J Neurophysiol 107: 1556–1563, 2012.
First published December 14, 2011; doi:10.1152/jn.00985.2011.

Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence

Dobromir A. Rahnev,[1] **Brian Maniscalco,**[1] **Bruce Luber,**[2] **Hakwan Lau,**[1,3] **and Sarah H. Lisanby**[2]

[1]*Columbia University, Department of Psychology, New York, New York;* [2]*Duke University, Department of Psychiatry, Durham, North Carolina; and* [3]*Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Netherlands*

Submitted 28 October 2011; accepted in final form 11 December 2011

Rahnev DA, Maniscalco B, Luber B, Lau H, Lisanby SH. Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J Neurophysiol* 107: 1556–1563, 2012. First published December 14, 2011; doi:10.1152/jn.00985.2011.—The relationship between accuracy and confidence in psychophysical tasks traditionally has been assumed to be mainly positive, i.e., the two typically increase or decrease together. However, recent studies have reported examples of exceptions, where confidence and accuracy dissociate from each other. Explanations for such dissociations often involve dual-channel models, in which a cortical channel contributes to both accuracy and confidence, whereas a subcortical channel only contributes to accuracy. Here, we show that a single-channel model derived from signal detection theory (SDT) can also account for such dissociations. We applied transcranial magnetic stimulation (TMS) to the occipital cortex to disrupt the internal representation of a visual stimulus. The results showed that consistent with previous research, occipital TMS decreased accuracy. However, counterintuitively, it also led to an increase in confidence ratings. The data were predicted well by a single-channel SDT model, which posits that occipital TMS increased the variance of the internal stimulus distributions. A formal model comparison analysis that used information theoretic methods confirmed that this model was preferred over single-channel models, in which occipital TMS changed the signal strength or dual-channel models, which assume two different processing routes. Thus our results show that dissociations between accuracy and confidence can, at least in some cases, be accounted for by a single-channel model.

TMS; perception; vision; Bayesian inference; signal detection theory; modeling

IN MOST PSYCHOPHYSICAL TASKS, there is typically a positive association between confidence and accuracy (Busey et al. 2000; Fleet et al. 1987). When a task is easier, subjects usually perform better and give higher confidence for their responses (Pleskac and Busemeyer 2010). Nevertheless, there have also been several studies that reported dissociations between accuracy and confidence. For example, Lau and Passingham (2006) used metacontrast masking and found that two different stimulus-onset asynchronies, between the stimulus and the prime subjects, had the same level of accuracy but gave significantly different confidence ratings. Dissociation between accuracy and confidence has also been demonstrated in patients with lesions of the primary visual cortex, which resulted in a condition known as "blindsight" (Weiskrantz 1986).

In a recent psychophysics study (Rahnev et al. 2011), we showed that spatial attention boosted visual discrimination accuracy but lowered the corresponding visibility ratings (akin to confidence). These data, although counterintuitive, were nicely fit by a model based on signal detection theory (SDT) (Green and Swets 1966). According to SDT, high-confidence/visibility ratings are given when the internal signal exceeds a criterion set on the standard decision axis (Fig. 1*A*). The critical assumption of the model was that lack of attention increased the trial-by-trial variability of the internal signal, thus increasing the spread of the internal distributions. Due to the increased variability, the signal from unattended targets exceeded the criteria for high confidence more often, thus leading to higher overall confidence/visibility ratings for unattended stimuli (Fig. 1*B*). Interestingly, the model can easily generalize from attention to other processes that change the internal variability of the signal. According to the model, if the criteria for high confidence remain constant, an increase in signal variability should lead to a decrease in discrimination accuracy and an increase in confidence (Fig. 1*B*). To ensure that subjects do not consciously adjust their criteria to compensate for the change in signal variability, ideally, the difference between the conditions of interest should be subtle and mainly focused on increasing variability in one of the conditions.

Although transcranial magnetic stimulation (TMS) to the visual cortex is often used to completely "knock out" conscious perception (e.g., Boyer et al. 2005; Breitmeyer et al. 2004; Kastner et al. 1998; Koivisto et al. 2010, 2011; Ro et al. 2004), some recent studies have shown that low-intensity stimulation can effectively inject noise to the visual system (Ruzzoli et al. 2010; Schwarzkopf et al. 2011; although, see Harris et al. 2008; Ruzzoli et al. 2011). In the present study, we applied single-pulse TMS at an intensity below the threshold for the conscious perception of phosphenes. We carefully adjusted the stimulus location on the screen to place the small target within the TMS-influenced region (as reflected by phosphene perception at higher intensity). To anticipate, we found that single-pulse TMS to occipital cortex decreased discrimination accuracy but nevertheless, increased confidence ratings. Confirming our model (Rahnev et al. 2011), computational modeling showed that TMS seemed to influence mainly the variability but not the mean signal intensity of the visual percept.

MATERIALS AND METHODS

Subjects. Eleven subjects participated in the experiment. Four of them did not perceive phosphenes under high-intensity occipital stimulation with TMS. Because our task procedure critically depended on identifying the location of the effects of TMS within the visual field (see *Stimuli and task* below), these subjects were excluded from the analysis. One subject performed the task at chance level and was

---

Address for reprint requests and other correspondence: D. A. Rahnev, Columbia Univ., Dept. of Psychology, 406 Schermerhorn Hall, 1190 Amsterdam Ave., MC 5501, New York, NY 10027 (e-mail: drahnev@gmail.com).

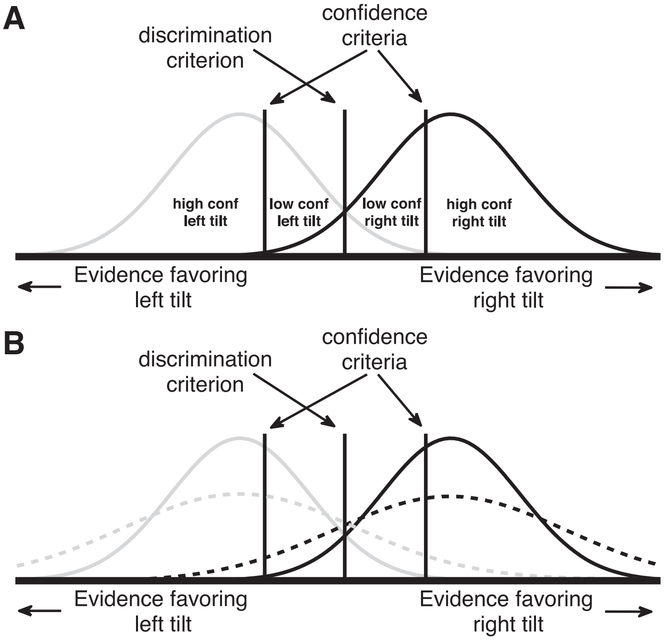

Fig. 1. Discrimination between 2 classes of stimuli within the framework of signal detection theory (SDT). *A*: according to SDT, left- and right-tilted bars produce Gaussian distributions of internal signals along a single dimension. A discrimination criterion is used to distinguish between these 2 classes of stimuli. Confidence (conf) responses are given based on additional confidence criteria. In our task, subjects needed to place 2 such criteria. The 4 types of responses would then naturally fall between consecutive confidence and discrimination criteria. *B*: transcranial magnetic stimulation (TMS) to the occipital cortex may increase the noise in the distributions (see dashed distributions). This can lead to a decrease in d'—a measurement of subjects' capacity to discriminate the orientation of the bar—and a simultaneous increase in the number of trials judged with high confidence.

also excluded from the analyses. This left us with six subjects (four female; all subjects 19–40 years old). Subjects were required to have normal or corrected-to-normal vision. All subjects were screened with psychiatric, physical, and neurological examinations, urine drug screens, and pregnancy tests for women of childbearing capacity. Potential subjects were excluded before the experiment if they had a history of current or past Axis I psychiatric disorder (including substance abuse/dependence), as determined by the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Axis I Disorders–Nonpatient, a history of neurological disease, or seizure risk factors. Subjects received detailed information about the potential side-effects of TMS and signed an informed consent statement approved by the New York State Psychiatric Institute Institutional Review Board.

Stimuli and task. Subjects were seated in a dark room, ~100 cm away from a computer monitor. They were required to fixate on a small white cross for the duration of the experiment (Fig. 2). The task was to indicate whether a small bar (0.2° visual angle) was tilted 45° to the left or to the right. The bar was presented for 33 ms (two computer display refresh frames) and appeared just below the fixation cross. The exact location varied from subject to subject, as we placed the bar at the location in the visual field where each subject experienced phosphenes from TMS. The small size of the bar was used to ensure that the stimulus fell completely within the stimulated region of the visual cortex (Boyer et al. 2005) (see description of TMS methods in *TMS session* below). Stimuli were generated using Psychophysics Toolbox (Brainard 1997) in MATLAB (MathWorks, Natick, MA) and were shown on a MacBook (13-in. monitor size, 1,200 × 800 pixel resolution, 60 Hz refresh rate).

The fixation cross changed from white to black during the period when the bar was present. This change was the cue to subjects that the stimulus had appeared and that they needed to respond, guessing if necessary. Subjects responded by pressing one of four keyboard buttons that indicated both their decision about the tilt of the bar (left/right), as well their confidence in their decision (high/low). We did not use a more-graded scale for confidence, to keep the task as simple as possible and because a continuous confidence scale would have complicated our signal detection theoretic modeling (see *Model specifications* below). Before the start of the experiment, subjects were encouraged to use both levels of confidence as evenly as possible. This instruction was not repeated once the experiment started, to avoid subjects deliberately adjusting their confidence ratings by using cognitive strategies rather than perceptual processes. Trial duration during the TMS experiment was set at 5 s. We chose this relatively long interval to ensure that the TMS effect of the previous pulse had largely subsided by the time the next pulse was delivered, so that cumulative effects of TMS over the course of a block would be minimized.

In an initial training session, subjects practiced with the task over the course of 558 trials. This session occurred on a different day and in a separate room than the subsequent TMS experiment. In the training session, each trial began 500 ms after the response to the previous trial was given. Based on the training data, we determined for each subject three levels of luminance contrast for the tilted bar. A medium contrast level was chosen so that d'—a measurement of subjects' capacity to discriminate the orientation of the bar—would be between one and two (roughly corresponding to 70–85% correct responses). The lowest contrast was then fixed at 75% of the medium contrast, and the higher contrast was fixed at 125% of the medium contrast. We used three contrast levels, because such variability in stimulus strength encourages subjects to use both sides of the confidence scale.

In the TMS session, the contrast of the bar was chosen pseudorandomly in each trial so that the total number of presentations with each contrast was identical. For one subject (the first subject in our experiment), only the medium contrast level was used. Subjects were not informed explicitly about the presence of multiple contrast levels.

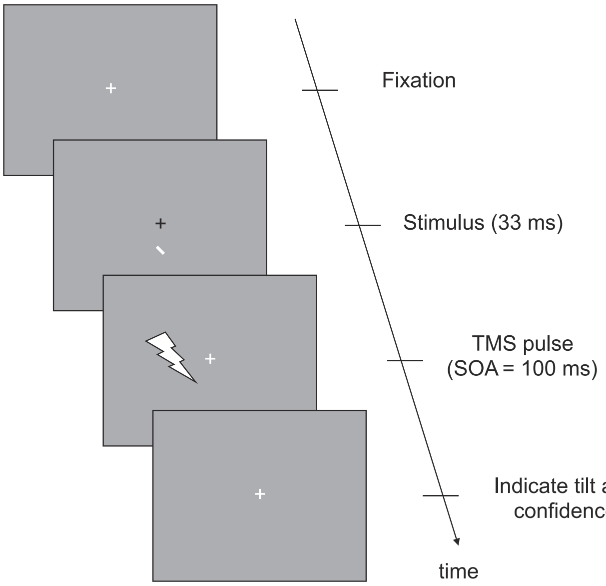

Fig. 2. Experimental task. The stimulus was a small white bar presented below fixation (exact location was chosen for each subject according to his or her phosphene location). The TMS pulse was delivered 100 ms after the onset of the stimulus. Throughout the experiment, subjects were asked to fixate on a small white cross that changed color to black when the stimulus appeared. Subjects gave their response with a single button press, indicating both the tilt of the bar (left/right) and their confidence (high/low). The duration of each trial was exactly 5 s. SOA, stimulus-onset asynchronies.

TMS session. Before the start of the TMS experiment, subjects were given 20 trials of practice with the task without TMS. This was done to familiarize the subjects with the TMS environment and to make sure that they remembered all of the details of the task.

TMS was delivered with a figure-eight coil (9 cm diameter), powered by a Magstim 200 stimulator (Magstim, Whitland, UK). This coil provides weaker but more focused stimulation, compared with the round coil, which was used in many previous studies that found complete suppression of visual stimuli (Amassian et al. 1989; Boyer et al. 2005; Corthout et al. 1999).

The precise location and intensity of the TMS were determined just prior to the beginning of the main experiment. We first used a hunting procedure to determine the optimal location for stimulation on the occipital cortex. We started by explaining to the subjects what phosphenes are and how they could determine if their subjective experiences would be considered to be phosphenes. We then delivered a pulse of 70% of the maximal stimulator output to a location ~2 cm above the inion. The main axis of the coil was oriented parallel to the sagittal plane and the coil handle extended ventrally. If subjects failed to perceive a phosphene, we stimulated around the initial site. If subjects still did not perceive any phosphenes, we increased the stimulation intensity by 10% and repeated the procedure. As mentioned above, four subjects did not perceive phosphenes, even at 100% of stimulator output, and were subsequently excluded from the experiment.

Once a subject perceived phosphenes, we moved the coil in 2-cm steps in lateral, medial, inferior, and superior directions until we found the place on the occipital cortex that resulted in the strongest and clearest phosphenes. The spot was then marked on the head, and the rest of the experiment was done using that location.

We then proceeded to determine each individual's phosphene threshold in the following manner. Starting at 40% of the maximum stimulator output, we delivered three TMS pulses to the predetermined location on the occipital cortex. If the subject did not perceive a phosphene, we increased the intensity by 5%. If the subject perceived a phosphene at least once, we continued to deliver TMS pulses until either five positive or five negative responses were given. If we received five negative responses first, we again increased the intensity by 5% and repeated the procedure. Once five positive responses were provided at a given intensity, that intensity level was chosen as the subject's "phosphene threshold." Afterward, all stimulation was delivered at 80% of the individual phosphene threshold. The resulting mean intensity of stimulation during the experiment was 52.5% (SD = 5.3%) of maximum stimulator output. We used jumps of 5% to minimize the time to find the phosphene threshold as in previous studies (Koivisto et al. 2010). However, it is important to note that this procedure may have slightly overestimated the phosphene threshold with up to 4%. Nevertheless, even with such overestimation, our final stimulation intensity was lower than 90% of the real phosphene threshold. No subject reported seeing phosphenes or any other visual phenomena during the experiment; in fact, no subject noticed a difference in his or her perception of the visual stimuli. The intensity of TMS to the control site (vertex) was always the same as for the occipital cortex. No leg movement was elicited by vertex stimulation in any of the subjects.

The first subject completed 200 trials, separated into two blocks, terminating the experiment after two blocks due to fatigue. The second subject completed 600 trials, separated in six blocks, taking almost 3 h to complete the experiment. The data from these two subjects did not differ in any systematic way from the data from the rest of the subjects. Subsequently, we revised the total trial number and block size to prevent fatigue and excessive experiment duration. All subsequent subjects completed 468 trials, separated into six blocks of 78 trials.

The TMS pulse was always delivered 100 ms after the onset of the bar stimulus. This interval was chosen based on a number of previous studies that found that TMS had maximum effect on visual perception between 80 and 120 ms after stimulus onset (Amassian et al. 1989; Boyer et al. 2005; Corthout et al. 1999; Kammer 2007; Kastner et al. 1998; Luber et al. 2007; Maccabee et al. 1991; Miller et al. 1996). In one-half of the blocks, subjects received a TMS pulse to the occipital cortex; in the other one-half of the blocks, the subjects received a pulse at the same stimulus intensity to the vertex of the head. The vertex was an active control site, chosen because TMS applied there was not expected to interfere with the visual task but controlled for the ancillary effects of TMS, including startle, acoustic artifact, and somatosensory sensation. The order of the occipital and vertex TMS blocks was counterbalanced across subjects.

After the experiment, the subjects were debriefed about the purpose of the study and asked about the side-effects of TMS. No subject reported more than mild discomfort from the TMS, and the discomfort did not continue after the end of the experiment. Visual analog scales were used to assess the subjects' emotional state before and after TMS, and no significant differences were found.

Data analysis. The signal detection theoretic measure d' was calculated by first coding each trial as a hit, miss, false alarm, or a correct rejection. Trials, in which subjects reported that the stimulus was right tilted, were coded as hits if the bar was indeed right tilted and as false alarms if the bar was left tilted. Trials, in which subjects reported that the stimulus was left tilted, were coded as misses if the bar was right tilted and as correct rejections if the bar was left tilted. Hit rate (HR) was computed as hits/(hits + misses) and false-alarm rate (FAR) was computed as false alarms/(false alarms + correct rejections). Finally, d' = z(HR) − z(FAR), where z is the inverse of the cumulative standard normal distribution that transforms HR and FAR into z scores.

Computational modeling assumptions. We fit the behavioral results using five different models. All models were based on SDT. Two of them stipulated different channels for confidence ratings and stimulus discrimination. The other three were based on a single detection theoretic process.

In each model, we made standard assumptions: *1*) the two stimuli used in the experiment gave rise to internal signals normally distributed along some decision axis; *2*) perceptual decisions were made by comparing the signal on the decision axis with a criterion; *3*) confidence judgments were made by comparing the signal on the decision axis with multiple criteria, corresponding to the confidence ratings available to subjects in this experiment (Fig. 1*A*); and *4*) criteria for perceptual decisions and confidence ratings were set in the same way for occipital and vertex TMS. The last assumption is justified for several reasons. First, none of the subjects reported that they had consciously perceived the task during occipital TMS to be any harder than during vertex stimulation. The relatively small difference in capacity that we measured (see RESULTS) also suggests that two conditions were not noticeably different to the subjects. Finally, previous research (Gorea and Sagi 2000) has demonstrated that when lower and higher visibility stimuli are presented together in a block, subjects tend to use a single set of criteria for both. In the current experiment, the accuracy of differences between the occipital and vertex TMS trials was closer than those found between conditions in Gorea and Sagi's experiment, which makes it even less likely that subjects were able to shift their confidence criteria between them.

Model specifications. The first three models postulated a single channel with two Gaussian distributions (corresponding to left- and right-tilted bars) with the difference in means (μ) and SDs (σ).

In *Model 1*, we allowed occipital TMS to affect only the distance (μ) between the Gaussian distributions. The SD σ was set to an arbitrary constant value of 1; the numerical value of this parameter, on its own, is unimportant, because its contribution is to be determined within the context of other parameters. The model had five free parameters: $\mu_{\text{occipital}}$, μ_{vertex}, and the location of each of the three criteria levels used for discrimination and confidence judgments (see Fig. 1*A*).

Model 2 tested if the primary effect of occipital TMS was inducing noise in the internal representations. It allowed occipital TMS to affect only σ with no influence on μ. The SD for vertex TMS (σ_{vertex}) was again set to 1. The model had five free parameters: μ (which had the same value for occipital and vertex TMS), $\sigma_{\text{occipital}}$, and the location of each of the three criteria levels used for discrimination and confidence judgments.

Model 3 allowed occipital TMS to affect both the signal and the noise of the internal representations. Thus in this model, occipital TMS changed both μ and σ. As in *Model 2*, the SD for vertex TMS (σ_{vertex}) was set to 1. Overall, the model included six free parameters: $\mu_{\text{occipital}}$, μ_{vertex}, $\sigma_{\text{occipital}}$, and the location of each of the three criteria levels used for discrimination and confidence judgments.

The last two models included "conscious" and "unconscious" channels. In each of the two channels, the left- and right-tilted bars gave rise to a Gaussian distribution. The distance between the centers of these Gaussian distributions were defined as μ_c and μ_u for the conscious and unconscious channels, respectively. Similarly, σ_c and σ_u were the SDs of the distributions for the conscious and unconscious channels, respectively.

Model 4 allowed occipital TMS to affect μ_c only. This is a model similar to previous accounts of the blindsight phenomenon (Boyer et al. 2005; Weiskrantz 1986), which postulated that lesions to the primary visual cortex may affect a conscious cortical channel but leave intact an unconscious subcortical channel. Without loss of generality, we set σ_c and σ_u equal to 1. This left the model with six free parameters: μ_u (which had the same value for occipital and vertex TMS), $\mu_{\text{c-occipital}}$, $\mu_{\text{c-vertex}}$, and the location of each of the three criteria levels used for discrimination and confidence judgments.

However, one can argue that it is possible that the unconscious channel was also affected by occipital TMS. Therefore, in *Model 5*, we allowed occipital TMS to affect the unconscious channel as well. The model was equivalent to the first model in all other respects. It had seven free parameters: the free parameter μ_u from *Model 4* was now modeled as two parameters: $\mu_{\text{u-occipital}}$ and $\mu_{\text{u-vertex}}$, whereas the remaining five parameters were the same as in *Model 4*.

Model fitting. We fit the models to the data using a maximum-likelihood estimation approach, which has previously been used within a signal detection framework (Dorfman and Alf 1969). Briefly, the likelihood of a set of SDT parameters given the observed data can be calculated using the multinomial model. Formally

$$L\left(\theta \middle| data\right) \propto \prod_{i,j} Prob_{\theta}\left(Resp_i \middle| Stim_j\right)^{n_{data}(Resp_i|Stim_j)}$$

where each $Resp_i$ is a behavioral response that a subject may produce on a given trial, and each $Stim_j$ is a type of stimulus that may be shown on a given trial. $Prob_{\theta}(Resp_i\ Stim_j)$ denotes the probability with which the subject produces the response $Resp_i$ after being presented with $Stim_j$, according to the signal detection model specified with parameters θ. $n_{data}(Resp_i\ Stim_j)$ is a count of how many times a subject actually produced $Resp_i$ after being shown $Stim_j$.

In the current study, the subjects had four possible responses from which to select (two stimulus classification options·two levels of confidence), and there were two stimulus types (left- or right-tilted bars).

Note that the models were not fit to summary statistics of performance, such as percent correct or average visibility. Rather, the models were fit to the full distribution of probabilities of each response type contingent on each stimulus type. From this full behavioral profile of stimulus-contingent response probabilities, we could derive various summary statistics.

We fit all models under consideration to the observed data by finding the maximum-likelihood parameter values θ. Maximum-likelihood fits were found using a simulated annealing procedure (Kirkpatrick et al. 1983). Model fitting was conducted separately for each subject's data.

Formal model comparison. The maximum likelihood associated with each model characterizes how well that model captures patterns in the empirical data. However, comparing models directly in terms of likelihood can be misleading; greater model complexity can allow for tighter fits to the data but can also lead to undesirable levels of overfitting, i.e., the erroneous modeling of random variation in the data. The Akaike Information Criterion (AIC), motivated by considerations from information theory, provides a means for comparing models on the basis of their maximum-likelihood fits to the data, while correcting for model complexity (Burnham and Anderson 2002). We used AICc, a variant of AIC, which corrects for finite sample sizes: $\text{AICc} = -2\cdot\log[\text{L}(\theta\ \text{data})] + 2\cdot\text{K}\cdot[\text{n}/(\text{n} - \text{K} - 1)]$, where K is the number of parameters in the model, and n is the number of observations being fit. Lower values of AICc are desirable, because they indicate a higher model likelihood and/or a lower model complexity (lower number of parameters).

We used the likelihood of each model, given the data, as a basis for model comparison: $\text{L}(\text{model}_i\ \text{data}) \propto e^{-1/2\cdot(\text{AICc\_i} - \text{AICc\_min})}$.

AICc_i is the AICc for model I, and AICc_min is the lowest AICc exhibited by all models under consideration. These model likelihoods can be scaled to sum to 1, and the resulting "Akaike weights" reveal the relative weight of evidence for each model as evaluated by its fit to the data, correcting for model complexity.

We replicated the above analysis using the Bayesian Information Criterion (BIC) in place of AICc, where $\text{BIC} = -2\cdot\log[\text{L}(\theta\ \text{data})] + \text{K}\cdot\log(\text{n})$.

The two different ways of performing model selection yielded similar results.

RESULTS

We first analyzed our data using repeated-measures ANOVA to test for the effects of TMS site and contrast separately. Since our first subject only had one contrast level, that subject was excluded from the ANOVA. With the use of the remaining five subjects, we first analyzed the accuracy data (percent correct) and found a significant main effect of TMS site [$F(1,4) = 14.01$, $P = 0.02$], a significant main effect of contrast [$F(2,3) = 10.31$, $P = 0.045$], and no interaction between TMS site and contrast [$F(2,3) = 0.62$, $P = 0.596$]. We then analyzed the confidence data and again found a significant main effect of TMS site [$F(1,4) = 10.16$, $P = 0.033$], a significant main effect of contrast [$F(2,3) = 13.89$, $P = 0.03$], and no interaction between TMS site and contrast [$F(2,3) = 0.93$, $P = 0.485$]. Thus occipital TMS has a significant influence on both accuracy and confidence. Since our main interest was in the effect of occipital vs. vertex TMS and since contrast did not interact with TMS site, we pooled across contrast levels for each subject. This pooling allowed us to analyze the data from all six subjects simultaneously.

When we pooled across the three contrast levels and included all subjects, a paired sample *t*-test showed that compared with vertex TMS, occipital TMS decreased discrimination capacity d' [$t(5) = 4.49$, $P = 0.006$]. The decrease was relatively small (0.16 on average, which corresponds to an ~2.5% decline in correct answers), but this was expected, based on our procedure and choice of low-intensity, localized stimulation (see MATERIALS AND METHODS). Nevertheless, the result was consistent: it was present in each of our six subjects (Fig. 3). Similarly, in line with this detection theoretic analysis, the proportion of correct answers was lower for occipital TMS than vertex [$t(5) = 4.37$, $P = 0.007$]. Also, occipital TMS did not bias subjects toward one stimulus type [right- or left-tilted bars; $t(5) = 0.63$, $P = 0.55$]. On the other hand, confidence

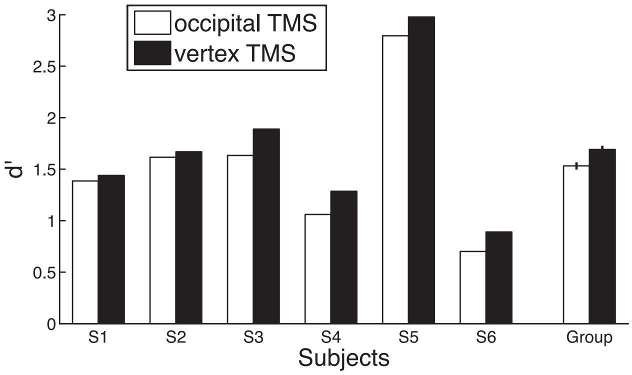

Fig. 3. Effects of occipital TMS on capacity d'. TMS to the occipital cortex decreased d' for each of the 6 subjects (S1–S6). The decrease was significant in the group as a whole ($P = 0.006$). The group means are plotted on the right side of the figure. The error bars represent the SE of the difference between the means.

ratings increased after occipital TMS compared with vertex stimulation [$t(5) = 3.94$, $P = 0.01$]. The effect was again consistent across our subjects, with five of them showing an increase and one showing no difference in confidence between the two conditions (Fig. 4).

Computational modeling results. To explain these somewhat counterintuitive results, we fit five different models to the data (see MATERIALS AND METHODS). The first three models postulated a single processing channel, in which occipital TMS is posited to affect the signal only (*Model 1*), the noise only (*Model 2*), or the signal and the noise (*Model 3*), respectively. The other two models were dual-channel models, which included conscious and unconscious channels that operate in parallel, with confidence rating being specifically dependent on the conscious channel. In *Model 4*, occipital TMS affected the signal of the conscious channel only, whereas in *Model 5*, occipital TMS affected the signal in both the conscious and unconscious channels.

Since the five models differed in complexity, we used AIC to compare the fit with observed data, while punishing models for extra complexity (number of parameters). The results showed that the best-fitting model was *Model 2*, in which occipital TMS affected only the variance of the internal representations (Fig. 5). The model's average Akaike weight was two times higher than any of the other models. These results were confirmed by using a slightly different metric for punishing the complexity of the models—BIC (see MATERIALS AND METHODS)—which gave similar results, with *Model 2* emerging as the preferred model. Thus our findings do not depend on the specific measure that we used to compare among the models.

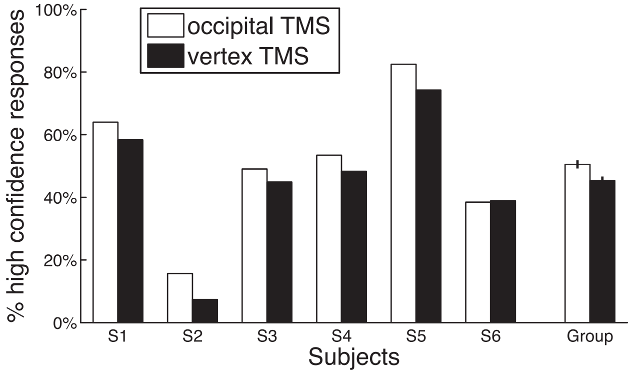

Fig. 4. Effects of occipital TMS on confidence ratings. Unlike capacity d', confidence ratings increased after occipital TMS. The effect was present for 5 of the 6 subjects and was significant in the group as a whole ($P = 0.01$). The group means are plotted on the right side of the figure. The error bars represent the SE of the difference between the means.

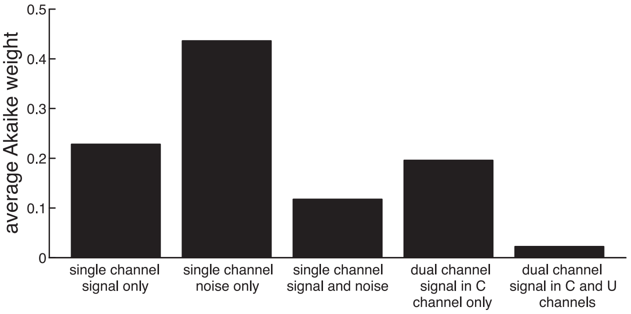

Fig. 5. Model selection results. Formal model comparison was conducted using Akaike Information Criterion, which rewards models for closely fitting, observed data, while punishing them for the degree of complexity. Higher average Akaike weight indicates that the corresponding model is more likely. *Models 1–5* are displayed in order from left to right and are indexed by a short description indicating whether the corresponding model postulates 1 or 2 processing channels and what parameters are affected by occipital TMS (C channel, "conscious" channel; U channels, "unconscious" channels; for more detail about each model, see MATERIALS AND METHODS). *Model 2*, which assumes that TMS only influenced the variability of the visual signal within a single process but not the average signal strength, was the clear winner, with average Akaike weights ~2× higher than the 2nd-best model. A complimentary analysis using the Bayesian Information Criterion produced similar results, with *Model 2* winning over the rest of the models.

We further investigated the parameter fits for each model. The best fit in *Model 2* occurred when occipital TMS increased the SD of the Gaussian distributions by ~12%. Furthermore, all models included three free parameters corresponding to the discrimination criterion and the two confidence criteria. The average discrimination criterion was close to zero for all models, which is in line with our finding that subjects were largely unbiased in discriminating between left- and right-tilted bars. The average values for the two confidence criteria were also virtually identical across the five models.

DISCUSSION

Summary of results. We delivered low-intensity TMS to the occipital cortex to disturb the normal visual processing of a small tilted bar. The manipulation was successful in decreasing subjects' ability to identify the tilt of the bar. Interestingly, occipital TMS also led to higher confidence ratings compared with stimulation to a control site (vertex). These counterintuitive results were parsimoniously explained by a signal detection model, in which TMS increased the variance of the underlying signal representation.

Site of stimulation within the visual cortex. It has sometimes been assumed that occipital TMS primarily targets the visual area V1 (Boyer et al. 2005; Laycock et al. 2007; Silvanto et al. 2005a, b). However, a recent study (Thielscher et al. 2010) found that the exact site of stimulation is likely to be V2d or even V3. In this study, we are largely agnostic with respect to the exact regions in the visual hierarchy that were stimulated, since the net effect of increased noise could be accomplished at several levels in the hierarchy.

Does TMS affect noise or mean signal intensity for visual perception? The current results, especially our computational modeling findings, suggest that single-pulse occipital TMS increases the noise of the internal representations in our discrimination task. This conclusion is in line with several studies that have argued that TMS acts by adding neural noise to the perceptual process rather than by affecting signal strength (Ruzzoli et al. 2010; Schwarzkopf et al. 2011). Nevertheless, two other studies argued for the opposite conclusion (Harris et al. 2008; Ruzzoli et al. 2011). This suggests that the precise influence of TMS may depend on the specific stimulation and task parameters. Note that the psychological noise increase observed in our study can be the result of complex neuronal influences that are not necessarily described well as simple addition of random noise to neuronal firing. For example, psychological noise can be the result of TMS suppressing activity of certain subpopulations of neurons, while increasing the activity in another (Siebner et al. 2009; Silvanto and Muggleton 2008). The effects of TMS are likely to be complex and depend on specific stimulus and TMS parameters. A more systematic approach is needed to establish the precise effect of TMS for neuronal and psychological processes in different task contexts and under various stimulation procedures. Because of the complexity of this issue, our conclusion that TMS induced noise is not meant to be generalized to all other TMS studies. It is restricted to this study, in which we focus on the trial-by-trial variability of the internal perceptual signal.

Previous TMS studies that investigated confidence. Several previous studies found that suprathreshold occipital TMS had detrimental effects on both confidence and visibility ratings. For example, Boyer et al. (2005) found that TMS to the occipital cortex led to a substantial amount of trials, in which subjects were unaware of the orientation (61%) or the color (70%) of the stimulus. Nevertheless, in these trials, subjects performed better than chance in guessing the orientation or color, respectively. Although this study suggests that objective performance may go beyond what subjects experienced subjectively after TMS, the different methodology makes it hard to directly compare the results of Boyer et al. (2005) with ours. In particular, the purpose of their study was to abolish the conscious percept, and they used stronger stimulation intensities and a circular coil, which typically lead to more-intense but less-focal effects. Our study, on the other hand, aimed to deliver localized and low-intensity stimulation by using a figure-eight coil and intensity below the threshold for phosphene perception. And yet, interestingly, when only the low-confidence trials in our study are considered, our subjects also performed better than chance (66% correct responses), which parallels the results of Boyer et al. (2005).

In another study, Koivisto et al. (2011) found that suprathreshold TMS led to a decrease in both accuracy and subjective awareness, the latter of which is usually considered to be akin to confidence ratings (Szczepanowski and Pessoa 2007). However, in that study, the researchers achieved a much higher level of suppression—accuracy decreased by ~20% compared with the no-TMS condition. Our single-channel SDT explanation is likely to be restricted to small changes in variance, which are not subjectively noticeable (such that subjects do not consciously adjust their criteria for giving high-confidence responses). Thus it is perhaps not surprising that our model would not apply to the results of Koivisto et al. (2011), where one may expect that the large decrease in accuracy led subjects to consciously adjust their confidence ratings. In comparison, in our experiment, accuracy only decreased by 2.5%.

Finally, Koivisto et al. (2010) used subthreshold TMS but still found decreases in both accuracy and visibility ratings with occipital TMS. However, in that study, the authors used motion stimuli, for which processing depends on a different visual area [middle temporal (MT)/V5]. A further difference with our study was the fact that the motion stimulus was several times larger than the small bar that we used, which was designed to fall completely within the region where subjects perceived phosphene upon TMS. Finally, it is possible that despite their strong association (Szczepanowski and Pessoa 2007), visibility and confidence ratings may dissociate in the context of occipital TMS.

Effect of subthreshold TMS on accuracy. As noted above, Koivisto et al. (2010) found that subthreshold TMS to the occipital cortex can lead to a decrease in accuracy. Nevertheless, two other studies (Abrahamyan et al. 2011; Schwarzkopf et al. 2011) found an increase in accuracy after subthreshold TMS. Unlike the present study, Schwarzkopf et al. (2011) used motion stimuli and targeted MT rather than the primary visual cortex and delivered a triple-pulse TMS (pulse gap of 50 ms). Abrahamyan et al. (2011) used much larger stimuli than in the current study (6.5° compared with 0.2° visual angle in our study; a difference of over 30 times) and used a two-interval, forced-choice detection task, which requires very different neural computation compared with our single-stimulus discrimination task (Macmillan and Creelman 2005). Overall, the differences among the current study and the previous studies (Abrahamyan et al. 2011; Koivisto et al. 2010; Schwarzkopf et al. 2011) prevent any conclusions about the general effects of subthreshold TMS on accuracy, as these effects likely depend on factors such as size and type of stimulus used, stimulation site, stimulation procedure, and precise stimulation intensity. Therefore, our data should not be considered as evidence that subthreshold TMS has a general detrimental effect on performance outside of our paradigm. More systematic research is needed to pinpoint the exact effects of subthreshold TMS and how they depend on the factors mentioned above.

Mechanisms for confidence ratings. Dissociations between accuracy and confidence have often been explained by dual-channel models (Del Cul et al. 2009; Jacoby 1991; Jolij and Lamme 2005; Morewedge and Kahneman 2010). In dual-channel models, typically, one channel supports conscious processing (this is often assumed to be a cortical channel that goes through the primary visual cortex), whereas the other is largely unconscious (a subcortical channel that bypasses the visual cortex). Thus within the context of experiments involving confidence ratings, only the conscious channel contributes to confidence ratings, whereas both channels may contribute to the perceptual decision. It is easy to see how dual-channel models can account for dissociations between confidence and accuracy. For example, the phenomenon blindsight (Weiskrantz 1986) could be due to a disruption in the conscious channel (thus resulting in a lack of conscious visual experience), whereas the remaining visual processing would be based on the unconscious channel that is left largely intact. Alternatively, for the data presented here, it could be that under TMS, the signal in the unconscious channel went down (leading to a decrease in accuracy), and that signal in the conscious channel went up (leading to an increase in confidence). We do not challenge that dual-channel models are plausible in some contexts. However, in the present study, we show that a single-channel model can more parsimoniously account for the data.

The fact that the data fit with a single-channel model does not mean that all relevant activity takes place within the visual cortex. Critical to this model is that high confidence is generated when the signal crosses the relevant criteria. In this study, perhaps because the intensity of TMS was weak and did not lead to supraliminal visual sensation (it was below phosphene threshold), the modeling results suggest that subjects did not adjust their confidence criteria. However, in other contexts, subjects may deliberately adjust their confidence criteria. We have argued elsewhere that such a criterion-setting mechanism may depend on the prefrontal cortex (Lau and Rosenthal 2011).

Could subjects have deliberately adjusted their confidence criteria? Our results show that occipital TMS led to lower accuracy but higher confidence. Since we instructed subjects to use both confidence ratings (high and low) as evenly as possible, one concern is whether this instruction may have encouraged subjects to deliberately adjust their criteria for high confidence in occipital and vertex TMS blocks. There are several reasons why we believe that this was not the case.

Importantly, all of our subjects reported that they did not notice a difference in the appearance of the stimuli in the occipital and vertex stimulation conditions. Many of them were even commenting that they were certain that occipital TMS did not have any influence on them. Thus it is likely that they did not make a conscious effort to make any adjustments of their confidence judgments between the occipital and vertex stimulation conditions in the experiment.

Furthermore, several aspects of the data are also incompatible with the interpretation that subjects deliberately adjusted their confidence criteria. As can be seen in Fig. 4, three of the six subjects used the confidence levels in a very biased manner (overusing either the high- or low-confidence responses), which suggests that they did not continuously re-adjust their confidence criteria to achieve a somewhat even use of high- and low-confidence ratings. Additionally, if such deliberate confidence adjustment actually occurred, one would expect that confidence would decrease for vertex TMS and increase for occipital TMS over the course of the experiment. The reason for this is that adjustments to the confidence criteria would become easier, later in the experiment when subjects have seen trials with both occipital and vertex TMS. However, our results showed that the difference in confidence was highest between the first vertex and first occipital TMS blocks (8.6% more high-confidence trials in the occipital TMS block) compared with the second (2.1% difference) and third (3.6% difference) blocks. This pattern of results is the opposite of what one would expect if the difference in confidence were due to subjects' deliberately adjusting their confidence criteria.

Finally, if subjects adjusted their confidence ratings, one would predict that confidence should be the same across the occipital and vertex stimulation conditions. It is hard to see why, in the case of occipital stimulation, that subjects would overadjust, such that confidence ratings would become higher than in the case of vertex stimulation.

Although none of the above arguments rules out conclusively the interpretation that subjects consciously adjusted their confidence ratings, these arguments make such an interpretation highly unlikely. Further experiments, in which TMS to the occipital cortex and vertex is interleaved within each block, are needed to conclusively settle this issue.

ACKNOWLEDGMENTS

We thank David Zhou and Linda Bahdo for assistance with the TMS sessions.

GRANTS

Funding for H. Lau was provided by the Human Frontiers Science Project (HFSP Short-Term Fellowship) and from the Netherlands Organization for Scientific Research (NWO Open Competition).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: D.A.R., B.M., B.L., H.L., and S.H.L. conception and design of research; D.A.R. and B.L. performed experiments; D.A.R., B.M., and H.L. analyzed data; D.A.R., B.M., B.L., H.L., and S.H.L. interpreted results of experiments; D.A.R. prepared figures; D.A.R. and H.L. drafted manuscript; D.A.R., B.M., B.L., H.L., and S.H.L. edited and revised manuscript; D.A.R., B.M., B.L., H.L., and S.H.L. approved final version of manuscript.

REFERENCES

Abrahamyan A, Clifford CW, Arabzadeh E, Harris JA. Improving visual sensitivity with subthreshold transcranial magnetic stimulation. *J Neurosci* 31: 3290–3294, 2011.

Amassian VE, Cracco RQ, Maccabee PJ, Cracco JB, Rudell A, Eberle L. Suppression of visual perception by magnetic coil stimulation of human occipital cortex. *Electroencephalogr Clin Neurophysiol* 74: 458–462, 1989.

Boyer JL, Harrison S, Ro T. Unconscious processing of orientation and color without primary visual cortex. *Proc Natl Acad Sci USA* 102: 16875–16879, 2005.

Brainard DH. The Psychophysics Toolbox. *Spat Vis* 10: 433–436, 1997.

Breitmeyer BG, Ro T, Ogmen H. A comparison of masking by visual and transcranial magnetic stimulation: implications for the study of conscious and unconscious processing. *Conscious Cogn* 13: 829–843, 2004.

Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer, 2002.

Busey TA, Tunnicliff J, Loftus GR, Loftus E. Accounts of the confidence-accuracy relation in recognition memory. *Psychon Bull Rev* 7: 26–48, 2000.

Corthout E, Uttl B, Walsh V, Hallett, Cowey MA. Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *Neuroreport* 10: 2631–2634, 1999.

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

Dorfman DD, Alf E. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating-method data. *J Math Psychol* 6: 487–496, 1969.

Fleet ML, Brigham JC, Bothwell RK. The confidence-accuracy relationship: the effects of confidence assessment and choosing. *J Appl Soc Psychol* 17: 171–187, 1987.

Gorea A, Sagi D. Failure to handle more than one internal representation in visual detection tasks. *Proc Natl Acad Sci USA* 97: 12380–12384, 2000.

Green D, Swets J. *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons, 1966.

Harris JA, Clifford CW, Miniussi C. The functional effect of transcranial magnetic stimulation: signal suppression or neural noise generation? *J Cogn Neurosci* 20: 734–740, 2008.

Jacoby LL. A process dissociation framework: separating automatic from intentional uses of memory. *J Mem Lang* 30: 513–541, 1991.

Jolij J, Lamme VA. Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proc Natl Acad Sci USA* 102: 10747–10751, 2005.

Kammer, T. Masking visual stimuli by transcranial magnetic stimulation. *Psychol Res* 71: 659–666, 2007.

Kastner S, Demmer I, Ziemann U. Transient visual field defects induced by transcranial magnetic stimulation over human occiptal lobe. *Exp Brain Res* 118: 19–26, 1998.

Kirkpatrick S, Gelatt CD Jr, Vecchi, MP. Optimization by simulated annealing. *Science* 220: 671–680, 1983.

Koivisto M, Mantyla T, Silvanto J. The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *Neuroimage* 51: 828–834, 2010.

Koivisto M, Railo H, Salminen-Vaparanta N. Transcranial magnetic stimulation of early visual cortex interferes with subjective visual awareness and objective forced-choice performance. *Conscious Cogn* 20: 288–298, 2011.

Lau H Rosenthal D. Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci* 15: 365–373, 2011.

Lau HC Passingham RE. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc Natl Acad Sci USA* 103: 18763–18768, 2006.

Laycock R, Crewther DP, Fitzgerald PB, Crewther SG. Evidence for fast signals and later processing in human V1/V2 and V5/MT+: a TMS study of motion perception. *J Neurophysiol* 98: 1253–1262, 2007.

Luber B, Stanford A, Malaspina D, Lisanby SH. Revisiting the backward masking deficit in schizophrenia: individual differences in performance and modeling with transcranial magnetic stimulation. *Biol Psychiatry* 62: 793–799, 2007.

Maccabee PJ, Amassian VE, Cracco RQ, Cracco JB, Rudell AP, Eberle LP, Zemon V. Magnetic coil stimulation of human visual cortex: studies of perception. *Electroencephelogr Clin Neurophysiol* 43: 111–120, 1991.

Macmillan N, Creelman C *Detection Theory: A User's Guide*. New York: Cambridge University Press, 2005.

Miller BM, Fendrich R, Eliassen JC, Demirel S, Gazzaniga MS. Transcranial magnetic stimulation: delays in visual suppression due to luminance changes. *Neuroreport* 7: 1740–1744, 1996.

Morewedge CK, Kahneman D. Associative processes in intuitive judgment. *Trends Cogn Sci* 14: 435–440, 2010.

Pleskac TJ, Busemeyer JR. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev* 117: 864–901, 2010.

Rahnev D, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H. Attention induces conservative subjective biases in visual perception. *Nat Neurosci* 14: 1513–1515, 2011.

Ro T, Shelton DJM, Lee OL, Chang E. Extrageniculate mediation of unconscious vision in transcranial magnetic stimulation induced blindsight. *Proc Natl Acad Sci USA* 101: 9933–9935, 2004.

Ruzzoli M, Abrahamyan A, Clifford CWG, Marzi CA, Miniussi C, Harris JA. The effects of TMS on visual motion sensitivity: an increase in neural noise or a decrease in signal strength? *J Neurophysiol* 106: 138–143, 2011.

Ruzzoli M, Marzi CA, Miniussi C. The neural mechanisms of the effects of transcranial magnetic stimulation on perception. *J Neurophysiol* 103: 2982–2989, 2010.

Schwarzkopf DS, Silvanto J, Rees G. Stochastic resonance effects reveal the neural mechanisms of transcranial magnetic stimulation. *J Neurosci* 31: 3143–3147, 2011.

Siebner HR, Hartwigsen G, Kassuba T, Rothwell JC. How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex* 45: 1035–1042, 2009.

Silvanto J, Cowey A, Lavie N, Walsh V. Striate cortex (V1) activity gates awareness of motion. *Nat Neurosci* 8: 143–144, 2005a.

Silvanto J, Lavie N, Walsh V. Double dissociation of V1 and V5/MT activity in visual awareness. *Cereb Cortex* 15: 1736–1741, 2005b.

Silvanto. J, Muggleton NG. New light through old windows: moving beyond the "virtual lesion" approach to transcranial magnetic stimulation. *Neuroimage* 39: 549–552, 2008.

Szczepanowski R, Pessoa L. Fear perception: can objective and subjective awareness measures be dissociated? *J Vis* 7: 1–17, 2007.

Thielscher A, Reichenbach A, Uğurbil K, Uludağ K. The cortical site of visual suppression by transcranial magnetic stimulation. *Cereb Cortex* 20: 328–338, 2010.

Weiskrantz L. *Blindsight: A Case Study and Implications*. Oxford, UK: Oxford University Press, 1986.