Causal evidence for frontal cortex organization for perceptual decision making

Dobromir Rahnev[a,b,1], Derek Evan Nee[b], Justin Riddle[c], Alina Sue Larson[c], and Mark D'Esposito[b,c]

[a]Department of Psychology, Georgia Institute of Technology, Atlanta, GA 30332; [b]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720; and [c]Department of Psychology, University of California, Berkeley, CA 94720

Edited by Randolph Blake, Vanderbilt University, Nashville, TN, and approved March 24, 2016 (received for review November 15, 2015)

Although recent research has shown that the frontal cortex has a critical role in perceptual decision making, an overarching theory of frontal functional organization for perception has yet to emerge. Perceptual decision making is temporally organized such that it requires the processes of selection, criterion setting, and evaluation. We hypothesized that exploring this temporal structure would reveal a large-scale frontal organization for perception. A causal intervention with transcranial magnetic stimulation revealed clear specialization along the rostrocaudal axis such that the control of successive stages of perceptual decision making was selectively affected by perturbation of successively rostral areas. Simulations with a dynamic model of decision making suggested distinct computational contributions of each region. Finally, the emergent frontal gradient was further corroborated by functional MRI. These causal results provide an organizational principle for the role of frontal cortex in the control of perceptual decision making and suggest specific mechanistic contributions for its different subregions.

perception | frontal cortex | hierarchy | TMS | fMRI

The frontal cortex has extensive connections with most other cortical and subcortical structures, placing it in a unique position to orchestrate a wide range of processes (1). Even though, historically, only a few studies have investigated the involvement of the frontal cortex in perceptual processes, a large amount of recent research has demonstrated that the frontal cortex has a critical role in the control of perceptual decision making (2–5). Despite these empirical findings, the unique contributions of different functional subdivisions within frontal cortex for perceptual decision making remain underspecified.

We propose that a frontal organization for perception emerges when one considers the temporal structure of perceptual decision making. Perceptual judgments consist of subsequent stages, such as selection, criterion setting, and evaluation processes (3, 4, 6). Here, we use the term "selection processes" to refer to mechanisms that allow the individual to direct resources to a specific object, feature, or part of space; "criterion setting processes" to refer to mechanisms that allow the individual to exert control over the final perceptual decision by adjusting the criteria for making the decision; and "evaluation processes" to refer to mechanisms that allow the individual to determine the likelihood that a perceptual judgment was correct. The temporal dependency between these three processes is evident when considering that the stimulus needs first be selected before decision criteria can be applied, and that both of these processes need to occur before evaluation can fully take place. It is likely that these processes partially overlap in some cases (e.g., criterion setting can be initialized, even if not fully completed, before the stimulus selection has concluded), but such partial overlaps do not undermine the general temporal structure of these three processes.

How is the frontal cortex organized to support and control these three stages of perceptual decision making? Several organizational principles of the frontal cortex have emerged in recent years. Notably, convergent evidence points to a rostrocaudal (i.e., anterior-to-posterior) gradient in the frontal cortex such that rostral regions support more abstract representations that build on the representations in caudal areas (1, 7–10). In particular, Fuster and Bressler (11) argue that progressively rostral regions are critical for progressively later stages of the perception/action cycle. Despite the emphasis on both perception and action, this representational structure of the frontal cortex has been studied virtually exclusively with regard to cognitive control over action, and has not directly been linked to the processes underlying perceptual decision making. We address this gap and specifically investigate whether selection, criterion setting, and evaluation processes necessary for perceptual decision making are controlled by the caudal, midlateral, and rostral frontal cortex, respectively.

Most research on the role of frontal cortex in perception has thus far been correlational. However, because the same regions of frontal cortex often support a variety of cognitive functions (1), such studies cannot conclusively establish the degree of specialization of different subregions of frontal cortex. In addition, previous studies have typically focused on a single one of these three perceptual processes, and thus could not directly compare their dependence on regions within the frontal cortex.

Here, we move beyond these limitations, and use causal techniques to explore the theoretically driven question of whether successive perceptual processes are controlled by progressively rostral regions of the frontal cortex. We designed a strong test of this hypothesis by a priori defining for each subject three regions along the lateral frontal cortex that are involved in the three proposed perceptual decision-making processes, and then targeted these regions with transcranial magnetic stimulation (TMS) to disrupt their function. This causal approach allowed us to move beyond previous correlational studies, which have found widespread frontal cortex activity during perceptual decision-making

Significance

The frontal cortex has long been understood as the seat of higher level cognition. Recent research, however, highlights its role in modulating perception. Here, we present a theoretical framework for frontal involvement in perceptual decision making and test it with the causal technique of transcranial magnetic stimulation. We find that progressively rostral regions of frontal cortex are involved in the control of progressively later stages of perceptual decision making. These causal findings are further corroborated by functional MRI and simulations of a dynamic model of decision making. Our results point to a critical role of the frontal cortex in the control of perceptual processes and reveal its intrinsic organization in support of modulating perception.

Author contributions: D.R. and M.D. designed research; D.R., J.R., and A.S.L. performed research; D.R., D.E.N., and J.R. analyzed data; and D.R., D.E.N., J.R., A.S.L., and M.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 5771.

[1]To whom correspondence should be addressed. Email: drahnev@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522551113/-/DCSupplemental.

tasks (12), and to test directly the necessity of each region for the control of each processing stage. Our task required subjects to deploy spatial attention to engage selection processes (6), follow speed/accuracy instructions to engage criterion setting processes (13), and provide metacognitive judgments to engage evaluation processes (14). We found clear evidence for frontal cortex organization such that progressively rostral regions were necessary for controlling later stages of processing during perceptual decision making. This emergent gradient was corroborated by simulations derived from a dynamic model of decision making that suggested specific computational contributions of each frontal region, as well as functional MRI (fMRI) data that extended the TMS results.

Results

We designed a task in which the processes of selection, criterion setting, and evaluation could be clearly identified (Fig. 1*A*). On each trial, subjects were instructed to attend selectively to one of two peripheral stimuli (selection). The task was to indicate the orientation (clockwise/counterclockwise) of a grating embedded in noise while adjusting the decision criterion so as to emphasize either speed or accuracy (criterion setting). After making their choice, subjects indicated their level of confidence (evaluation).

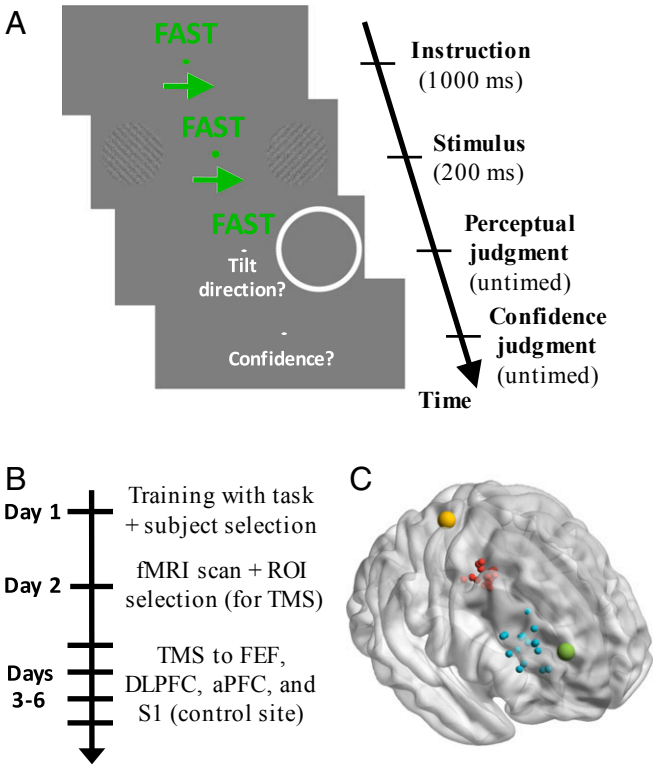

Fig. 1. Task, experiment time line, and TMS locations. (*A*) Trial sequence. Each trial began with a 1-s instruction to attend to either the left or right stimulus, as well as to emphasize either speed or accuracy. The grating stimuli were presented for 200 ms, and a postcue indicated which stimulus subjects should respond to. The postcue was on the attended side 66.7% of the time. Responses regarding stimulus orientation (clockwise/counterclockwise) and confidence (on a 1–4 scale) were untimed. The following trial began 1 s later. (*B*) Experiment time line. ROI, region of interest. (*C*) Approximate location of S1 is depicted in yellow (the target was identified in the postcentral gyrus). FEF (red) and DLPFC (blue) were localized separately for each subject based on individual fMRI activations (each dot represents a different subject). Finally, the site for aPFC stimulation (green) was common across subjects and based on Fleming et al. (4). All targets were identified in the right hemisphere. The *y* coordinates for each region did not overlap: S1: −33, FEF: [−10, 2], DLPFC: [26, 48], aPFC: 53.

Each subject received training (day 1), then performed the task during the collection of fMRI data (day 2), and finally received TMS to one of four different sites before performing the same task (days 3–6; Fig. 1*B*). Based on the fMRI data, for each subject, we identified three progressively more rostral sites in frontal cortex: putative frontal eye fields (FEFs), dorsolateral prefrontal cortex (DLPFC), and anterior prefrontal cortex (aPFC), as well as the primary somatosensory cortex (S1), which served as a control site (Fig. 1*C*). We then delivered continuous theta-burst stimulation (cTBS) to each of these regions for each subject on different days. cTBS has been demonstrated to produce a decrease in the excitation level in the stimulated cortex (15), likely through processes akin to long-term depression.

TMS Evidence for Frontal Organization for Perception. TMS did not influence overall task performance as measured by overall accuracy, reaction time (RT), or confidence ($P > 0.05$ for all pairwise comparisons between any of the four sites; Table S1), suggesting it is unlikely that frontal cortex is necessary for the low-level visual processing. We now turn to the frontal cortex involvement in the control of selection, criterion setting, and evaluation processes.

Selection (spatial cue). The first critical component of the task was a requirement to control the way stimuli were selected for processing: a cue indicated which of two stimuli to attend. Subjects successfully followed the spatial cue as demonstrated by faster RTs for attended compared with unattended stimuli during the fMRI session [RT difference = 128 ms, $t(16) = 8.52$, $P = 2 * 10^{-7}$]. A decreased ability to engage this selection process following TMS would manifest itself as a smaller RT difference between attended and unattended stimuli (6). We predicted that TMS to the most caudal frontal site (putative FEF) would exhibit this effect based on previous work (reviewed in 16). Consistent with this prediction, we found a significant difference in performance between different TMS sites [$\chi^2(3) = 10.6$, $P = 0.01$, mixed-effects model (17); Fig. 2*A*]. A planned post hoc *t* test confirmed that the RT difference between attended and unattended stimuli was significantly decreased after FEF stimulation compared with the control site [RT difference = 102 ms, $t(16) = 2.89$, $P = 0.011$], corresponding to an effect size of d = 0.7. Exploratory analyses also demonstrated a significant difference in this selection effect between FEF TMS and both DLPFC TMS [RT difference = 77 ms, $t(16) = 2.25$, $P = 0.039$, d = 0.55] and aPFC TMS [RT difference = 75 ms, $t(16) = 3.1$, $P = 0.007$, d = 0.75]. No significant differences were found between S1, DLPFC, and aPFC ($P > 0.05$ for all pairwise comparisons, RT differences < 28 ms). Thus, these findings strongly suggest that the selection process depends on the caudal frontal cortex (putative FEF) but not on the more rostral frontal regions.

Criterion setting (speed/accuracy instruction). The second critical component of the task involved a requirement to set a perceptual criterion by emphasizing on different trials either speed or accuracy. Such adjustment of the response threshold has long been considered an important example of how decision criteria are set in perceptual decision making (5, 13). Subjects successfully followed the instructions as demonstrated by a large RT difference between accuracy and speed trials during the fMRI session [RT difference = 370 ms, $t(16) = 5.19$, $P = 9 * 10^{-5}$]. A decreased ability to set the response criterion appropriately would manifest in a smaller RT difference between the two types of trials. We predicted that TMS to the middle of the rostrocaudal frontal gradient (DLPFC) would interfere with the control of the criterion setting process, based on previous work (5). Consistent with this prediction, we found a significant difference in performance between different TMS sites [$\chi^2(3) = 15.3$, $P = 0.002$, mixed-effects model; Fig. 2*B*)]. A planned post hoc *t* test confirmed that the RT difference between accuracy and speed instructions was significantly decreased following DLFPC TMS compared with the control site [RT difference = 55 ms, $t(16) = 3.31$,

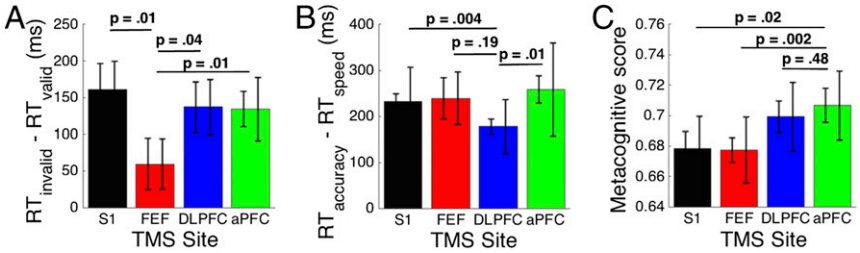

Fig. 2. TMS results. (*A*) TMS to FEF decreased subjects' ability to follow the spatial cue, as quantified by the RT difference between unattended and attended stimuli. (*B*) TMS to DLPFC decreased subjects' ability to follow speed/accuracy instruction, as quantified by the RT difference between accuracy and speed trials. (*C*) TMS to aPFC increased subjects' metacognitive scores, as quantified by the Type 2 AUC curve. The increase was similar but smaller for DLPFC. The left error bars represent the within-subject SE for the comparison with FEF (*A*), DLPFC (*B*), and aPFC (*C*). The error bar for the comparison site is the same as the S1 error bar. The right error bars represent the between-subject SE, and are not indicative of the significance of the effects.

$P = 0.004$, d = 0.8]. Exploratory analyses also demonstrated a significant difference in this effect between DLPFC TMS and aPFC TMS [RT difference = 81 ms, $t(16) = 2.74$, $P = 0.01$, d = 0.66] but not between DLPFC TMS and FEF TMS [RT difference = 62 ms, $t(16) = 1.38$, $P = 0.19$, d = 0.34]. No significant differences were found between S1, FEF, and aPFC ($P > 0.05$ for all pairwise comparisons, RT differences < 27 ms). These results suggest a critical role for DLPFC (located in the middle part of the rostrocaudal gradient in frontal cortex) in the control of the criterion setting process.

Evaluation (metacognitive ratings). The third critical component of the task required subjects to evaluate their perceptual judgments by providing a confidence rating. We investigated the extent to which these confidence ratings were linked to subjects' accuracy, which is a measure of subjects' metacognitive ability. This correspondence was determined as the area under the type 2 receiver operating characteristic curve (Type 2 AUC) (18) (*Materials and Methods*). We predicted that TMS to the most rostral area of frontal cortex (aPFC) would impair subjects' metacognitive scores, based on previous work (4, 18). However, the observed effect was in the opposite direction such that TMS to the rostral part of frontal cortex improved metacognition. Indeed, we found a significant difference in Type 2 AUC between different TMS sites [$\chi^2(3) = 11$, $P = 0.01$, mixed-effects model; Fig. 2*C*]. A planned t test demonstrated that the metacognition score was significantly higher after aPFC TMS compared with the control site [Type 2 AUC difference = 0.03; $t(16) = 2.51$, $P = 0.02$, d = 0.61]. Exploratory analyses showed that subjects' metacognitive scores were also higher after aPFC TMS compared with FEF TMS [Type 2 AUC difference = 0.03; $t(16) = 3.61$, $P = 0.002$, d = 0.88], although there was no significant difference between TMS to aPFC and DLPFC [$t(16) = 0.72$, $P = 0.48$, d = 0.17]. Comparing the other three sites (S1, FEF, and DLPFC), we found that TMS to DLPFC led to significantly higher metacognitive scores compared with TMS to FEF [Type 2 AUC difference = 0.02; $t(16) = 2.39$, $P = 0.03$, d = 0.58], although no significant differences were found in the other two comparisons ($P > 0.05$ in both cases).

The finding that TMS to DLPFC affected metacognition, despite our prediction that only TMS to aPFC would do so, could be partly due to the fact that DLPFC was localized in a very anterior location for most subjects (Fig. 1*C*). Thus, this finding does not necessarily contradict the possibility that metacognitive sensitivity depends primarily on the rostral part of frontal cortex.

Due to our unexpected findings (aPFC TMS leading to improved rather than impaired metacognitive performance), we sought to confirm that our results were not due to the specific measure of metacognition that we chose. We repeated our analyses with three more measures: meta-d′ (19), a simple correlation between confidence and accuracy [also known as phi (20)], and the difference in confidence between correct and incorrect trials. All three measures showed the exact same pattern of results (Table S2). Specifically, aPFC TMS led to significantly higher metacognitive scores than both the control site and FEF for each measure ($P <$ 0.05 in all cases).

Comparing the three measures. The three results above suggest a selective association between FEF, DLPFC, and aPFC and the processes of selection, criterion setting, and evaluation in perceptual decision making, respectively. To corroborate this conclusion further, we found a significant interaction [$\chi^2(6) = 16.3$, $P = 0.01$, mixed-effects model] between the TMS site (S1, FEF, DLPFC, and aPFC) and the task component (selection, criterion setting, and evaluation). However, because not all pairwise comparisons were significant for each measure, we cannot conclude the existence of a complete triple dissociation among these three regions.

Simulating the TMS Effects with a Dynamic Model of Decision Making. Our results suggest that caudal, middle, and rostral frontal cortex have differential contributions to perceptual decision making. To understand the functional role of each region better, we performed simulations using an adapted model of perceptual decision making introduced by Kepecs et al. (21) and De Martino et al. (22), wherein evidence is accumulated separately for each of the two choices, and the decision is made when one of the accumulators reaches a bound (23). Confidence is then assigned as the noise-corrupted difference between the winning and losing accumulators (Δe, the difference in evidence; Fig. 3*A*) such that higher difference indicates higher confidence. The critical parameters of the model are (*i*) the drift rate, which determines how quickly evidence accumulates for each choice; (*ii*) the bound, which determines how much evidence is needed to make a decision; and (*iii*) the confidence noise, which determines the strength of the association between confidence and accuracy.

This modeling framework provides a natural way to operationalize the processes of selection, criterion setting, and evaluation using the above parameters (Fig. 3*A*). First, selection is defined as the process of enhancing the sensitivity for one stimulus over another. In the framework of our model, this process is equivalent to boosting the drift rate for the correct choice for the attended, but not the unattended, stimulus. Second, the requirement to set the response criterion according to the speed/accuracy instructions is naturally modeled by an adjustment of the bound to be higher for accuracy compared with speed instructions. Third, we observed significant variability in the metacognitive scores (from 0.58 to 0.83 in the fMRI session), which points to the existence of confidence noise that varies between subjects (22). This confidence noise controls how tightly the metacognitive ratings follow a subject's decision accuracy such that a greater amount of this type of noise leads to lower metacognitive scores.

Our simulations demonstrated that changes to these three parameters of the model can qualitatively reproduce our frontal

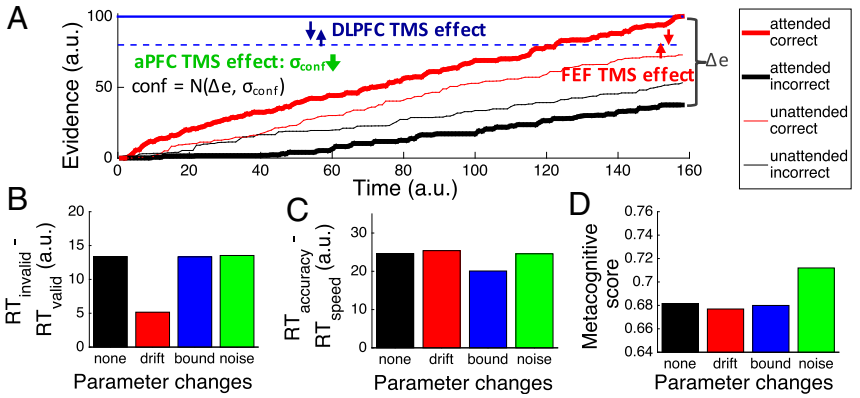

Fig. 3. Dynamic model of perceptual decision making. (*A*) Three critical parameters in our model were drift rate (the amount of perceptual evidence), bound (the decision criterion that controls how quickly subjects give their response), and confidence noise (the amount of noise added to the metacognitive decision). The figure depicts the evidence traces for an attended trial (thick lines) and an unattended trial (thin lines), as well as the decision criteria for accuracy focus (solid blue line) and speed focus (dashed blue line). The results of TMS to FEF, DLPFC, and aPFC were reproduced by changes in the difference between drift rates for attended and unattended stimuli (red arrows), the difference in the bound between the accuracy and speed instructions (blue arrows), and the confidence noise across all trials (green arrow), respectively. We performed four simulations runs changing each of these parameters, as well as a control simulation with default parameters. The predicted pattern of RT difference between unattended and attended stimuli (*B*), accuracy and speed instructions (*C*), and the metacognitive scores (*D*) was found, suggesting that TMS to different frontal brain regions affected different parameters within our dynamic decision model. a.u., arbitrary units.

TMS effects. First, the smaller difference in RT between attended and unattended targets after FEF TMS is reproduced by a smaller difference in the drift rate between attended and unattended conditions (red arrows in Fig. 3*A* and results in Fig. 3*B*). Second, the smaller difference in RT between accuracy and speed instructions after DLPFC TMS is reproduced by a smaller difference in the bound between speed and accuracy focus (blue arrows in Fig. 3*A* and results in Fig. 3*C*). Finally, the unexpected finding of higher metacognitive score after aPFC TMS is reproduced by a decrease in the confidence noise (Fig. 3*D*). Our simulations assumed that TMS to each of these regions affected only a single parameter of the model, which is why the simulated data do not perfectly reflect the empirical results (Fig. 2). For example, the metacognitive score after DLPFC TMS increased compared with our control site, but this increase is not reflected in the simulations. However, what is important here is the demonstration that the TMS effects on the processes of selection, criterion setting, and evaluation can be naturally understood computationally in the context of our model of dynamic decision making.

Frontal Organization Corroborated by fMRI. Our TMS results and model simulation were consistent with our predictions that progressively rostral frontal regions are involved in progressively later processing stages during perceptual decision making. Because, as we noted above, the three stages are temporally organized, another prediction is that more rostral frontal regions will become active later in the course of each trial of our task. We tested this prediction by using the fMRI data from day 2 to characterize the activity in frontal cortex during the (*i*) instruction, (*ii*) stimulus/perceptual judgment, and (*iii*) confidence epochs of the task. We do not claim that the selection, criterion setting, and evaluation processes occur exclusively during the instruction, stimulus/perceptual judgment, and confidence epochs of the task, respectively. Instead, a temporal hierarchy exists whereby the stimulus needs first be selected before decision criteria can be applied, and both of these processes need to occur before evaluation can take place. This temporal hierarchy implies that each process should peak later than the previous one, even in the absence of one-to-one correspondence between the three processes and the three task epochs. The design of our task was optimized for the TMS effects rather than this particular analysis, but the results confirmed our prediction nonetheless. Specifically, we found a clear rostrocaudal gradient such that the activity in progressively rostral frontal regions peaked during progressively later epochs of our task (Fig. 4).

We first examined the brain activity during each of the three epochs of the task (Fig. 4*A*). The whole-brain activation patterns for each task epoch are shown and discussed in greater detail in Fig. S1 (we note that the pattern of activity in the left hemisphere was similar to the right hemisphere, and we provide a link to complete unthresholded maps; *Materials and Methods*). Here, we focus on the results in the frontal cortex. We found that frontal cortex activity during the instruction epoch was mostly constrained to a caudal region, activity during the stimulus/perceptual judgment epoch extended from caudal to midlateral frontal regions, and activity during the confidence epoch extended across the entire lateral surface of the frontal cortex.

Critically, we found that progressively rostral frontal regions were activated maximally during progressively later task epochs (Fig. 4*B*). Indeed, we observed a significant interaction between region (FEF, DLPFC, aPFC) and task epoch (instruction, stimulus/perceptual judgment, confidence) [$F(4,40) = 22.16$, $P < 0.00001$, repeated measures ANOVA]. Specifically, FEF activity was greatest early in each trial, DLPFC activity was greatest in the middle of the trial, and aPFC activity was greatest at the end of the trial. The most caudal frontal region, FEF, was more active during the instruction [$t(20) = 2.09$, $P = 0.049$, d = 0.46] and stimulus/perceptual judgment [$t(20) = 4.31$, $P = 0.0003$, d = 0.94] epochs, compared with the confidence epoch. FEF activity during the instruction and stimulus/perceptual judgment epochs was not significantly different ($P = 0.99$), which may be explained by the observation that FEF is responsive to stimulus presentation (16). The middle frontal region, DLPFC, was more active during the stimulus/perceptual judgment epoch compared with both the instruction [$t(20) = 4.52$, $P = 0.0002$, d = 0.99] and confidence [$t(20) = 2.33$, $P = 0.03$, d = 0.51] epochs. Finally, the most rostral frontal region, aPFC, was less active during the instruction epoch compared with both the stimulus/perceptual judgment [$t(20) = 7.32$, $P = 4 * 10^{-7}$, d = 1.6] and confidence [$t(20) = 6.88$, $P = 1*10^{-6}$, d = 1.5] epochs. aPFC activations during the stimulus/perceptual judgment and confidence epochs were not significantly different ($P = 0.33$), which may be partly due to the evaluation process starting immediately after making the perceptual decision internally, which is likely a few hundred milliseconds before the

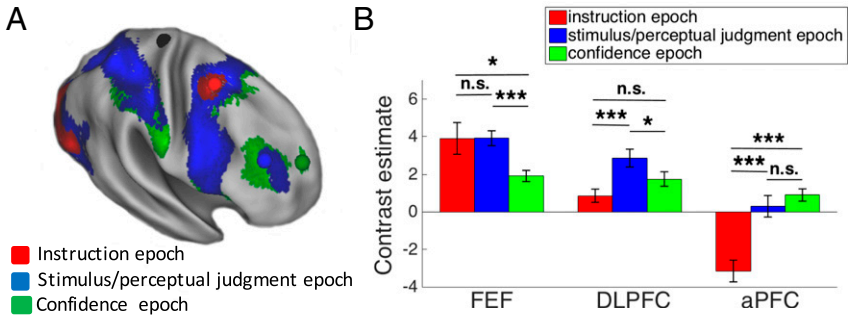

Fig. 4. fMRI results. (*A*) Brain activity corresponding to the instruction, stimulus/perceptual judgment, and confidence epochs. A caudal-to-rostral gradient is apparent with later epochs of the trial activating preferentially more rostral regions. The colored spheres are the mean locations of the stimulated S1 (black), FEF (red), DLPFC (blue), and aPFC (green) sites. (*B*) Mean blood-oxygenation-level–dependent (BOLD) contrast estimate for each trial epoch (beta value difference between the regressor for the relevant epoch and regressor for the "rest" period) is shown for each of the three regions, demonstrating that caudal regions are active earlier in the trial, whereas rostral regions are active later in the trial. Error bars represent SE. $*P < 0.05$; $***P < 0.001$.

button press that we used as an external indicator of the end of the stimulus/perceptual judgment epoch.

The above results were obtained by creating separate generalized linear model (GLM) models for each task epoch (*SI Materials and Methods*) to identify the full extent of activity during each task epoch. In a control analysis, we analyzed all three task epochs in the same GLM and obtained very similar results (Figs. S2 and S3).

Discussion

Despite numerous studies demonstrating the involvement of the frontal cortex in various high-level perceptual processes (2–5), the roles of distinct areas within frontal cortex during perceptual decision making remain underspecified. In this study, we provide a principle for frontal cortex functional organization based on the temporal organization of perception in the processes of selection, criterion setting, and evaluation. More specifically, convergent evidence from TMS and fMRI demonstrated that there are distinct frontal regions along a rostrocaudal (i.e., anterior-to-posterior) gradient that are necessary for the control of progressively later stages of the perceptual decision-making process.

Our results based on a causal intervention with TMS provide a critical addition to the literature on the contribution of frontal cortex to perceptual decision making that is largely based on correlational studies. Using correlational techniques, some studies claimed that relatively caudal regions of the frontal cortex are important for some of the later perceptual stages of processing. For example, speed/accuracy signals were found in FEF neurons (24), and confidence signals were found in supplementary eye field neurons (25). However, in our study, disruption of caudal frontal cortex function with TMS did not have a significant effect either on speed/accuracy or on confidence. It is possible that these differences are due to interspecies variation in the organization of frontal cortex and/or the substantial difference in the tasks used. Another important possibility is that because the perceptual decision was indicated via a saccade in both of these studies, the speed/accuracy and confidence signals were passed to the eye movement effector system but were nevertheless computed in more anterior areas of frontal cortex. This possibility is consistent with a recent study in which monkeys indicated the perceptual decisions using their hands and speed/accuracy signals were present in the primary motor cortex even though it is unlikely that these signals originated there (26). More studies that use causal interventions in both humans and monkeys are needed to determine the etiology of the discrepancies between our and these previous studies.

The functional gradient revealed in our data has strong implications regarding the general organization of the frontal cortex. A critical mass of studies has suggested the existence of a rostrocaudal gradient in the frontal cortex (1, 7–10). Although these studies differ in the details of the type of processes or representations being linked to each PFC subregion, each proposes a hierarchical organization with more rostral regions involved in the processing of more abstract representations (1, 7). Other studies, however, have proposed that the lateral frontal cortex is homogeneous in function without a functional gradient (12, 27, 28). This debate is complicated by the correlational nature of most previous studies. However, two previous studies of patients with focal brain lesions found causal support for a rostrocaudal gradient in frontal cortex (9, 10). The current results extend these previous patient studies by providing causal evidence from healthy human subjects in support of a rostrocaudal functional organization of frontal cortex.

Simulations based on a dynamic computational model of perceptual decision making (21–23) were able to reproduce the observed empirical TMS effects. The decrease in the RT advantage for attended stimuli following FEF TMS could be reproduced by decreasing the difference in drift rate between attended and unattended stimuli. Thus, one possibility is that the caudal frontal cortex biases the processing of visual information such that one stimulus is favored over another through a process akin to gain amplification (16, 29). This possibility is further corroborated by the known connectivity of FEF to early visual areas that respond to the visual stimulus (30). The decrease in the RT difference between accuracy and speed focus following DLPFC TMS could be reproduced by decreasing the difference in the decision bound between the two conditions. One possibility is that DLPFC is involved in the adjustment of the decision criterion. Such a role is facilitated by the wide connectivity of DLPFC with higher visual and parietal (as well as premotor and subcortical) areas (5). Finally, the improved metacognitive performance after aPFC TMS could be reproduced by decreasing the noise term in confidence decisions, consistent with a role of aPFC in metacognitive evaluations. This type of metacognitive process likely requires communication only with other high-level regions, such as frontal and parietal cortices, which is consistent with the connectivity pattern of aPFC (31). In summary, even though our simulations were intended as, and should only be seen as, a proof of concept, they are consistent with a rostrocaudal organization of frontal cortex function in relation to visual perception. A similar idea has been put forth in the context of linking perception with action (1).

Surprisingly, we found improvement in metacognition after aPFC TMS. Despite the unexpected nature of this result, it is actually in line with a pair of recent studies. The first one reported similar metacognitive enhancement after aPFC TMS on a memory task (32). The second one showed that monkeys with lesions to

rostral frontal cortex showed behavioral improvements on certain tasks (33). Specifically, they remained more focused in exploiting the current task when faced with various interruptions, potentially suggesting a role for rostral frontal cortex in reallocating cognitive resources for new purposes. Nevertheless, metacognitive impairment after TMS to a more posterior site in middle frontal gyrus has also been reported (34). Critically, in our study, average confidence ratings were not affected by aPFC TMS; instead, what was improved was the correspondence between the trial-by-trial confidence ratings and accuracy. Several types of explanations have been provided for TMS-induced performance improvements. For example, if TMS suppresses the noise more than the signal, behavioral performance would improve rather than decline (35). Another possibility is that behavioral performance can improve if TMS disrupts processes that are normally detrimental to the experimental task (36). This last explanation also fits with the monkey data discussed above (33). In partial support of this last possibility, we previously suggested a role for aPFC in decreasing the amount to which confidence on a previous trial biases the confidence rating on a current trial, a phenomenon dubbed "confidence leak" (37). Such confidence leak is likely beneficial in most everyday tasks but is suboptimal in laboratory tasks in which successive trials are independent and the previous stimulus should therefore be ignored during the current decision. Additional analyses (*SI Results*) demonstrated that aPFC TMS decreased the amount of confidence leak, which could have contributed to the improvements in metacognition. Nevertheless, in the absence of direct neural evidence, each of these explanations remains speculative. Regardless of the explanation of our finding, it does support a critical role for aPFC in metacognition, and is consistent with the existence of a rostrocaudal gradient in frontal cortex for perception.

Materials and Methods

Forty-one subjects were tested in an initial screening session. Twenty-one of these subjects were able to perform the task appropriately by following both the attentional and speed/accuracy instructions, and were therefore invited to participate in the five additional days of testing. Four subjects were unable to complete all six sessions; thus, a total of 17 subjects completed the study (11 females and 6 males, average age = 23.06 y, age range: 21–30 y). All participants had normal or corrected-to-normal vision. They received detailed information about the potential side effects of TMS and provided written informed consent. All procedures were approved by the University of California, Berkeley Committee for the Protection of Human Subjects.

All behavioral data and codes that reproduce every analysis and figure are freely available at https://github.com/DobyRahnev/TBS-to-PFC. In addition, unthresholded fMRI maps are uploaded at neurovault.org/collections/599.

ACKNOWLEDGMENTS. This work is supported by NIH Grants MH63901 and NS79698.

1. Fuster JM (2008) *The Prefrontal Cortex* (Academic, London), 4th Ed.
2. Rahnev D, Lau H, de Lange FP (2011) Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *J Neurosci* 31(29):10741–10748.
3. Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. *Nat Rev Neurosci* 9(6):467–479.
4. Fleming SM, Huijgen J, Dolan RJ (2012) Prefrontal contributions to metacognition in perceptual decision making. *J Neurosci* 32(18):6117–6125.
5. van Veen V, Krug MK, Carter CS (2008) The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *J Cogn Neurosci* 20(11): 1952–1965.
6. Posner MI (1980) Orienting of attention. *Q J Exp Psychol* 32(1):3–25.
7. Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci* 10(9):659–669.
8. Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302(5648):1181–5.
9. Azuar C, et al. (2014) Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions. *Neuroimage* 84:1053–1060.
10. Badre D, Hoffman J, Cooney JW, D'Esposito M (2009) Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nat Neurosci* 12(4):515–522.
11. Fuster JM, Bressler SL (2012) Cognit activation: A mechanism enabling temporal integration in working memory. *Trends Cogn Sci* 16(4):207–218.
12. Crittenden BM, Duncan J (2014) Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cereb Cortex* 24(2):532–540.
13. Wickelgren W (1977) Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol (Amst)* 41:67–85.
14. Fleming SM, Dolan RJ, Frith CD (2012) Metacognition: Computation, biology and function. *Philos Trans R Soc Lond B Biol Sci* 367(1594):1280–1286.
15. Huang Y-Z, Edwards MJ, Rounis E, Bhatia KP, Rothwell JC (2005) Theta burst stimulation of the human motor cortex. *Neuron* 45(2):201–206.
16. Vernet M, Quentin R, Chanes L, Mitsumasu A, Valero-Cabré A (2014) Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Front Integr Neurosci* 8:66.
17. Pinheiro JC, Bates DM (2000) *Mixed-Effects Models in Sand S-PLUS* (Springer, New York).
18. Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329(5998):1541–1543.
19. Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 21(1):422–430.
20. Kornell N, Son LK, Terrace HS (2007) Transfer of metacognitive skills and hint seeking in monkeys. *Psychol Sci* 18(1):64–71.
21. Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455(7210):227–231.
22. De Martino B, Fleming SM, Garrett N, Dolan RJ (2013) Confidence in value-based choice. *Nat Neurosci* 16(1):105–110.
23. Vickers D (1970) Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* 13(1):37–58.
24. Heitz RP, Schall JD (2012) Neural mechanisms of speed-accuracy tradeoff. *Neuron* 76(3):616–628.
25. So N, Stuphorn V (2016) Supplementary Eye Field Encodes Confidence in Decisions Under Risk. *Cereb Cortex* 26(2):764–782.
26. Thura D, Cisek P (2016) Modulation of Premotor and Primary Motor Cortical Activity during Volitional Adjustments of Speed-Accuracy Trade-Offs. *J Neurosci* 36(3):938–956.
27. Duncan J (2013) The structure of cognition: Attentional episodes in mind and brain. *Neuron* 80(1):35–50.
28. Reynolds JR, O'Reilly RC, Cohen JD, Braver TS (2012) The function and organization of lateral prefrontal cortex: A test of competing hypotheses. *PLoS One* 7(2):e30284.
29. Gregoriou GG, Rossi AF, Ungerleider LG, Desimone R (2014) Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nat Neurosci* 17(7):1003–1011.
30. Ruff CC, et al. (2006) Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Curr Biol* 16(15):1479–1488.
31. Passingham RE, Wise SP (2012) *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight* (Oxford Psychology, Oxford).
32. Ryals AJ, Rogers LM, Gross EZ, Polnaszek KL, Voss JL (2016) Associative Recognition Memory Awareness Improved by Theta-Burst Stimulation of Frontopolar Cortex. *Cereb Cortex* 26(3):1200–1210.
33. Mansouri FA, Buckley MJ, Mahboubi M, Tanaka K (2015) Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices. *Proc Natl Acad Sci USA* 112(29):E3940–E3949.
34. Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn Neurosci* 1(3):165–175.
35. Allen CPG, et al. (2014) Enhanced awareness followed reversible inhibition of human visual cortex: A combined TMS, MRS and MEG study. *PLoS One* 9(6):e100350.
36. Tadin D, Silvanto J, Pascual-Leone A, Battelli L (2011) Improved motion perception and impaired spatial suppression following disruption of cortical area MT/V5. *J Neurosci* 31(4):1279–1283.
37. Rahnev D, Koizumi A, McCurdy LY, D'Esposito M, Lau H (2015) Confidence Leak in Perceptual Decision Making. *Psychol Sci* 26(11):1664–1680.
38. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.
39. Yokoyama O, et al. (2010) Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neurosci Res* 68(3):199–206.
40. Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL (2000) Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* 11(6 Pt 1):735–759.
41. Rahnev D, et al. (2013) Continuous theta burst transcranial magnetic stimulation reduces resting state connectivity between visual areas. *J Neurophysiol* 110(8): 1811–1821.
42. Rahnev DA, Maniscalco B, Luber B, Lau H, Lisanby SH (2012) Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J Neurophysiol* 107(6):1556–1563.
43. Bates D, Maechler M, Bolker BM, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv:1406.5823.
44. Fox J, Weisberg HS (2010) *An R Companion to Applied Regression* (Sage, Thousand Oaks, CA), 2nd Ed.
45. Donner A, Klar N (2000) *Design and Analysis of Cluster Randomization Trials in Health Research* (Arnold Publishing, London).

Supporting Information

Rahnev et al. 10.1073/pnas.1522551113

SI Materials and Methods

Session Sequence. The experiment took place over 6 d. On day 1, subjects received behavioral training with the task. We used the data from that day to exclude subjects who were not able to follow the attention or speed/accuracy instructions (as indicated by roughly equal RTs for attended/unattended stimuli or speed/accuracy instructions). The data from this day were not analyzed further, so any biases induced from this selection procedure did not affect our results. On day 2, subjects performed the same task in the MRI scanner to allow for the identification of the regions of interest (ROIs) to be targeted with TMS. Finally, on days 3–6, subjects performed the task after receiving TMS to one of four different regions. Three of them were ROIs in the frontal cortex: putative FEF, DLPFC, and aPFC, and the final one was a control region: S1. The order of the stimulation sites in days 3–6 was pseudorandomized across subjects. All sessions were separated by at least a week.

Task. Subjects were instructed to fixate on a small central square throughout the experiment. Each trial began with a 1,000-ms presentation of the attentional and speed/accuracy instructions. The attentional cue was an arrow (length = 3°, height = 1°) that indicated the side (left/right) to which subjects should attend (Fig. 1*B*). The speed/accuracy instruction consisted of the word "FAST" or "ACCURATE" presented in Arial font. To help subjects follow the speed/accuracy instruction, when the word FAST was presented, both the word and the arrow cue were colored in green, whereas when the word ACCURATE was presented, both the word and the arrow cue were colored in red. Two stimuli were then presented for 200 ms, while the cues were still being displayed. The stimuli were gray-scale gratings (diameter = 3°) displayed 9° to the left and right of fixation and consisting of a noisy background composed of uniformly distributed intensity values (8% contrast) on top of which we added a grating (0.5 cycles per degree). Each grating was tilted at 45° or 135° from vertical. The stimuli were then replaced by a postcue in the form of a white circle (diameter = 4°) that appeared around the location of one of the gratings but without inducing backward masking (due to its larger size). Subjects' task was to indicate the tilt of the grating in the postcued location (clockwise vs. counterclockwise from vertical). The postcue was presented at the attended location on 66.67% of the trials (valid cue trials) and at the unattended location on the remaining 33.33% of the trials (invalid cue trials). Subjects were informed about this contingency and encouraged to deploy their attention accordingly. After the decision was made via a button press, subjects indicated with a second button press their confidence using a four-point scale, where 1 was defined as low confidence and 4 was defined as high confidence. Subjects were instructed to attempt to use the whole confidence scale. During behavioral testing (days 1 and 3–6), subjects used the 1–4 keys on a computer keyboard, whereas they used a button box in the MRI scanner. Feedback was provided during training (discussed below) but not during the actual experimental sessions.

On each day of testing, subjects completed four runs, each consisting of four blocks of 30 trials for a total of 480 trials. Participants were given 15-s breaks between blocks and unlimited breaks between runs. We fully counterbalanced the speed/accuracy instruction, the direction of the attention (left/right), the validity of the attentional cue (valid/invalid), and the orientation of the postcued stimulus (clockwise/counterclockwise) over the 480 trials. The orientation of the non-postcued stimulus was chosen randomly on each trial, and was thus also independent of all of the above factors.

Subjects received extensive training on day 1 and shorter additional training at the beginning of days 2–6. On day 1, subjects completed 267 trials during which the different components of the task were introduced and the contrast of the stimuli was gradually adjusted. Trial-by-trial feedback was provided, except during the last 30 trials. The contrast was adjusted during this training session, as well as during the 480 test trials. Contrast was then fixed for days 2–6 (mean = 8.4%, SD = 2.7%). On day 2, participants were given two blocks of 24 trials of practice outside the scanner, as well as an additional complete run of 120 trials in the scanner that included trial-by-trial feedback. Finally, on days 3–6, participants were given one block of 48 trials of practice with no feedback before receiving TMS.

Stimuli were generated using Psychophysics Toolbox (38) in MATLAB (MathWorks). During the behavioral testing (days 1 and 3–6), subjects were seated in a dim room 60 cm away from the computer monitor (19-inch display, 1,024 × 768 pixel resolution, 60-Hz refresh rate). During the fMRI experiment (day 2), participants were 57 cm away from a screen mounted to the rf coil where the stimuli were back-projected via a liquid crystal display (LCD) projector. The lights in the scanner room were turned off.

Eye tracking was performed at 60 Hz with an Avotec system (Arrington Research) inside the scanner during day 2. These data showed that subjects were able to maintain fixation and that there was no difference in eye position with attentional or speed/accuracy instruction ($P > 0.2$ for both). During the TMS sessions (days 3–6), eye tracking was not possible for logistical reasons but subjects' eyes were recorded using a desktop video camera. Recordings were visually inspected for the presence of large eye movements.

fMRI Acquisition. MRI scanning was done at the Henry H. Wheeler, Jr. Brain Imaging Center at the University of California, Berkeley. Images were acquired on a Siemens TIM/Trio 3 T MRI System using a 12-channel receive-only head coil, with a single-shot gradient echo-planar imaging (EPI) sequence [repetition time (TR) = 2,000 ms; echo time (TE) = 24 ms; 37 descending slices; voxel size = 3 × 3 × 3 mm; slice thickness = 3.5 mm; interslice gap = 0.50 mm; flip angle = 70°; field of view = 224 mm, matrix = 64 × 64, fat suppression and prescan normalization included]. A high-resolution T1-weighed structural 3D magnetization-prepared rapid acquisition gradient-echo (MP-RAGE) sequence was also acquired for all subjects [160 slices, slice thickness = 1 mm, TR = 2,300 ms, TE = 2.98 ms, flip angle = 9°, matrix = 256 × 256, field of view = 256 mm, inversion time (TI) = 900 ms].

After participants were positioned in the scanner, we acquired a diffusion tensor imaging scan for 10 min while subjects were completing the practice run. We then collected four runs of the task (each lasting about 10 min), followed by a 10-min resting state scan and a 5-min structural scan. The diffusion and resting state scans were not used in the current analyses. The scanner collected two dummy volumes before EPI recording began, but we additionally discarded the first two acquired volumes to ensure scanner equilibration.

fMRI Preprocessing. All analyses were performed using SPM8 (Wellcome Department of Imaging Neuroscience, London). Preprocessing consisted of converting the raw DICOM images to NIFTI format, slice timing correction to the onset of the first slice, realignment through rigid-body registration to correct for

head motion, coregistration of the functional and anatomical images, segmentation of the anatomical image, normalization to Montreal Neurological Institute space using the gray matter image obtained from the segmentation, interpolation of functional images to 2 × 2 × 2 mm, and smoothing with a Gaussian kernel with a full-width at half-maximum of 4 mm for the individual-level analyses used for defining the ROIs and 8 mm for the group-level analyses.

Regressors for the first-level analysis of evoked activity were obtained by convolving the unit impulse time series for each condition with the canonical hemodynamic response function.

A first model was used to identify the frontal cortex ROIs that would be targeted later with TMS. The model included four regressors, reflecting the combination of the attentional cue (valid/invalid) and the speed/accuracy instruction (fast/accurate); thus, each trial was encompassed by exactly one of these regressors. Each regressor's onset coincided with the onset of the cues on the corresponding trial and offset coinciding with subject's first response (mean duration = 1.81 s). We included six nuisance regressors related to head motion: three regressors related to translation and three regressors related to rotation of the head.

Defining ROIs for TMS Targeting. We used individual task activations to identify FEF and DLPFC in every participant based on the main effect of task. In particular, we used the contrast task > background (by taking the average activation of the four task regressors) using $P < 0.001$ uncorrected (for a few subjects, more conservative or liberal thresholds were used). This contrast did not consistently reveal activations in aPFC; therefore, we opted to define this site based on previous studies (4, 18, 39) that all converged on a very similar spot in the rostral frontal cortex. Finally, S1 was identified based on its known anatomical location in the postcentral gyrus.

All regions were defined in the right hemisphere of the frontal cortex, which is thought to be dominant in perceptual tasks (30). The putative FEF was defined as the site of maximal activation near the junction of superior frontal sulcus and ascending limb of precentral sulcus (16). The average coordinates across our subjects were [$x = 28$ (SD = 3.9), $y = -3.4$ (SD = 2.9), $z = 51.3$ (SD = 4.4)]. The putative DLPFC was defined in the midlateral frontal cortex (5). In cases in which more than one locus of activity was found in that area, we chose the site in the middle frontal gyrus. The average coordinates across our subjects were [40.2 (SD = 5.3), 35.7 (SD = 6.3), 27.5 (SD = 8)]. Finally, aPFC was defined anatomically at [27 53 25] based on the results in a study by Fleming et al. (4), where it was determined that this region is involved in generation of confidence ratings [note that similar coordinates were reported by Fleming et al. (18) and Yokoyama et al. (39)]. We created 3-mm spheres for each site and used them to guide the placement of the TMS coil.

Analyses of the Activity During Each Task Epoch. To determine the timing of the involvement of different frontal regions in the perceptual process, we created three additional models. For these analyses, we included all 21 subjects who were scanned, regardless of whether they completed all six sessions of the experiment. Each model included just two regressors: one corresponding to the part of the trial epoch of critical interest and one corresponding to the rest periods between blocks. In the "instruction epoch" model, the first regressor encompassed the 1-s period when the instructions were presented (onset coincided with cue onset; duration was always 1 s); in the "stimulus/perceptual judgment epoch" model, the first regressor encompassed the period from stimulus onset until the decision was made (i.e., the first button press; onset coincided with stimulus onset; duration varied on each trial: mean = 0.81 s, SD = 0.22 s), and in the "confidence epoch" model, the first regressor encompassed the period from when the decision was made until when the confidence was given (i.e., the second button press; onset coincided with first button press; duration varied on each trial: mean = 1.06 s, SD = 0.49 s). The same six motion regressors as above were also included in each model. The contrast of interest was "trial epoch" > rest, where trial epoch corresponds to the respective first regressor in each model. Whole-brain contrasts were threshold at $P < 0.05$ family-wise error–corrected, using a voxel-level threshold of $P < 0.01$ and cluster correction of 244 voxels according to simulations using AlphaSim. ROI analyses were performed by defining spheres of 5 mm around the average coordinates for FEF, DLPFC, and aPFC reported above.

The task that we used was optimized to encompass the maximum number of trials to increase the power for finding TMS effects. Thus, this task was not optimal for cleanly separating the activity during each task epoch. Nevertheless, it was critical that the same task was used in the scanner on day 2 as for the TMS sessions (days 3–6) because making changes to the task could have shifted the loci of activations, and thus could have resulted in suboptimal localization of the TMS targets. However, although we cannot be sure that a regressor for a given task epoch is not contaminated by the processes occurring during an adjacent task epoch, any pattern of difference that we find (such as what is reported in Fig. 4) between different brain regions cannot be due to this limitation.

Another potential limitation for this analysis is the possibility that different ROIs may have different hemodynamic response delays. Substantially longer delays for the more anterior sites in the frontal cortex may result in apparent activations that are biased for the later epochs, which is exactly the pattern of the observed results. Although this issue remains to be fully settled by studies that provide reliable estimates for the hemodynamic delays across areas in the frontal cortex, there are several reasons why we think that differences in the hemodynamic delay are unlikely to have had a major influence on our results. First, all our areas of interest in the frontal cortex receive blood supply from the same vascular system: the superior branch of the middle cerebral artery. Therefore, the differences in hemodynamic delay between these areas are likely to be small. Second, hemodynamic delay differences in early sensory and motor areas (the regions where such delays can be computed most precisely) tend to be up to half a second (40). As noted above, the differences between our areas of interest are likely to be substantially smaller than this value. However, given that each epoch lasted around 1 s, even differences on the order of half a second cannot account for our results.

As an additional control analysis, we included all three task epoch regressors described above in a single GLM. The results found were very similar to the main results from the separate GLMs (Figs. S2 and S3). Importantly, we note that the lack of independence between the GLM regressors for each task epoch would only decrease the difference in activity between different task epochs. Thus, the pattern of results that we found (activity in rostral regions peaking for later epochs) cannot be explained by correlations between regressors.

TMS. TMS was delivered with a Magstim Super Rapid Stimulator (Magstim Company Ltd.) connected to two booster modules, using a figure-of-eight coil with a diameter of 70 mm. We used a standard offline TMS sequence called cTBS that is theorized to reduce cortical excitability (15). cTBS involves delivering five bursts of three 50-Hz pulses every second for a total of 600 pulses over 40 s. The stimulation was delivered at 80% of the individual motor threshold (resulting in an average of intensity of 35.5%, SD = 2.7% of maximum stimulator output). No arm, leg, or other movement was elicited by the stimulation of any of our targeted sites.

We determined the motor threshold immediately before each delivery of cTBS using a method similar to Rahnev et al. (41, 42). Briefly, we first used a "hunting procedure" to determine the location where suprathreshold single pulses induce maximal hand twitch. Then, starting at 30% of the maximum stimulator output,

we delivered single-pulse TMS until we reached the lowest intensity that resulted in motor-evoked potentials larger than 50 μV peak-to-peak in the targeted muscle on five of 10 consecutive trials. This intensity was chosen as the participant's motor threshold. Throughout this procedure and subsequent application of cTBS, the main axis of the coil was always oriented at 45° offset from the posterior-anterior direction with the coil handle extending posteriorly.

The exact site of stimulation was determined using a frameless stereotaxic localization system (Brainsight; Rogue-Research, Inc.). The coil was oriented such that the induced magnetic field was orthogonal to the skull. The experimenters were highly trained in TMS delivery and generally allowed less than 1 mm and 1° deviation from the target over the course of the 40-s stimulation.

Behavioral Analyses. Unless otherwise specified, analyses were performed in MATLAB. Our main question of interest was in investigating the effects of cTBS applied to different frontal sites on measures related to attention, speed/accuracy instructions, and confidence. As a measure of attention, we used the RT difference between attended and unattended stimuli:

$$\Delta_{attention} = \mathrm{RT}_{\mathrm{unattended}} - \mathrm{RT}_{\mathrm{attended}}.$$

The reasoning is that if TMS interferes with subjects' ability to deploy attention to the attended side, then the RT difference between unattended and attended stimuli will decrease. To avoid interactions with speed/accuracy instructions, only trials with instructions to be accurate were used.

In a similar fashion, our measure of successfully following the speed/accuracy (SA) instructions was the RT difference between trials with accuracy and speed focus:

$$\Delta_{SA} = \mathrm{RT}_{\mathrm{accuracy}} - \mathrm{RT}_{\mathrm{speed}}.$$

As above, to avoid interaction with attention, we analyzed only trials with valid cuing.

Finally, to measure the quality of the confidence ratings, we determined their correspondence to accuracy on a trial-to-trial basis (thus obtaining an estimate of subjects' metacognitive capacity). We used a widely used nonparametric measure, the Type 2 AUC, that plots confidence as a function of task accuracy (18). We further repeated these analyses using three alternative measures: meta-d′ (19), the correlation between confidence and accuracy [phi (20)], and the difference between confidence on correct and incorrect trials.

All measures were computed for each day of TMS stimulation for each subject. To determine how TMS affected each of the above measures, we applied linear mixed-effects analysis using the lme4 package (43) in R (R Core Team, 2012). TMS site was included as a fixed effect, whereas an intercept for subjects was entered as a random effect. *P* values were obtained for regression coefficients using the *car* package (44). Planned follow-up tests were performed using paired *t* tests.

Despite our methodology of using the data from day 1 to exclude subjects who did not perform the task well, the remaining subjects still exhibited a large variability in their $\Delta_{attention}$, Δ_{SA}, and metacognitive scores. For example, on day 2 (fMRI day), $\Delta_{attention}$ varied from −32 ms to 297 ms, whereas Δ_{SA} varied from 30 ms to 1,243 ms. Small values of these parameters indicate that the attentional or speed/accuracy instructions modulated subjects' responses only to a small extent. For such cases, it is impossible for TMS to produce a sizeable behavioral effect. To account for this issue, we weighted the parameter values obtained during the TMS days (days 3–6) by the corresponding parameter value on the fMRI day (day 2). Note that this procedure does not carry any inherent bias that will make any particular TMS site appear to produce higher or lower values compared with any other TMS site. All weights were positive, with the exception of a single $\Delta_{attention}$ value of −32, which was replaced by a 0. These weights, extracted from the corresponding parameter values on the fMRI day, were included in both the linear mixed-effects analyses and the paired *t* tests. The exact formulas used can be seen in the codes that we provided and are described in detail elsewhere (45). Briefly, the one-sample weighted *t* test is conducted by computing the weighted mean:

$$\mu_{weighted} = \frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$$

and weighted variance:

$$\sigma^2_{weighted} = \frac{\left(\sum_{i=1}^{n} w_i * x_i^2\right) * \left(\sum_{i=1}^{n} w_i\right) - \left(\sum_{i=1}^{n} w_i * x_i\right)^2}{\left(\sum_{i=1}^{n} w_i\right)^2 - \sum_{i=1}^{n} w_i^2},$$

then by computing the weighted SE:

$$SE_{weighted} = \frac{\sigma_{weighted}}{\sqrt{N}},$$

and, finally, by computing the *t* statistic:

$$t = \frac{\mu_{weighted}}{SE_{weighted}}.$$

The mixed-effects analyses are conducted in an equivalent manner.

Finally, to check for interactions between the site of TMS and measure affected, we performed another linear mixed-effects analysis on the data from all three measures. One difficulty for this type of analysis is that these measures were fundamentally different: Two were RT differences, and one was a metacognitive score. Therefore, to be able to compare these measures directly, we normalized the data from each measure (mean of 0 and SD of 1). We entered the TMS site, the measure type, and their interaction as fixed effects and an intercept for subjects as a random effect. Additionally, by-subject random slopes were included for the effects of TMS site and measure type.

Simulations. We found that TMS to each of our three different frontal sites affects a corresponding behavioral measure of interest. We sought to account for these effects using simulations based on a dynamic model of decision making that can account for choice, RT, and confidence. We adapted a simple model used by Kepecs et al. (21) and De Martino et al. (22), which is a variant of the "race models" (23) wherein separate accumulators code each possible outcome and the first accumulator that reaches a threshold determines the perceptual decision. Briefly, we modeled the stimulus on each trial as a normally distributed random variable $s(t) \in N(\mu_{tilt}, \sigma_{tilt})$, where the sign and magnitude of μ_{tilt} are determined by the actual stimulus orientation and the observed level of performance, respectively. The evidence (e) for left (L) and right (R) tilt accumulates according to the following simple rule:

$$e_{L/R}(t) = \int_0^t s_{L/R}(t)\, dt,$$

where

$$s_R(t) = \begin{cases} s(t), s(t) \geq 0 \\ 0, s(t) < 0 \end{cases} \text{ and } s_L(t) = \begin{cases} 0, s(t) \geq 0 \\ -s(t), s(t) < 0 \end{cases}.$$

The decision is made when one of the two accumulators e_L or e_R reaches a boundary θ. In this model, confidence is computed as

the difference Δe between the two accumulators at decision time $\Delta e = |e_R(t_\theta) - e_L(t_\theta)|$. Following De Martino et al. (22), we included an extra noise term in the confidence rating such that, on every trial, the reported confidence was drawn from a Gaussian distribution with a mean of Δe: conf = $N(\Delta e, \sigma_{conf})$. Intuitively, this noise term made the confidence rating less precise, thus reproducing the imperfect metacognitive performance exhibited by our subjects. We used the same parameters as Kepecs et al. (21) and De Martino et al. (22): $\sigma_{stim} = 1$, $\theta = 100$, and $dt = 1$. In addition, μ_{tilt} was set at 0.02 for unattended stimuli and 0.1 for attended stimuli, θ was set at 89 for speed instruction (the value of 100 above was used for the accuracy instruction), and σ_{conf} was set to 8. This model produced a continuous measure of confidence; to translate it to the 1–4 scale rating used in the experiment, we used additional criteria C such that a confidence rating of i was given when $C_{i-1} \leq \Delta e \leq C_i$, where $C_0 = 0$, $C_1 = 8$, $C_2 = 16$, $C_3 = 24$, and $C_4 = \infty$. The results we obtained were robust to a range of parameter magnitudes, and the exact values above should not be considered as quantitatively precise model fits.

We simulated 100,000 trials for each condition from the 4 (TMS site: S1, FEF, DLPFC, aPFC) × 2 (attended vs. unattended) × 2 (speed vs. accuracy instruction) design. The simulation of the results of S1 TMS used the parameters above. For the FEF TMS simulations, we used drift rates that were less different between the attended and unattended targets (exact values were 0.042 for unattended targets and 0.072 for attended targets). For the DLPFC TMS simulations, we used bounds that were less different between the accuracy and speed instructions (exact values were 90 for speed and 99 for accuracy instructions). Finally, for aPFC TMS simulations, we used lower confidence noise (we set σ_{conf} to 4). The results of these simulations are displayed in Fig. 3 *B–D*, which demonstrates our main point in performing these analyses: namely, that the effects of TMS to the three frontal sites correspond to different parts of the decision-making process. We emphasize that even though we chose the exact values of the parameters to produce an approximate fit to the average data, any similar manipulation would have produced the same pattern of results.

SI Results

Confidence leak was computed in similar fashion to experiments 3 and 4 in a study by Rahnev et al. (37) by taking the absolute value of the Fisher transform of the lag-1 confidence autocorrelation. As with the analyses on $\Delta_{attention}$, Δ_{SA}, and the metacognitive scores, we performed weighted t tests using as weights the confidence leak scores from the fMRI day. The results showed that confidence leak was significantly decreased after aPFC TMS compared with S1 TMS [$t(20) = 2.83$, $P = 0.012$], suggesting that TMS to aPFC may have increased the metacognitive scores, in part, by suppressing irrelevant processes, such as the ones that give rise to the phenomenon of confidence leak.

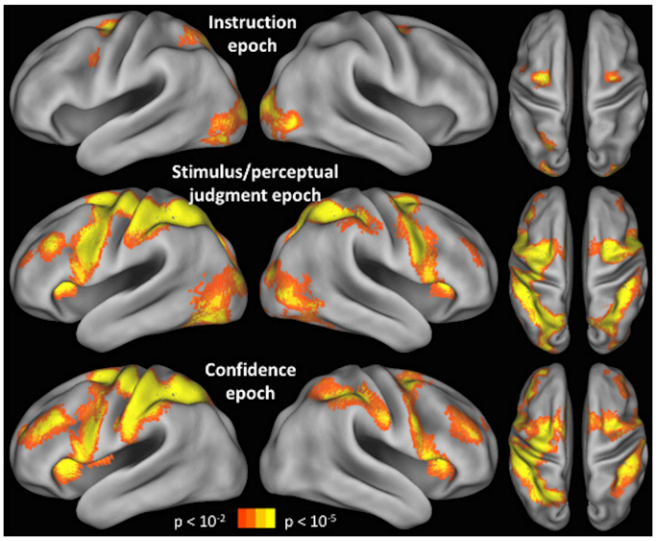

Fig. S1. Activations for each trial epoch. The activations during the instruction, stimulus/perceptual judgment, and confidence epochs are shown for the whole brain. Unthresholded maps are available on NeuroVault (*Materials and Methods*). The frontal cortex gradient is present in both hemispheres. In addition to the activations in the frontal cortex, a complex pattern of results is evident across the rest of the cortex. Part of this pattern can be explained by the amount of visual stimulation present on the screen during each epoch (e.g., the confidence period does not produce visual cortex activations because no stimuli were presented at that time). The activations in parietal cortex are interesting but are beyond the scope of the current investigation.

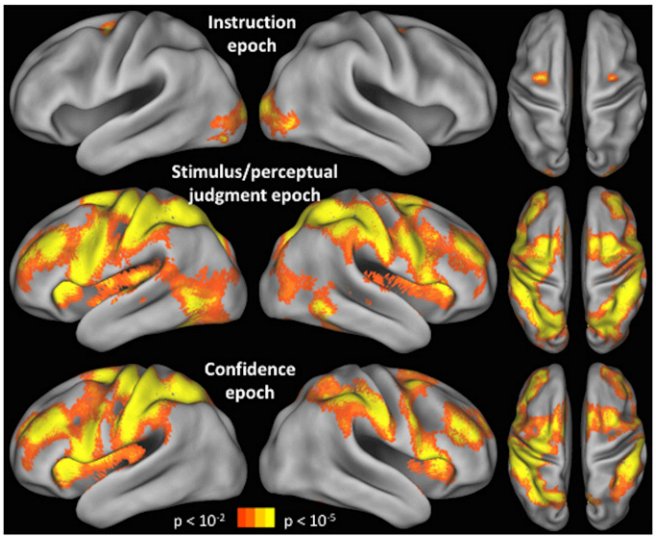

Fig. S2. Activations for each trial epoch in a single model. The activations during the instruction, stimulus/perceptual judgment, and confidence epochs are shown for the whole brain, but for a model in which a single GLM was used (Fig. S1 was produced by three separate GLMs corresponding to each task epoch). A similar pattern of results is revealed by this analysis.

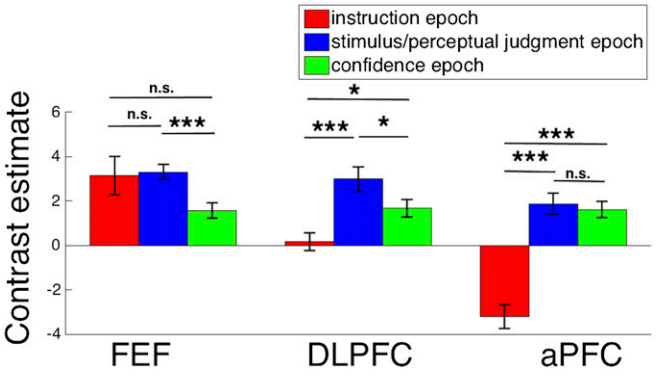

Fig. S3. ROI analyses in the single GLM model. The same ROI analyses as in Fig. 4*B*, but for the single GLM, are presented. As can be seen, the results are very similar and again point to a clear progression such that activity in progressively rostral regions peaks during progressively later epochs of the task. $*P < 0.05$; $***P < 0.001$. n.s., not significant.

Table S1. Performance on the task

Condition	S1	FEF	DLPFC	aPFC
RT				
Invalid, fast	658 (166)	686 (216)	619 (176)	637 (215)
Invalid, accurate	972 (385)	888 (316)	872 (318)	933 (482)
Valid, fast	579 (166)	590 (216)	558 (176)	540 (215)
Valid, accurate	811 (385)	829 (316)	735 (318)	798 (482)
Accuracy				
Invalid, fast	0.63 (0.12)	0.63 (0.07)	0.64 (0.10)	0.63 (0.10)
Invalid, accurate	0.63 (0.12)	0.64 (0.11)	0.65 (0.13)	0.61 (0.11)
Valid, fast	0.77 (0.12)	0.78 (0.07)	0.77 (0.10)	0.78 (0.10)
Valid, accurate	0.78 (0.12)	0.78 (0.11)	0.78 (0.13)	0.77 (0.11)
Confidence				
Invalid, fast	2.12 (0.42)	2.08 (0.34)	2.06 (0.39)	2.09 (0.43)
Invalid, accurate	2.34 (0.40)	2.41 (0.34)	2.33 (0.48)	2.33 (0.45)
Valid, fast	2.60 (0.42)	2.63 (0.34)	2.64 (0.39)	2.64 (0.43)
Valid, accurate	2.80 (0.40)	2.83 (0.34)	2.84 (0.48)	2.81 (0.45)

The table reports the average RTs (in milliseconds), accuracy (in proportion correct), and confidence for each of the four conditions: valid/invalid cue × fast/accurate instruction. The numbers in parentheses indicate SDs. Note that these SDs are between-subject, whereas all statistics are run within-subject. The data are pooled across targets on the left and right of fixation because no significant differences were found for presentation side.

Table S2. Metacognitive measures

Measure	S1	FEF	DLPFC	aPFC
Type 2 AUC	0.68 (0.09)	0.68 (0.09)	0.70 (0.09)	0.71 (0.09)
Meta-d′	1.31 (0.71)	1.26 (0.58)	1.43 (0.67)	1.50 (0.64)
Phi	0.28 (0.13)	0.28 (0.13)	0.30 (0.14)	0.33 (0.14)
Confidence difference	0.67 (0.35)	0.64 (0.36)	0.75 (0.37)	0.79 (0.42)

The table reports the results for four metacognitive measures: Type 2 AUC, meta-d′, phi, and the confidence difference between correct and incorrect trials. As can be seen, all four measures indicate the same pattern of results.