS. Real cTBS was delivered with the coil placed tangentially to the scalp with the handle pointing posteriorly. In sham cTBS sessions, the coil was placed perpendicularly to the scalp, an ineffective position for the delivery of conditioning pulses, which provided comparable acoustic stimuli to the real cTBS condition. The coil was positioned with the handle at 45° to the sagittal plane. The current flow in the initial rising phase of the biphasic pulse in the biphasic pulse induced a posterior-to-anterior current flow in the underlying cortex.

The basic TBS pattern was a burst containing three pulses of 50 Hz magnetic stimulation given in 200 ms intervals (i.e. at 5 Hz). In the continuous theta burst stimulation paradigm (cTBS), a 20 s train of uninterrupted TBS is given (300 pulses or 100 bursts). Physiological studies have shown that this produces a decrease in corticospinal excitability which lasts for about 20 min (Huang et al., 2005), when applied to the primary motor cortex, M1. This new rTMS paradigm has been developed recently in our laboratory and has the advantage of being a rapid and efficient method of conditioning, which has effects on corticospinal excitability that have been shown to involve similar mechanisms to long-term potentiation/depression (LTP/LTD) with NMDA dependence (Huang et al., 2007), as well as effects on behavior and learning (Huang et al., 2005; Talelli, Greenwood, & Rothwell, 2007).

The site of cTBS stimulation was located 5 cm anterior to the "motor hot spot" on a line parallel to the midsagittal line. This DLPFC location has been used in previous studies and can be shown consistently on structural scans (Mottaghy, Gangitano, Sparing, Krause, & Pascual-Leone, 2002; Rounis et al., 2006; Figure 1B). The position of the motor hot spot was defined functionally as the point of maximum evoked motor response in the slightly contracted right FDI. The active motor threshold was defined as the lowest stimulus intensity that elicited at least five twitches in 10 consecutive stimuli given over the motor hot spot, while the subject was maintaining a voluntary contraction of about 20% of maximum using visual feedback.

The use of such low subthreshold intensity (80% AMT) had the advantage of decreased spread of stimulation away from the targeted site, thus keeping the area that was stimulated with the conditioning pulses more focal (Pascual-Leone, Valls-Solé, Wassermann, & Hallett, 1994; Münchau, Bloem, Irlbacher, Trimble, & Rothwell, 2002). Also, a previous study on the prefrontal cortex that applied intensity above motor threshold reported unpleasant vagal reactions in subjects (Grossheinrich et al., 2009). However, even at that higher intensity there was no adverse effects on mood, seizure or epileptiform observed in the recorded electroencephalogram. This suggests that our stimulation at this lower intensity should be safe to our subjects.

Data analysis

Metacognitive sensitivity (i.e. the efficacy with which visibility ratings distinguish between correct and incorrect responses) was assessed using two separate methods. The first method followed previous studies (e.g., Kolb & Braun, 1995; Kornell, Son, & Terrace, 2007) in using the correlation between accuracy and subjective rating as a measure of metacognitive sensitivity. We used the correlation coefficient phi, which quantifies the degree of correlation between two binary variables, to calculate the correlation between task accuracy (correct/incorrect) and stimulus visibility (clear/unclear). Phi is equivalent to Pearson's r computed for two binary variables, and like r it ranges from –1 (perfect negative correlation) to +1 (perfect positive correlation). We calculated phi for the 300 trials pre and post real and sham TMS for each subject. We predicted that TMS would hinder metacognitive sensitivity, and thus that there would be a TMS (real/sham) × time (pre/post) interaction.

We also performed a signal detection theoretic (SDT) analysis to estimate metacognitive sensitivity. On this analysis, we estimate a value called meta-d', which is the amount of signal available for one's metacognitive disposal (i.e. available for doing the confidence/visibility rating task). This measure is in the same scale as d', i.e., the signal that is available for the primary stimulus classification task, so the two can be directly compared. If meta-$d' < d'$, it means that some signal that is available for the primary stimulus classification task is lost for metacognition, which means that metacognitive sensitivity is not perfect.

The need to perform an SDT analysis is due to the fact that phi can be shown to generate non-regular receiver operating characteristic (ROC) curves, which in turn implies an underlying threshold model of

detection (Swets, 1986). The ROC profile and threshold model of phi are not in good agreement with the standard SDT model (Macmillan & Creelman, 2005), nor with a recent treatment of the "type 2" signal detection model that characterizes metacognitive performance within an SDT framework (Galvin et al., 2003). (In the following, we use the terms "metacognitive" and "type 2" interchangeably.) The consequence of this is that phi may confound sensitivity and response bias, rather than being a pure measure of sensitivity. Thus, we also performed a type 2 SDT analysis of the data.

The fundamental idea in SDT is to use observed hit rate (HR) and false alarm rate (FAR) data to estimate an observer's sensitivity (d') and response bias (c) (Macmillan & Creelman, 2005; Figure 2A). Likewise, the fundamental idea of a type 2 SDT analysis is to use type 2 HR (i.e. the frequency with which correct responses are endorsed with high confidence) and type 2 FAR (i.e. the frequency with which incorrect responses are endorsed with high confidence) to estimate the sensitivity and response bias an observer's confidence ratings (or, analogously, visibility ratings) exhibit in classifying first-order stimulus judgments as correct or incorrect. However, the underlying formalisms for the type 2 SDT model are quite complex (Galvin et al., 2003), and there is as yet no widespread agreement on how to perform a proper type 2 SDT analysis. One could estimate the type 2 ROC curve, and measure type 2 sensitivity by the area under the curve. However, this measure would depend on type 1 d' (Galvin et al., 2003), and any observed change in type 2 sensitivity could therefore be confounded. We therefore developed a measure called meta-d', i.e. the signal that is available for the type 2 task, and compared it against type 1 d'. Below is how we calculate this meta-d' measure.

We characterize type 2 sensitivity by capitalizing on the observation that, on the classic SDT model, the sensitivity component of type 2 HR and FAR is already determined by the sensitivity for stimulus discrimination, d', in conjunction with the criterion for stimulus classification response, c (Galvin et al., 2003). That is, once d' and c for the classic SDT model are fixed, a type 2 ROC curve defining the tradeoff between type 2 HR and FAR is already implied for each type of stimulus response (Figure 2B). Thus, we characterize type 2 sensitivity as the d' an ideal SDT observer with a fixed stimulus classification criterion c would require in order to produce the observed type 2 HR and FAR data (Figure 2C). This measure (call it meta-d') can then be compared to the observer's actual d' in order to assess how well the observer's observed type 2 sensitivity compares to the

theoretically ideal type 2 sensitivity, according to the classic SDT model. For an ideal SDT observer, meta-$d' = d'$; for suboptimal metacognitive sensitivity, meta-$d' < d'$; and for an observer whose confidence ratings are not diagnostic of judgment accuracy at all, meta-$d' = 0$. We predicted that the difference meta-$d' - d'$ would decrease more following real TMS than sham TMS, indicating that TMS hinders subjects' metacognitive sensitivity, relative to the theoretical ideal.

We estimated meta-d' for each subject, in each TMS × time condition, as follows. First, we estimated the SDT parameters c' (the stimulus classification criterion measured relative to d') and s (the ratio of standard deviations of internal evidence for the two stimulus classes) (Macmillan & Creelman, 2005; Figure 2A). Holding c' and s constant, we estimated values for meta-d' and the type 2 criterion $c_{\text{conf | response="square left/diamond right"}}$ that would minimize the sum of squared errors between observed and modeled type 2 HR and FAR for all trials where the subject classified the stimulus as "square left/diamond right." We then estimated values for meta-d' and the type 2 criterion $c_{\text{conf | response="diamond left/square right"}}$ that would minimize the sum of squared errors between observed and modeled type 2 HR and FAR for all trials where the subject classified the stimulus as "diamond left/square right." (Compare to Figure 2, where, for example, "square left/diamond right" corresponds to "S1" and "diamond left/square right" corresponds to "S2"). Thus, we generated two estimates of meta-d', corresponding to the subject's type 2 HR and FAR conditional on each stimulus classification type. The two estimates were combined via a weighted average, where the weight of each meta-d' estimate was determined by the number of trials used to estimate it. The mean SSE corresponding to each meta-d' estimate was 9.1×10^{-5}, indicating that this approach provided an excellent fit to the observed type 2 HR and FAR data. Minimization of SSE was achieved using the Optimization Toolbox in MATLAB® (MathWorks, Natick, MA).

Because we are testing a directional hypothesis in a 2 × 2 factorial design (i.e., metacognitive sensitivity is reduced following real TMS more than following sham TMS), we report halved p-values for the TMS × time interaction on phi and meta-$d' - d'$.

RESULTS

In the following we present ANOVA analyses with within-subject factors of TMS (real/sham) and time (pre/post) for several independent variables of interest such as accuracy and response time for correct
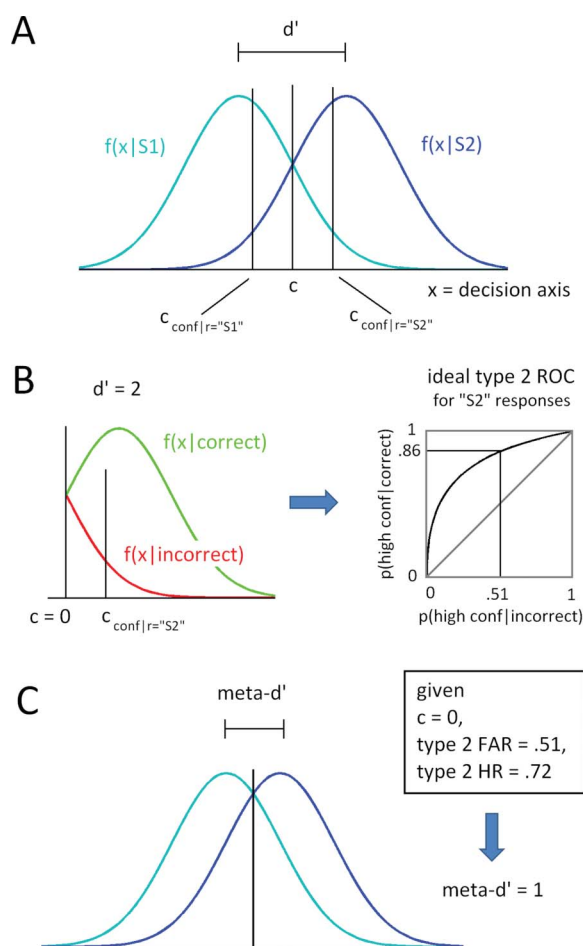
Figure 2. SDT analysis of type 2 (metacognitive) performance. The basic idea of this analysis is to compute meta-d', a measure of the signal that is available for one's metacognitive disposal (i.e., available for making subjective ratings). This measure is in the same scale as d', i.e., the signal available for the primary forced-choice task, such that the two can be directly compared. If meta-$d' < d$, it means that some signal that is available for the primary forced-choice task is lost in the rating, i.e. the subject is not metacognitively perfect. (A) The classic SDT model. The observer must discriminate between stimulus classes S1 and S2. Each stimulus presentation generates a value on an internal decision axis, corresponding to the evidence in favor of S1 or S2. Evidence generated by each stimulus class is normally distributed across the decision axis, and the normalized distance between these distributions (d') measures how well the observer can discriminate S1 from S2. The observer sets a decision criterion c, such that all signals exceeding c are labeled "S2" and all those failing to exceed c "S1." The observer also sets criteria $c_{\text{conf} | r="S1"}$ and $c_{\text{conf} | r="S2"}$ to determine confidence ratings (higher ratings for signals farther from c). For expositional ease in this hypothetical example, we set $d' = 2$, $c = 0$, and s (the ratio of standard deviations for the two distributions) = 1; in the actual analysis, d', c, and s are estimated from data. (B) Type 2 sensitivity from d' and c. Consider only trials where the observer responds "S2." Then the S2 distribution corresponds to the distribution of evidence for correct responses, and the S1 distribution corresponds to the distribution of evidence for incorrect responses. All trials surpassing $c_{\text{conf} | r="S2"}$ are endorsed with high confidence. Sweeping the $c_{\text{conf} | r="S2"}$ criterion across the decision axis generates different values for type 2 false alarm rate, $p(\text{high confidence} | \text{incorrect})$, and type 2 hit rate, $p(\text{high confidence} | \text{correct})$, and thus generates a type 2 ROC curve. (Similar considerations hold for "S1" responses.) Thus, d' and c are jointly sufficient to determine type 2 sensitivity for each response type, according to the standard SDT model. (C) Characterizing type 2 sensitivity. The analysis from (A) and (B) can be reversed to characterize metacognitive sensitivity. Suppose that the observer produced type 2 FAR = .51, type 2 HR = .72. We then ask: "What d' would an SDT-optimal observer with stimulus classification criterion c require in order to produce the observed type 2 FAR and type 2 HR?" The answer is meta-d', a characterization of the sensitivity with which confidence ratings discriminate between correct and incorrect judgments. In the example in the figure, meta-$d' = 1$ even though observed $d' = 2$, indicating suboptimal metacognitive sensitivity.

trials. None of these analyses exhibited a main effect of TMS condition (F values < 1.7), indicating that the real and sham TMS sessions were comparable on baseline task performance.

Stimulus contrast was adjusted online in order to control classification accuracy; thus, as expected,

frequency of correct responses did not vary as a function of time, $F(1, 19) = 2.45$, $MSE = 0.001$) or the TMS × time interaction, $F(1, 19) = 0.002$, $MSE = 0.001$ (Figure 3A). A more insightful measure of stimulus classification performance is the mean contrast required to keep classification accuracy
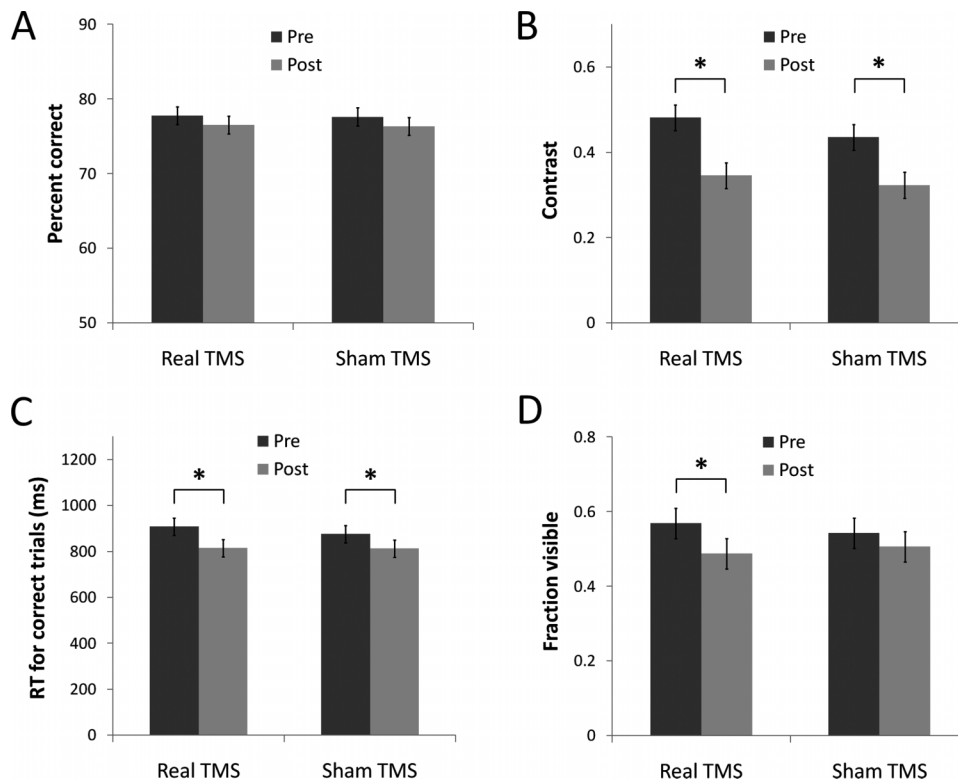
Figure 3. Task performance. (A) Percent correct. Percent correct was controlled by titration of stimulus contrast, such that stimulus judgments were about 75% correct throughout the experiment (see "Method"). Therefore the lack of any significant effects on these values is trivial. (B) Mean stimulus contrast. Stimulus contrast was determined online by the computer program (see "Method"), such that if subjects performed better than 75% correct, the contrast was reduced, and if subjects performed worse than 75% correct, the contrast was increased. There was a main effect of time on contrast ($p < .001$), indicating a perceptual learning effect; had the computer not been programmed to adjust task difficulty online, subjects would have shown improved accuracy over time. However, perceptual learning was not affected by TMS (TMS × time interaction, $F = 0.73$). (C) Reaction time for correct responses. Perceptual learning was also evident in reaction time data. Subjects were quicker to make correct responses in the second half of the experiment (main effect of time, $p = .016$). However, again, this learning effect was not modulated by TMS (TMS × time interaction, $F = 0.79$). (D) Mean visibility ratings. Visibility ratings decreased over time ($p = .005$), but the TMS × time interaction on visibility was not significant ($p = .4$). See the discussion for caveats about the visibility rating analysis. "Real pre": performance level before real TMS. "Real post": after real TMS. "Sham pre": before sham TMS. "Sham post": after sham TMS. *$p < .05$. Error bars represent 1 *SEM*.

constant. The stimulus contrast generated by the performance staircasing algorithm reduced over time, $F(1, 19) = 88.06$, $MSE = 0.003$, $p < .001$, suggesting a perceptual learning effect: Over time, subjects required a lower level of contrast in order to maintain the same level of response accuracy. However, the TMS × time interaction was not significant, $F(1, 19) = 0.73$, $MSE = 0.004$, indicating that the TMS treatment had no effect on stimulus classification performance (Figure 3B). Likewise, reaction time for correct trials improved over time, $F(1, 19) = 7.04$, $MSE = 17863$, $p = .016$, but was not sensitive to TMS, $F(1, 19) = 0.80$, $MSE = 6064$ (Figure 3C).

Similarly, mean visibility ratings decreased over time, $F(1, 19) = 9.92$, $MSE = 0.008$, $p = .005$, but independently of the TMS manipulation, $F(1, 19) = 1.2$,

$MSE = 0.007$ (Figure 3D). We address this null finding more fully in the discussion.

As hypothesized, TMS significantly impaired metacognitive sensitivity. A TMS × time interaction was evident for the correlation between accuracy and visibility, phi, $F(1, 19) = 3.64$, $MSE = 0.002$, $p = .036$ (Figure 4A). Investigation of this interaction revealed that phi was lowered following real TMS, one-tailed paired t-test, $t(19) = 4.13$, $p < .001$, but not sham TMS, $t(19) = 0.77$.

The bias-free SDT measure of metacognitive sensitivity, meta-$d' - d'$, also exhibited a TMS × time interaction effect, $F(1, 19) = 5.51$, $MSE = 0.125$, $p = .015$ (Figure 4B). The difference between observed and ideal type 2 sensitivity decreased following real TMS, one-tailed paired t-test, $t(19) = -3.1$, $p = .006$, but not sham, $t(19) = 0.37$. Metacognitive sensitivity
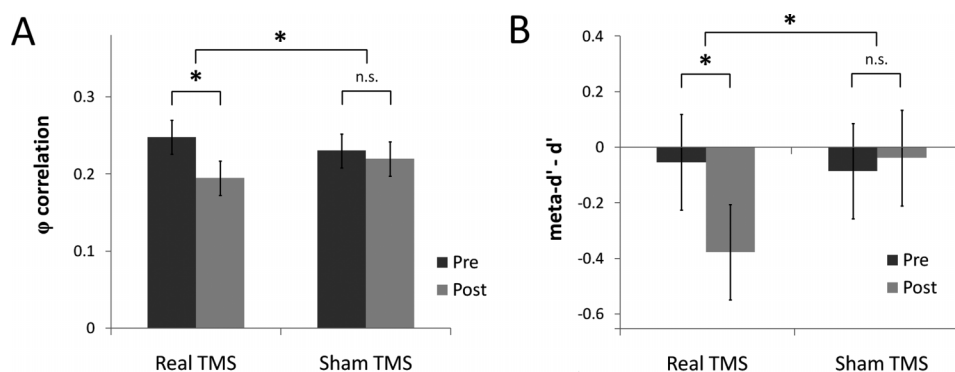
Figure 4. Effect of TMS on metacognitive sensitivity. (A) Correlation coefficient, phi. TMS significantly reduced phi, the correlation between stimulus classification accuracy and stimulus visibility. The effect of TMS is evident in a significant TMS × time interaction, $p = .036$; phi was lower following real TMS ($p < .001$) but not sham TMS ($p = .5$). (B) Divergence from optimal metacognitive sensitivity, meta-$d' – d'$. A signal detection theory analysis revealed that subjects' metacognitive sensitivity, relative to the optimal level of metacogntive sensitivity determined by their task performance (see Figure 2 and "Method"), was significantly impaired by TMS (interaction $p = .015$). Metacognitive sensitivity was lower following real TMS ($p = .006$) but not sham TMS ($p = .7$). Subjects exhibited significantly suboptimal metacognitive sensitivity following real TMS, i.e., meta-$d' – d' < 0$ ($p = .004$) but not in any other experimental condition (p values > .3). "Real pre": metacognitive performance before real TMS. "Real post": after real TMS. "Sham pre": before sham TMS. "Sham post": after sham TMS. *$p < .05$; n.s., not significant. Error bars represent 1 *SEM*.

was significantly suboptimal following real TMS, i.e. meta-$d' < d'$, one-tailed t-test, $t(19) = 2.93$, $p = .004$, but not in any other TMS × time condition (t values < 1).

There are several ways in which TMS could have impaired metacognitive sensitivity. One possibility is that TMS reduced visibility for correct trials, which would amount to a kind of relative blindsight (Lau & Passingham, 2006). Alternatively, TMS may

have increased visibility for incorrect trials, a kind of "hallucinatory" effect. A third possibility is that the reduction in metacognitive sensitivity was not specific to correct or incorrect trials. Thus, to better characterize the effect of TMS, we examined visibility ratings separately for correct and incorrect trials pre- and post-TMS (Figure 5A). We found a significant accuracy × time interaction, $F(1, 19) = 18.04$, $MSE = 0.002$, $p < .001$, driven by the fact that TMS
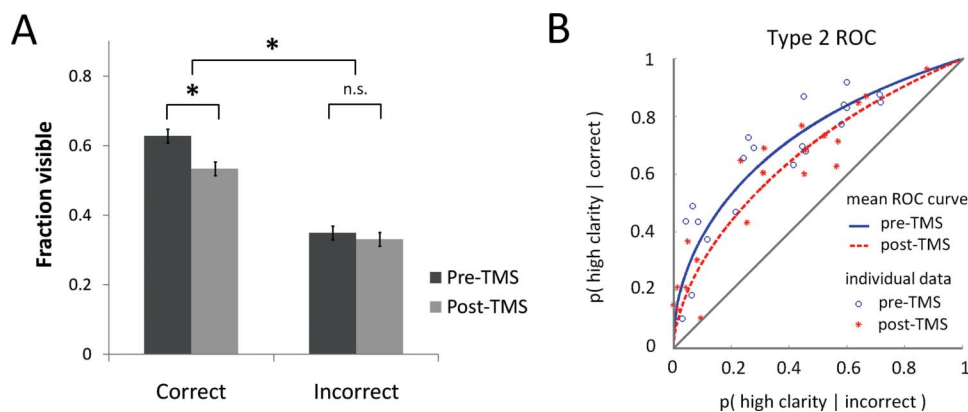

Figure 5. Nature of the TMS effect on metacognition. (A) Selective reduction of type 2 hit rate. Visibility ratings are displayed as a function of time (pre-/post-TMS) and accuracy (correct/incorrect) for the real TMS condition. TMS significantly reduced visibility for correct responses (two-tailed paired t-test, $p = .002$), but not for incorrect responses ($p = .5$). The time × accuracy interaction was significant, $p < .001$. These results suggest that TMS reduced metacognitive sensitivity (Figure 4) specifically by decreasing visibility ratings for correct responses (as opposed to increasing visibility ratings for incorrect responses). Thus, TMS induced a kind of relative blindsight, to the extent that TMS suppressed the reports of visibility for accurately processed stimuli. *$p < .005$; n.s., not significant. Error bars represent 1 *SEM*. (B) Type 2 ROC analysis. Individual data points indicate the type 2 hit rates and false alarm rates for every subject pre- and post-TMS. Type 2 ROC curves were estimated for each subject using estimates of meta-d', c', and s; the average of these ROC curves is plotted for the pre- and post-real TMS conditions. The distribution of individual data points and the fitted ROC curves indicate that TMS influenced metacognitive sensitivity rather than just response bias. Note that the ROC data is a reflection of meta-d', and thus is not as sensitive to the effect of TMS as the measure used in the analysis, meta-$d' – d'$ (Figure 4), since some variation in meta-d' is attributable merely to variation in d' (Figure 2).

reduced visibility for correct responses, two-tailed paired *t*-test, $t(19) = 3.54$, $p = .002$, but not incorrect responses, $t(19) = 0.72$. Thus, TMS impaired metacognitive sensitivity by selectively reducing the visibility of correctly classified stimuli.

DISCUSSION

Our results show that theta-burst TMS applied to bilateral DLPFC can reduce metacognitive sensitivity, i.e. the efficacy with which subjective visibility ratings distinguish between correct and incorrect stimulus judgments. This effect was driven specifically by a reduction in visibility for correct trials, rather than by a specific elevation of visibility for incorrect trials or by a nonspecific effect. In this sense, the direction of the effect is reminiscent of blindsight (Weiskrantz, 1997), where patients deny visual awareness even when they can perform visual discrimination tasks well above chance level. The effect of TMS was specific to metacognitive sensitivity; TMS did not disrupt stimulus classification performance, as measured by contrast level (Figure 3B) and reaction time for correct trials (Figure 3C).

We did not find a significant effect of TMS on averaged stimulus visibility itself. However, note that the effect of TMS is at least partially characterized by a change in visibility ratings, in that TMS reduced metacognitive sensitivity precisely by reducing visibility for correctly classified stimuli while leaving visibility for incorrectly classified stimuli unaffected (Figure 5A). Indeed, although the interaction was not significant, separate paired *t*-tests show a difference in visibility pre and post real TMS, two-tailed, $t(19) = 3.09$, $p = .002$, but no difference pre and post sham TMS, $t(19) = 1.47$. There are two reasons why the design of the current study may not have been ideal to statistically detect an effect of TMS on overall stimulus visibility. One is that stimulus visibility was affected by an experimental factor other than TMS, namely the contrast levels of the stimuli, which were adjusted online throughout the experiment in order to hold discrimination performance constant. Another reason is that subjects were instructed to use visibility ratings in a relative manner, in order to distinguish stimuli that were relatively more or less visible. The instruction to rate visibility in this relative way may have obscured the extent to which visibility ratings reflected absolute differences in stimulus visibility across experimental conditions. Nonetheless, these limitations are not important for the main focus of this study, which is the metacognitive sensitivity of visibility ratings.

One typical argument against studies of awareness is that the manipulation in question might have only changed subjects' criteria for producing subjective ratings, rather than changing awareness *per se*. A change in response criterion is not necessarily uninteresting, but, more importantly, this is not what we found. Our type 2 SDT analysis demonstrates that TMS reduced metacognitive sensitivity (i.e. the efficacy with which subjective visibility ratings discriminate between correct and incorrect judgments), rather than merely affecting metacognitive response bias (i.e. the overall propensity to give high visibility ratings). TMS reduced visibility for correct trials (type 2 HR) but not for incorrect trials (type 2 FAR) (Figure 5A), a pattern that cannot be attributed solely to changes in response bias. Likewise, our measure of type 2 sensitivity, meta-d' – d', is not sensitive to changes in type 2 response bias (Figure 2). We also demonstrate this point graphically in Figure 5B, which shows the type 2 ROC points for each subject, and mean fitted ROC curves, pre- and post-TMS. The distributions of type 2 ROC points and the fitted type 2 ROC curves differ, indicating lower type 2 sensitivity following TMS. If TMS only affected subjects' criteria for reporting high visibility, the type 2 ROC curves pre- and post-TMS should overlap (Macmillan & Creelman, 2005), contrary to our findings.

Because we only used visual stimuli, we cannot rule out the interesting possibility that the observed deficit in metacognitive sensitivity following TMS to DLPFC would apply to tasks in other modalities, e.g. auditory tasks. It may be that the observed reduction in metacognitive sensitivity is not specific to visual processes *per se*, but rather generalizes to other decision-making contexts involving confidence or perceptual clarity judgments. However, note that our analysis rules out the possibility that such a nonspecific effect could be carried by general influences on, for example, risk aversion or overall confidence level. Such differences would constitute differences in type 2 response bias, not type 2 sensitivity.

Our results extend previous work. Similarly to the present study, Del Cul, Dehaene, Reyes, Bravo, and Slachevsky (2009) showed that prefrontal lesions can affect subjective reports of visual experience more than visual task performance. Slachevsky et al. (2001, 2003) have shown that lesion to the prefrontal cortex can affect awareness in the monitoring of actions or sensory-motor readjustments. Other studies show that visual processing can be affected by lesion (Latto & Cowey, 1971) or TMS (Grosbras & Paus, 2003; Ruff et al., 2006) to the frontal eye field. Turatto, Sandrini, and Miniussi (2004) showed that TMS to the DLPFC can affect performance in change blindness. These

studies show that, contrary to what critics have argued (Pollen, 1995), disruption of activity in the prefrontal cortex can in fact influence awareness and visual processing. What is new in the present study is that it specifically highlights the role of the prefrontal cortex in supporting the metacognitive sensitivity of visual awareness.

The prefrontal cortex is associated with many important cognitive functions, and therefore our interpretation is not that it is completely specific to the metacognitive sensitivity of visual awareness. It is likely that bilateral theta-burst TMS to the DLPFC would impair performance in other tasks where metacognitive visual awareness is not required. For instance, as mentioned above, we think it is likely that it may also impair metacognitive awareness for auditory signals. Instead of applying the same TMS treatment to unrelated control tasks and hoping to show a negative result in those situations, we show that TMS impaired a specific *process* involved in our task, namely metacognitive awareness, but not other processes involved in the same task. It is important to note that performance in the stimulus classification task was not influenced by TMS under the stimulation parameters currently used. Thus, it is unlikely that TMS affected metacognitive sensitivity by means of nonspecific disturbances such as reductions in visual attention or general arousal.

As in Del Cul et al. (2009), one limitation of the present study is that we did not show that a similar effect could not be obtained in a control anatomical site. The lack of such control conditions is unfortunate and largely constrained by logistics (e.g., we did not have ethical approval for every brain region for this relatively new TMS protocol, and the leading authors have since relocated). However, given that the TMS was applied offline (i.e., not during task), and that the effect did not change basic task performance, it is unlikely that the results we obtained were due to the general distraction due to TMS. It is likely that TMS applied to an unrelated region, such as the somatosensory area, would not lead to our metacognitive effect. However, it remains an open question whether TMS applied to parietal areas that are connected to DLPFC would lead to similar results.

In any case, our conclusion is not that the neural circuitry that supports metacognitive visual awareness is completely localized in the DLPFC. Rather, we conclude that disruption of activity in this area can impair the metacognitive sensitivity of visual awareness. The present results show that the prefrontal cortex is functionally relevant to visual awareness, in that manipulation of the former can affect the latter.

Further, the data clarify in what way the prefrontal cortex might contribute. Activity in the DLPFC may play a relatively unimportant role in representing the visual signal itself, but it may be essential for some form of internal uncertainty monitoring that allows observers to be able to distinguish when visual processing is effective and when it is not. It is this introspective and metacognitive aspect of visual awareness for which the prefrontal cortex may be critical.

REFERENCES

Arfeller, C., Vonthein, R., Plontke, S. K., & Plewnia, C. (2009). Efficacy and safety of bilateral continuous theta burst stimulation (cTBS) for the treatment of chronic tinnitus: Design of a three-armed randomized controlled trial. *Trials, 10*, 74.

Arnold, D. H., Law, P., & Wallis, T. S. A. (2008). Binocular switch suppression: A new method for persistently rendering the visible 'invisible'. *Vision Research, 48*(8), 994–1001.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*(6527), 121–123.

Dehaene, S., Sergent, C., & Changeux, J. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the United States of America, 100*(14), 8520–8525.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain, 132*(9), 2531–2540.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*(5), 735–755; discussion, 755–808.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10*(4), 843–876.

Grosbras, M., & Paus, T. (2003). Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. *European Journal of Neuroscience, 18*(11), 3121–3126.

Grossheinrich, N., Rau, A., Pogarell, O., Hennig-Fast, K., Reinl, M., Karch, S., et al. (2009). Theta burst stimulation of the prefrontal cortex: Safety and impact on cognition, mood, and resting electroencephalogram. *Biological Psychiatry, 65*(9), 778–784.

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

Huang, Y., Rothwell, J. C., Edwards, M. J., & Chen, R. (2008). Effect of physiological activity on an NMDA-dependent form of cortical plasticity in human. *Cerebral Cortex, 18*(3), 563–570.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377*(6547), 336–338.

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, *18*(1), 64–71.

Latto, R., & Cowey, A. (1971). Visual field defects after frontal eye-field lesions in monkeys. *Brain Research*, *30*(1), 1–24.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18763–18768.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Marois, R., Yi, D., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron*, *41*(3), 465–472.

Mottaghy, F. M., Gangitano, M., Sparing, R., Krause, B. J., & Pascual-Leone, A. (2002). Segregation of areas related to visual working memory in the prefrontal cortex revealed by rTMS. *Cerebral Cortex*, *12*(4), 369–375.

Münchau, A., Bloem, B. R., Irlbacher, K., Trimble, M. R., & Rothwell, J. C. (2002). Functional connectivity of human premotor and motor cortex explored with repetitive transcranial magnetic stimulation. *Journal of Neuroscience*, *22*(2), 554–561.

Pascual-Leone, A., Valls-Solé, J., Wassermann, E. M., & Hallett, M. (1994). Responses to rapid-rate transcranial magnetic stimulation of the human motor cortex. *Brain*, *117*(4), 847–858.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261.

Pollen, D. A. (1995). Cortical areas in visual awareness. *Nature*, *377*(6547), 293–295.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, *3*(4), 261–270.

Rounis, E., Stephan, K. E., Lee, L., Siebner, H. R., Pesenti, A., Friston, K. J., et al. (2006). Acute changes in frontoparietal activity after repetitive transcranial magnetic stimulation over the dorsolateral prefrontal cortex in a cued reaction time task. *Journal of Neuroscience*, *26*(38), 9629–9638.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J., et al. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Current Biology*, *16*(15), 1479–1488.

Sahraie, A., Weiskrantz, L., Barbur, J. L., Simmons, A., Williams, S. C., & Brammer, M. J. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(17), 9406–9411.

Slachevsky, A., Pillon, B., Fourneret, P., Pradat-Diehl, P., Jeannerod, M., & Dubois, B. (2001). Preserved adjustment but impaired awareness in a sensory-motor conflict following prefrontal lesions. *Journal of Cognitive Neuroscience*, *13*(3), 332–340.

Slachevsky, A., Pillon, B., Fourneret, P., Renié, L., Levy, R., Jeannerod, M., et al. (2003). The prefrontal cortex and conscious monitoring of action: An experimental study. *Neuropsychologia*, *41*(6), 655–665.

Stefan, K., Gentner, R., Zeller, D., Dang, S., & Classen, J. (2008). Theta-burst stimulation: Remote physiological and local behavioral after-effects. *NeuroImage*, *40*(1), 265–274.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*(1), 100–117.

Szczepanowski, R., & Pessoa, L. (2007). Fear perception: Can objective and subjective awareness measures be dissociated? *Journal of Vision*, *7*(4), 10.

Talelli, P., Greenwood, R. J., & Rothwell, J. C. (2007). Exploring theta burst stimulation as an intervention to improve motor recovery in chronic stroke. *Clinical Neurophysiology*, *118*(2), 333–342.

Turatto, M., Sandrini, M., & Miniussi, C. (2004). The role of the right dorsolateral prefrontal cortex in visual change awareness. *NeuroReport*, *15*(16), 2549–2452.

Weiskrantz, L. (1997). *Consciousness lost and found: A neuropsychological exploration*. New York: Oxford University Press.