Behavioral/Cognitive

Limited Cognitive Resources Explain a Trade-Off between Perceptual and Metacognitive Vigilance

Brian Maniscalco,[1,2] Li Yan McCurdy,[3] Brian Odegaard,[4] and Hakwan Lau[4,5]

[1]Neuroscience Institute, New York University, New York, New York 10016, [2]Department of Psychology, Columbia University, New York, New York 10027, [3]Department of Biological and Biomedical Sciences, Yale University, New Haven, Connecticut 06520, and [4]Department of Psychology and [5]Brain Research Institute, University of California Los Angeles, Los Angeles, California 90095

Why do experimenters give subjects short breaks in long behavioral experiments? Whereas previous studies suggest it is difficult to maintain attention and vigilance over long periods of time, it is unclear precisely what mechanisms benefit from rest after short experimental blocks. Here, we evaluate decline in both perceptual performance and metacognitive sensitivity (i.e., how well confidence ratings track perceptual decision accuracy) over time and investigate whether characteristics of prefrontal cortical areas correlate with these measures. Whereas a single-process signal detection model predicts that these two forms of fatigue should be strongly positively correlated, a dual-process model predicts that rates of decline may dissociate. Here, we show that these measures consistently exhibited negative or near-zero correlations, as if engaged in a trade-off relationship, suggesting that different mechanisms contribute to perceptual and metacognitive decisions. Despite this dissociation, the two mechanisms likely depend on common resources, which could explain their trade-off relationship. Based on structural MRI brain images of individual human subjects, we assessed gray matter volume in the frontal polar area, a region that has been linked to visual metacognition. Variability of frontal polar volume correlated with individual differences in behavior, indicating the region may play a role in supplying common resources for both perceptual and metacognitive vigilance. Additional experiments revealed that reduced metacognitive demand led to superior perceptual vigilance, providing further support for this hypothesis. Overall, results indicate that during breaks between short blocks, it is the higher-level perceptual decision mechanisms, rather than lower-level sensory machinery, that benefit most from rest.

Key words: aPFC; metacognition; psychophysics; signal detection theory; vigilance; voxel-based morphometry

Significance Statement

Perceptual task performance declines over time (the so-called vigilance decrement), but the relationship between vigilance in perception and metacognition has not yet been explored in depth. Here, we show that patterns in perceptual and metacognitive vigilance do not follow the pattern predicted by a previously suggested single-process model of perceptual and metacognitive decision making. We account for these findings by showing that regions of anterior prefrontal cortex (aPFC) previously associated with visual metacognition are also associated with perceptual vigilance. We also show that relieving metacognitive task demand improves perceptual vigilance, suggesting that aPFC may house a limited cognitive resource that contributes to both metacognition and perceptual vigilance. These findings advance our understanding of the mechanisms and dynamics of perceptual metacognition.

Introduction

As an observer continuously performs a perceptual task, the observer's perceptual sensitivity tends to decline over time, an effect known as the vigilance decrement (Davies and Parasuraman, 1982; Warm, 1984; See et al., 1995). Research has suggested that limited cognitive resources (Kahneman, 1973; Matthews et al., 2000; Wickens, 2002) become depleted as a vigil progresses, and so the vigilance decrement is better accounted for by resource exhaustion than by mindlessness or task disengagement (Grier et al., 2003; Helton et al., 2005; Helton and Warm, 2008; Warm et al., 2008). Consistent with the resource depletion account, the

Received May 27, 2013; revised Nov. 29, 2016; accepted Dec. 9, 2016.

Author contributions: B.M. and H.L. designed research; B.M. and L.Y.M. performed research; B.M., L.Y.M., and H.L. analyzed data; B.M., L.Y.M., B.O., and H.L. wrote the paper.

This work was supported by Templeton Foundation Grant 6-40689 (to H.L.) and National Institutes of Health Grant R01NS088628 (to H.L.). We thank Ai Koizumi, Floris de Lange, and Dobromir Rahnev for comments on a previous version of this manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Brian Maniscalco, Neuroscience Institute, New York University, 550 1st Avenue, MSB 462, New York, NY 10016. E-mail: brian@psych.columbia.edu.

DOI:10.1523/JNEUROSCI.2271-13.2016

 0270-6474/17/371213-12$15.00/0

vigilance decrement is exacerbated by increasing task demands such as stimulus degradation, rate of stimulus presentation, and memory load (See et al., 1995) and is associated with depleted ratings of energetic arousal, elevated reports of stress, and declines in cerebral blood flow velocity (Warm et al., 2008).

A seemingly unrelated line of research involves the relationship between perceptual sensitivity and perceptual metacognition (e.g., confidence ratings). Recent work has developed a signal detection theory (SDT) analysis of confidence ratings (Galvin et al., 2003; Maniscalco and Lau, 2012, 2014), allowing for a bias-free measure of metacognitive sensitivity (i.e., an observer's ability to discriminate between her own correct and incorrect judgments, regardless of her tendency to report high confidence). Of particular interest is how such measures of metacognition are related to perceptual performance. A tacit assumption of the classical SDT analysis of confidence rating data is that perceptual decisions and confidence rating are based on the same underlying process (Galvin et al., 2003; Macmillan and Creelman, 2005; Maniscalco and Lau, 2012, 2014), and this view has received some empirical support (Kepecs et al., 2008; Kiani and Shadlen, 2009; Kepecs and Mainen, 2012). Other findings suggest that metacognition is subserved by high-level prefrontal mechanisms and is therefore partially dissociable from perceptual performance (Fleming et al., 2010, 2014; Pleskac and Busemeyer, 2010; Rounis et al., 2010; McCurdy et al., 2013).

In the current work, we bring these two lines of research together by investigating the joint behavior of SDT measures of perceptual and metacognitive sensitivity over time, to mediate between two potential hypotheses regarding the source of perceptual and metacognitive abilities. If a single process generates perceptual and metacognitive decisions, we should expect declines in perceptual sensitivity to be associated with declines in metacognitive sensitivity (Maniscalco and Lau, 2012, 2014). Conversely, if distinct processes generate perceptual and metacognitive decisions, we might expect vigilance decrements in perception and metacognition to be dissociable. Importantly, according to SDT, for an ideal observer, perceptual performance should be related to metacognitive performance such that $d' = \text{meta-}d'$ (Maniscalco and Lau, 2012, 2014). However, deviations from this expectation caused by sampling error and suboptimal metacognitive performance are to be expected, so a comparison between real data and an SDT model assuming no differences between these two measures may result in spurious findings attributable to an overly conservative null hypothesis. Thus, to arbitrate between our two hypotheses, rather than comparing changes in meta-d' and d' to an SDT model assuming a perfect correlation between the two measures, we conducted Monte Carlo simulations based on actual data to generate the noisy changes in each measure we should expect to see if SDT were true, and we compared our empirical results to these expectations.

To anticipate, we find a robust effect whereby changes in perceptual and metacognitive sensitivity over time are weakly or negatively correlated, contrary to the strong positive correlation predicted by a single-process view of perception and metacognition. Voxel-based morphometry analysis suggests that this finding is mediated by a common cognitive resource housed in anterior prefrontal cortex (aPFC), a region previously associated with visual metacognitive sensitivity (Fleming et al., 2010; McCurdy et al., 2013). Consistent with this account, we find that alleviating metacognitive task demands reduces the perceptual vigilance decrement. Thus, perception and metacognition appear to be distinct processes that can differentially access limited cognitive resources.

Materials and Methods

Experiment 1: evaluating the characteristics of decline in perceptual and metacognitive abilities

Data from this experiment were originally reported by Maniscalco and Lau (2012).

Participants. Thirty Columbia University students (including both males and females) participated in the experiment. Participants gave informed consent and were paid $10 for approximately 1 h of participation. The research was approved by Columbia University's Committee for the Protection of Human Subjects. Four participants were omitted from data analysis. One exhibited perfect task performance. The other three used an extreme confidence rating (lowest/highest rating) >89% of the time, an extreme bias in reporting confidence that renders meaningful analysis of metacognitive sensitivity difficult.

Materials and procedure. Participants were seated in a dimmed room 60 cm away from a computer monitor. Stimuli were generated using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in MATLAB (MathWorks) and were shown on an iMac monitor (LCD, 24 inch monitor size, 1920 × 1200 pixel resolution, 60 Hz refresh rate).

On every trial, two stimuli were presented simultaneously, one 4° to the left of fixation and one 4° to the right (Fig. 1*A*). Stimuli were presented on a gray background for 33 ms. Each stimulus was a circle (3° diameter) consisting of randomly generated visual noise. The target stimulus contained a randomly oriented sinusoidal grating (2 cycles/°) embedded in the visual noise. After stimulus presentation, participants provided a forced-choice judgment of whether the left or the right stimulus contained a grating. After stimulus classification, participants rated their confidence in the accuracy of their response on a scale of 1 through 4. Participants were encouraged to use the entire confidence scale. If the confidence rating was not registered within 5 s of stimulus offset, the next trial commenced automatically. (Such trials were omitted from all analyses.) There was a 1 s interval between the entry of the confidence rating and the presentation of the next stimulus. Participants were instructed to maintain fixation on a small crosshair (0.35° wide) displayed in the center of the screen for the duration of each trial.

At the start of each experimental session, participants completed two practice blocks (28 trials each) and one calibration block (120 trials). In the calibration block, the detectability of the grating in noise was adjusted continuously between trials on the basis of the participant's task performance using the QUEST threshold estimation procedure (Watson and Pelli, 1983). Target stimuli were defined as the sum of a grating with Michelson contrast C_{grating} and a patch of visual noise with Michelson contrast C_{noise}. The total contrast of the target stimulus, $C_{\text{target}} = C_{\text{grating}} + C_{\text{noise}}$, was set to 0.9. The nontarget stimulus containing only noise was also set to a Michelson contrast of 0.9. The QUEST procedure was used to estimate the ratio of the grating contrast to the noise contrast, $R_{g/n} = C_{\text{grating}}/C_{\text{noise}}$, which yielded 75% correct performance in the 2AFC task. Three independent threshold estimates of $R_{g/n}$ were acquired, with 40 randomly ordered trials contributing to each, and the median estimate of these was used to create stimuli for the main experiment. The main experiment (1000 trials) consisted of 10 blocks of 100 trials each, with a self-terminated rest period of up to a 1 min between blocks.

Experiment 2: evaluating the link between behavioral results and cortical characteristics

Data from this experiment were originally reported by McCurdy et al. (2013).

Participants. Forty-one Radboud University students (19 males, 22 females) participated in the experiment. Participants gave informed consent and were paid €8 for approximately 1 h of participation. The research was approved by the local ethics committee where the experiment was performed (CMO region Arnhem-Nijmegen, The Netherlands).

Materials and procedure. The experimental design was identical to Experiment 1, with the following exceptions. Blocks of the visual perception task were interleaved with blocks of a memory task. [Comparison of visual and memory task performance was explored by McCurdy et al. (2013); data from the memory task are not analyzed in this study.] Each participant completed two experimental sessions on 2 consecutive days. On day 1, participants completed two practice blocks of the visual task, a calibration block for the visual task, and two blocks of the visual task

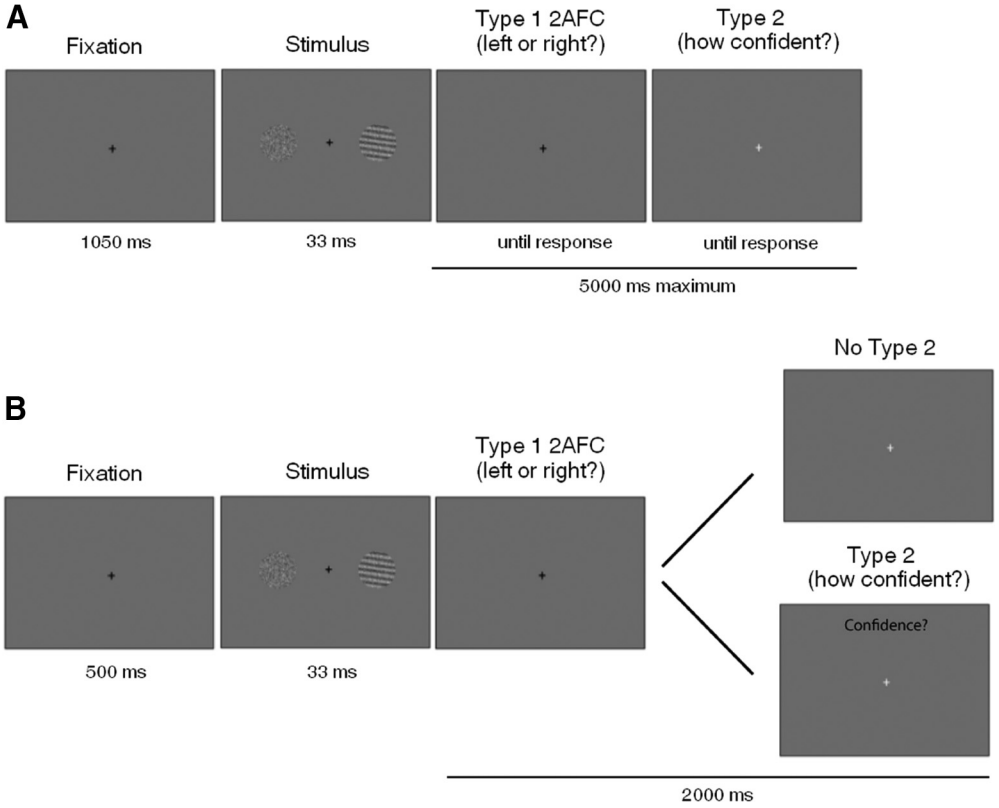

Figure 1. Design for Experiments 1–4. ***A***, In Experiments 1 and 2, subjects performed a spatial two-alternative forced-choice task. On each trial, two patches of visual noise simultaneously appeared to the left and right of fixation. One of these patches contained an embedded sinusoidal grating. Subjects first indicated whether the left or right patch contained the grating and then rated decision confidence on a scale of 1–4. Trial duration was determined by subject response time. ***B***, Experiment 3 was similar to Experiments 1 and 2, except that in even-numbered blocks of trials (partial type 2 blocks), subjects were not required to rate confidence for the first half (50 trials) of the block. A written cue appeared above fixation on all trials where subjects were required to rate confidence. Trial duration was fixed to be 2.533 s. In Experiment 4, subjects wagered points rather than rating confidence, such that they won or lost the number of points wagered depending on the accuracy of the left/right decision. Subjects were also provided with feedback about wagering performance after each block.

consisting of 102 trials each. On day 2, participants completed three more blocks of the visual task, using the stimulus settings acquired from the calibration block on day 1. In total across the 2 d, data were collected for 510 trials (five blocks of 102 trials each). As with Experiment 1, trial duration for the visual task was determined by response times, and participants experienced a self-terminated rest period of up to 1 min between blocks.

Rather than using a single value for the ratio of grating and noise contrast ($R_{g/n}$), as in the previous experiment, three different levels of $R_{g/n}$ were used across trials. The calibration block determined the highest level of $R_{g/n}$, $R^*_{g/n}$, and the two lower levels of contrast ratio were determined by multiplying $R^*_{g/n}$ by 0.75 and 0.5. In this study, all analyses for Experiment 2 collapse across contrast level to yield sufficient trials for SDT analysis.

Image acquisition. For 32 of the participants, a 1.5T Avanto MR-scanner (Siemens), using a 32-channel head coil, was used to acquire the T1-weighted anatomical MRI images (176 slices; echo time, 2.95 ms; TR, 2250 ms; voxel size, 1 mm isotropic). The remaining nine participants were scanned using different scanning parameters, and this was included as a covariate in the multiple regression design matrix in SPM8.

Voxel-based morphometry analysis. Voxel-based morphometry preprocessing was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm). Similar to the preprocessing protocol used by Fleming et al. (2010), the scans were first segmented into gray matter, white matter, and CSF in native space. DARTEL (Ashburner, 2007) was used to increase the accuracy of intersubject alignment by aligning and warping the gray matter images to an iteratively improved template. The DARTEL template was then registered to MNI space, and gray matter images were modulated such that their tissue volumes were preserved. Images were smoothed using an 8 mm full-width at half-maximum Gaussian kernel.

We conducted a whole-brain analysis and found a negative correlation between the change in metacognitive efficiency (meta-$d'_2 - d'_2$) − (meta-$d'_1 - d'_1$) and PFC at $p < 0.05$, where subscript indicates block half (e.g., d'_2 is d' computed from the second half of trials across all blocks). This initial whole-brain approach gave us reason to hypothesize that regions that have previously been implicated for sensory metacognition in prefrontal cortex would be involved. Thus, we then followed a previously established method, as described by McCurdy et al. (2013). We generated a T-statistic map reflecting the correlation between gray matter volume and meta-d'/d' and identified clusters using an initial threshold of $p < 0.001$ uncorrected. The resultant preprocessed gray matter images were analyzed using MarsBar version 0.42 software (marsbar.sourceforge.net). Small-volume correction was done to the clusters of interest by defining a 10 mm sphere of the two peak voxel coordinates identified by McCurdy et al. (2013) (peak voxel coordinate for left aPFC was [−12, 54, 16]; peak voxel coordinate for right aPFC was [32, 50, 7]; both survived cluster familywise error correction), and the gray matter volume in each sphere was calculated.

Experiment 3: investigating whether reduced metacognitive demand impacts perceptual vigilance

Participants. Twenty-one Columbia University students (including both males and females) participated in the experiment. Participants gave informed consent and were paid $10 for approximately 1 h of participation. The research was approved by Columbia University's Committee for the Protection of Human Subjects. One participant was omitted from data analysis because of using the highest confidence rating on 96% of all trials, an extreme bias in reporting confidence that renders meaningful analysis of type 2 data difficult.

Materials and procedure. The experimental design was identical to Experiment 1, with the following exceptions. The primary manipulation of

interest in Experiment 3 was that in even-numbered experimental blocks, participants did not provide confidence ratings in the first 50 of 100 trials in the block. We call these "partial type 2 blocks," as opposed to "whole type 2 blocks" in which confidence ratings were provided on all trials. Before each block, participants were instructed which kind of block was about to be presented. For partial type 2 blocks, the instruction read as follows: "Upcoming block: There will be NO CONFIDENCE RATING for the first 50 trials. Do not enter confidence ratings until you are prompted to do so." For whole type 2 blocks, the instruction read "Upcoming block: There will be confidence rating on EVERY trial" (Fig. 1*B*).

To clearly distinguish trials in which confidence ratings were and were not required, a text prompt reading "Confidence?" was displayed on every trial where confidence ratings were required. The prompt was displayed 6.4° above fixation.

Because some trials did not require confidence ratings, partial type 2 blocks would be shorter in duration than whole type 2 blocks if trial duration depended on participant response times, as it did in Experiments 1 and 2. Therefore, to standardize the temporal duration of the experiment, the duration of each trial and the duration of each break period were set constant. In Experiment 1, participants entered both the stimulus judgment and confidence rating in 2 s or less for 92% of all trials. Therefore, after each stimulus presentation in Experiment 3, there was a fixed response period of 2 s, during which participants had to enter the required stimulus and confidence responses. After the response period and before the next stimulus presentation, a crosshair was displayed for 0.5 s. Altogether, each trial lasted 2.533 s, a close match to the mean trial duration of 2.315 s in Experiment 1. Additionally, all break periods between blocks were set to 1 min. When only 10 s of break time were left, three auditory tones alerted participants to prepare for the upcoming block, and a timer counting down the remaining seconds of the break period was presented on the screen. In total, data were collected for 1000 trials (10 blocks of 100 trials each).

Experiment 4: adding incentive for metacognitive accuracy using a point-wagering system

Participants. Thirty-three Columbia University students (including both males and females) participated in the experiment. Participants gave informed consent and were paid $10 for approximately 1 h of participation. The research was approved by Columbia University's Committee for the Protection of Human Subjects. Six participants were omitted from data analysis because of using the highest point wager (see below) on 93% of all trials, an extreme bias in wagering that renders meaningful analysis of type 2 data difficult.

Materials and procedure. Experimental design was identical to Experiment 3, with the following exceptions. In Experiment 4, the confidence rating system was replaced with a point-wagering system. Participants were instructed that, after each stimulus identification response, they would sometimes be prompted to wager points on their stimulus decision. Participants could wager between 1 and 4 points. For correct trials, the number of wagered points was added to a running point tally, whereas for incorrect trials, the number of wagered points was subtracted from the tally.

Participants were instructed that their goal was to maximize the number of points they received over the whole course of the experiment. They were given the following guidelines for maximizing points: (1) get as many stimulus decisions correct as possible; (2) although the optimal wagering strategy is to wager 4 points for correct trials and 1 point for incorrect trials, the participant does not have perfect knowledge of which trials are incorrect, and high wagers for incorrect responses are costly. Thus, the optimal strategy is to wager points according to the best estimate of the likelihood that the stimulus response was correct, and so the entire wagering scale should be used to reflect variations in this estimated likelihood across trials. (We note that for an optimal observer performing at above chance, the true optimal strategy for maximizing points would be to always wager 4 points. However, subjects in the experiment used the whole point-wagering scale, as desired.)

For nonwagering trials, participants were instructed that correct trials would add 3 points to their tally and incorrect trials would subtract 3 points from their tally. Additionally, to incentivize participants to enter all required responses for each trial, they were informed that 10 points were subtracted from the tally for any trial where not all required responses were entered within the 2 s time limit.

During break periods, participants were provided with feedback on their wagering performance. They were shown how many points they had earned in the previous block, how many points they could have earned with an "optimal" wagering strategy (i.e., had they wagered 4 points for all correct responses and 1 point for all incorrect responses), and their overall wagering efficiency (the former quantity divided by the latter). The same information was provided for overall wagering performance across all blocks thus far completed. The text prompts used in Experiment 3 to inform participants which kind of block was about to come up, and to prompt them to enter wagers on trials where wagers were required, were the same in Experiment 4 except the word "confidence" was replaced by "wager." In total, data were collected for 1000 trials (10 blocks of 100 trials each).

Monte Carlo SDT simulations

We performed Monte Carlo SDT simulations to assess the extent to which observed changes in perceptual and metacognitive performance over time deviated from SDT expectation. In other words, since the expectation under SDT is that meta-$d' = d'$, it follows that under strict SDT expectation, Δmeta-$d' = \Delta d'$, where Δ indicates change across the first and second halves of a block of trials (e.g., $\Delta d' = d_2' - d_1'$). However, as a result of sampling error and suboptimal metacognitive performance, it is unreasonable to expect perfect equivalence between these measures. Therefore, to obtain an appropriate null distribution to compare our data against, we conducted Monte Carlo SDT simulations, which were structured so as to closely mirror key features of the empirical data across Experiments 1–4.

For each subject in Experiments 1–4, we binned together all trials occurring in the first half of an experimental block and computed d' (hereafter denoted d_1') and similarly binned together all trials occurring in the second half of an experimental block and computed d' (i.e., d_2'). (For Experiments 3 and 4, data were gathered only from blocks in which confidence ratings or wagers were provided on every trial.) Visual inspection of the scatterplot of d_2' versus d_1' suggested that these variables were roughly distributed as a bivariate normal distribution. Therefore, we computed the mean and covariance for d_1' and d_2' across all experiments and used a bivariate normal distribution with this mean and covariance as the basis for subsequent statistical sampling.

In Experiment 1, 500 trials contributed to each estimate of d_1' and d_2', whereas this number was reduced to 255 trials in Experiment 2 and 250 trials in Experiments 3 and 4 (after limiting the analysis to blocks where confidence ratings were provided on every trial). Therefore, in all simulations, 250 simulated "trials" contributed to the estimate of each SDT parameter for each simulated subject. Because the average number of subjects entered into the analysis for Experiments 1–4 was 28.5, each simulation contained data for 30 simulated subjects.

Simulation procedure. Simulations proceeded as follows. We simulated 2000 experiments, where each experiment had 30 simulated subjects, with a total of 500 simulated trials for each subject. For each subject, we first obtained "true" values for d_1' and d_2' by randomly sampling from the bivariate normal distribution described above. (If this resulted in any negative values, the sampling procedure was repeated until both d' values were positive.) These true d' values were used as the basis for subsequent sampling to obtain "simulated" values for d' and meta-d', as described below.

We also created a unique set of decision criteria for each subject. Decision criteria were initialized to values of -2, -1.75, -0.75, 0, 0.75, 1.75, and 2. To create different decision criteria for different simulated subjects, a small amount of random noise from $N(0, 0.5)$ was added to the initial values of the decision criteria. Decision criteria were then resorted to ensure they were in ascending order. Once the values of the decision criteria were determined for a simulated subject, these same criteria values were used for all simulated trials without any further variation.

For the first block half consisting of 250 trials, 125 simulated "S1" trials (corresponding to the experimental condition where the grating was on

the left) generated 125 sensory samples drawn from the normal distribution $N(-d_1'/2, 1)$. Another 125 simulated "S2" trials (corresponding to the experimental condition where the grating was on the right) generated 125 sensory samples drawn from $N(+d_1'/2, 1)$. (These reflect the normal distributions of sensory evidence contingent on stimulus presentation posited by SDT.) Each such sample was compared to the decision criteria, and this comparison determined the simulated subject's response for each trial (Macmillan and Creelman, 2005). Responses for the perceptual task could be either "S1" (i.e., grating was on the left) or "S2" (i.e., grating was on the right), and responses for the metacognitive task were a confidence rating ranging from values of 1 through 4. A similar procedure was used to simulate sensory samples and behavioral responses for the second block half of 250 trials.

Now that each trial was associated with a true stimulus configuration as well as the simulated subject's perceptual and metacognitive judgments, we were able to compute d' and meta-d' for the first and second block halves for each simulated subject using standard SDT analyses (Macmillan and Creelman, 2005; Maniscalco and Lau, 2012, 2014).

Modulation of simulated results using aPFC data from Experiment 2. Analysis of Experiment 2 suggested a model whereby aPFC gray matter volume is positively associated with both meta-d_1' and $\Delta d'$ (see Results and Fig. 5). To take these effects into account in the simulations, we used the following procedure. Using the data from Experiment 2, we applied a regression analysis to estimate the β values for the following equation:

$$\mathrm{aPFC}_{\mathrm{data}} = \beta_1 * \Delta d'_{\mathrm{data}} + \beta_0. \quad (1)$$

For the analysis of metacognitive performance, we defined the ratio of meta-d' to d' in the first block half as follows:

$$M_{1,\mathrm{data}} = \text{meta-}d'_{1,\mathrm{data}}/d'_{1,\mathrm{data}}. \quad (2)$$

Using data from Experiment 2, we applied another regression analysis to estimate the β values for the following equation:

$$M_{1,\mathrm{data}} = \beta_3 * \mathrm{aPFC}_{\mathrm{data}} + \beta_2. \quad (3)$$

On the basis of the β values obtained from the regression analysis in Equation 1 and the simulated values of $\Delta d'$, we assigned each simulated subject an aPFC volume:

$$\mathrm{aPFC}_{\mathrm{sim}} = \beta_1 * \Delta d'_{\mathrm{sim}} + \beta_0. \quad (4)$$

We then used the obtained value of $\mathrm{aPFC}_{\mathrm{sim}}$ to adjust the simulated subject's simulated value for M_1 as follows:

$$M_{1,\mathrm{adj}} = \beta_3 * \mathrm{aPFC}_{\mathrm{sim}} + M_{1,\mathrm{sim}}. \quad (5)$$

Since $\mathrm{aPFC}_{\mathrm{data}}$ was scaled in such a way that the mean value was 0, the coefficient β_2 derived from regression analysis in Equation 3 codes the mean value of M_1 in the data, which was 0.865. However, in SDT simulations, the mean value of M_1 was 0.996, consistent with the SDT expectation that meta-$d' = d'$ and therefore meta-$d'/d' = 1$ (Maniscalco and Lau, 2012, 2014). Thus, applying the β_2 coefficient to the simulated data would have been inappropriate. Instead, we replaced the β_2 coefficient (an estimate of the mean value of M_1 in the empirical data) with the actual value of M_1 derived for each simulated subject. This has the benefit of retaining natural between-subject sampling variation in the values of M_1 arising from the Monte Carlo simulation procedure when calculating the value for $M_{1,\mathrm{adj}}$.

Finally, we obtained a new value for meta-$d'_{1,\mathrm{sim}}$ using the following equation:

$$\text{meta-}d'_{1,\mathrm{adj}} = M_{1,\mathrm{adj}} * d'_{1,\mathrm{sim}}. \quad (6)$$

Results of the SDT + aPFC simulation displayed in Figure 7 are derived by taking the same simulated data used for the SDT simulation, with the exception that values of meta-$d'_{1,\mathrm{sim}}$ were replaced by the value of meta-$d'_{1,\mathrm{adj}}$ calculated for each simulated subject. This had the effect of modulating the simulated data set such that the relationships between simulated aPFC volume, $\Delta d'$, and meta-d_1' were similar to the relationships empirically observed in Experiment 2.

Analysis of correlations between Δmeta-d′ and Δd′. Each of the 2000 simulated experiments contained 30 simulated subjects, and so 30 values of $\Delta d'_{\mathrm{sim}}$ and $\Delta\text{meta-}d'_{\mathrm{sim}}$. For each simulated experiment, we calculated the Pearson's r correlation coefficient between $\Delta d'_{\mathrm{sim}}$ and $\Delta\text{meta-}d'_{\mathrm{sim}}$. To mitigate the influence of outliers, we excluded data from all simulated subjects with any d' value lower than 0.25 or higher than 3.

This resulted in 2000 simulated values for Pearson's r. We used these 2000 values to estimate the sampling distribution of r with and without the aPFC modulation of the SDT simulations, as displayed in Figure 7*C*. Estimates of the sampling distribution in turn allowed us to characterize the likelihood of the empirically observed r values in Experiments 1–4 under the null SDT model and the SDT model augmented by the aPFC findings.

Regression of Δmeta-d′ onto Δd′

For Experiments 1–4, one analysis of interest was to characterize the empirical relationship between $\Delta\text{meta-}d'$ and $\Delta d'$. Ideally, regressions between these variables should take into account that both are subject to sampling error. However, errors-in-variables approaches to regression typically require some knowledge or assumptions about the error structures of the dependent and independent variables.

We capitalized on the results of the Monte Carlo SDT simulations to characterize the error structures for these measures. As described above, for each simulated subject, we selected true values for $d_{1,\mathrm{true}}$ and $d'_{2,\mathrm{true}}$, and then repeatedly performed a sampling procedure using the SDT model parameterized with $d'_{1,\mathrm{true}}$ and $d'_{2,\mathrm{true}}$ to obtain corresponding simulated values $d'_{1,\mathrm{sim}}$, $d'_{2,\mathrm{sim}}$, meta-$d'_{1,\mathrm{sim}}$, and meta-$d'_{2,\mathrm{sim}}$. For each simulated subject, we calculated the sampling error for $\Delta d'$ as follows:

$$\mathrm{error}_{\Delta d'} = \Delta d'_{\mathrm{true}} - \Delta d'_{\mathrm{sim}}. \quad (7)$$

Likewise, since on the basic SDT model used here meta-$d' = d'$, it follows that $\Delta\text{meta-}d'_{\mathrm{true}} = \Delta d'_{\mathrm{true}}$. Thus, we calculated error for meta-d' as follows:

$$\mathrm{error}_{\Delta\text{meta-}d'} = \Delta d'_{\mathrm{true}} - \Delta\text{meta-}d'_{\mathrm{sim}}. \quad (8)$$

Sampling errors for $\Delta d'$ and $\Delta\text{meta-}d'$ were not correlated (Pearson's $r = -0.015$). Therefore, it was appropriate to use Deming regression to characterize their relationship (Deming, 1943). Deming regression requires knowing the value of the parameter δ, which is the ratio of the variances of error in the dependent and independent variables. On the basis of the simulation outcomes, we estimated that $\delta = \mathrm{var}(\mathrm{error}_{\Delta\text{meta-}d'})/\mathrm{var}(\mathrm{error}_{\Delta d'}) = 2.1535$. Therefore, for all regressions of $\Delta\text{meta-}d'$ onto $\Delta d'$ reported in this study, we used Deming regression with $\delta = 2.1535$.

Results

Within-session changes in meta-d′ and d′ are dissociable

Our behavioral paradigms across four experiments aimed to arbitrate between two possible hypotheses regarding the source of perceptual and metacognitive decisions: if a single process generates both perceptual and metacognitive decisions, declines in perceptual sensitivity should be associated with declines in metacognitive sensitivity, as vigilance decrements equally impact both domains. However, if distinct processes generate perceptual and metacognitive decisions, decrements in perception and metacognition should dissociate. As noted previously, an ideal observer according to SDT should exhibit meta-$d' = d'$, and it follows that under SDT expectation, $\Delta\text{meta-}d' = \Delta d'$. However, because of sampling error and suboptimal metacognitive performance, this perfect linear relationship is highly unlikely. Thus, we conducted Monte Carlo simulations based on the observed data to generate what one could reasonably expect if SDT assumptions were true. These simulations produced data reflecting our "null hypothesis" that, if supported, would indicate a single process likely generated both perceptual and metacognitive decisions. Alternatively, any violation of this trend would provide evidence for the alternative

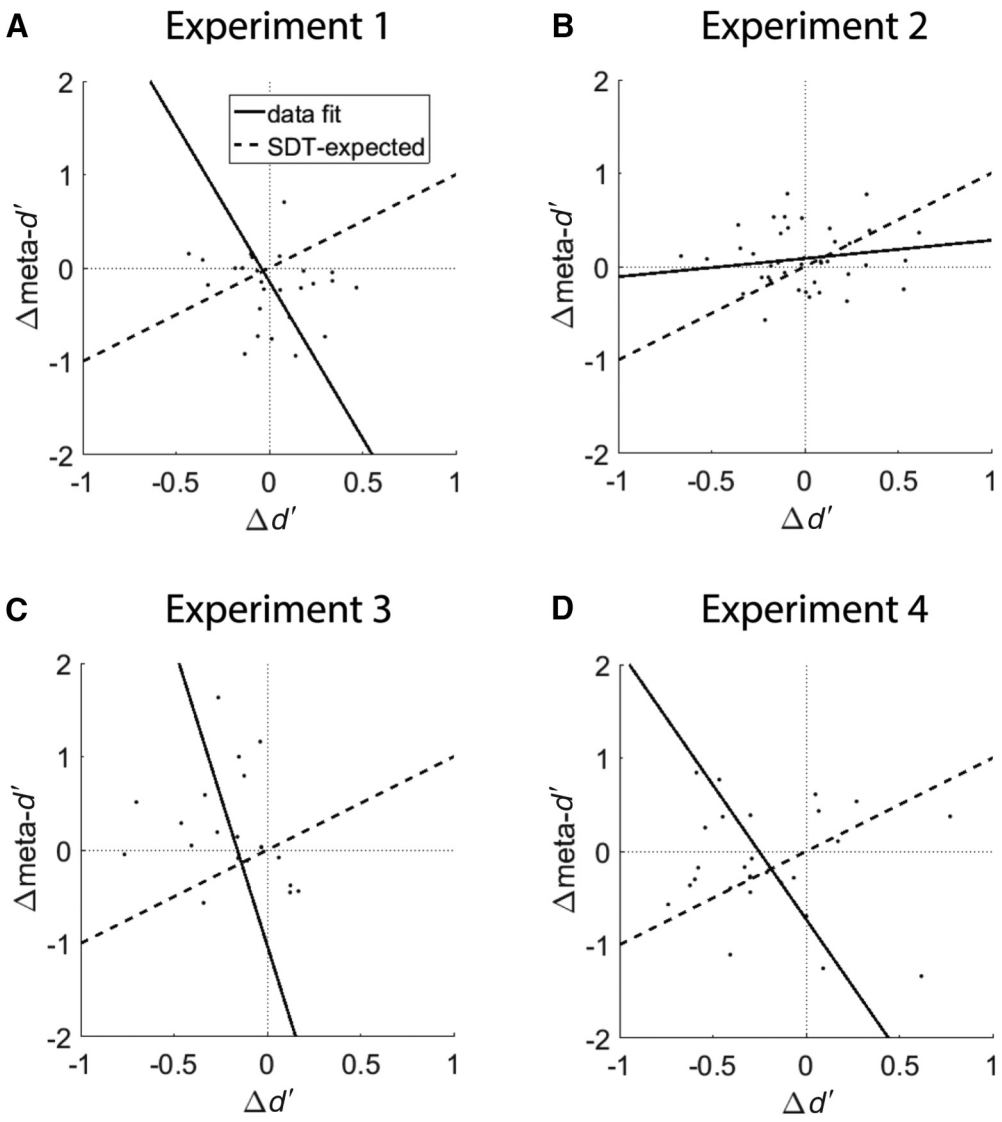

Figure 2. Individual results from Experiments 1–4 reveal a dissociation between SDT expectations and the relationship between changes in d' and changes in meta-d'. ***A***, Results from Experiment 1 displaying between-subject correlation of changes in perceptual and metacognitive performance and simulations based on SDT expectations. We computed the change in d' and meta-d' between the first and second halves of all blocks (i.e., $\Delta d' = d'_{\text{second block half}} - d'_{\text{first block half}}$; $\Delta\text{meta-}d' = \text{meta-}d'_{\text{second block half}} - \text{meta-}d'_{\text{first block half}}$) and found that these measures were inversely related, in stark contrast to SDT expectation. This suggests a trade-off effect whereby maintenance of perceptual performance comes at the expense of maintenance in metacognitive performance, and vice versa. ***B–D***, Data from Experiments 2–4. As in Experiment 1, the relationship between changes in d' and meta-d' failed to match SDT expectation.

hypothesis; namely, that distinct processes produced the perceptual and metacognitive decisions.

To assess the relationship between changes in perceptual task performance and changes in metacognitive performance, for each participant, we calculated d' and meta-d' using trials from the first and second halves of all blocks. We defined $\Delta d' = d'_2 - d'_1$ and $\Delta\text{meta-}d' = \text{meta-}d'_2 - \text{meta-}d'_1$, where subscripts indicate block half. As shown in Figure 2, across our four experiments, correlations between the changes in meta-d' and changes in d' from our behavioral data did not correspond to simulated data based on SDT expectations. Under the null hypothesis that changes in d' and meta-d' are generated by an SDT process, the observed Pearson's r correlation was 0.41 (Fig. 2, dotted lines). However, in three of the four experiments, changes in meta-d' were negatively correlated with changes in d' in the actual behavioral data, whereas in one experiment, they exhibited a weak, positive correlation that was much smaller than expected (Fig. 2, solid lines). Specifically, the observed Pearson's r correlations across our four experiments were the following: Experiment 1, $r = -0.18$ (Fig. 2*A*); Experiment 2, $r = 0.07$ (Fig. 2*B*); Experiment 3, $r = -0.22$ (Fig. 2*C*); Experiment 4, $r = -0.08$ (Fig. 2*D*).

Under the null hypothesis that changes in d' and meta-d' are generated by an SDT process with an expected $r = 0.41$, we estimate that the empirically observed correlation in experiment 1 ($r = -0.18$) corresponds to a one-tailed p value of 0.0015 (see Fig. 7*C*). Thus, according to SDT, the observed inverse relationship between $\Delta d'$ and $\Delta\text{meta-}d'$ is highly unlikely. The Deming regression slope relating $\Delta d'$ and $\Delta\text{meta-}d'$ was -3.12, lower than the SDT-expected value of 1. This trend held true in the other experiments: in Experiment 2, the empirically observed $r = 0.07$ corresponds to a one-tailed p value of 0.026 (see Fig. 7*C*). The Deming regression slope relating $\Delta d'$ and $\Delta\text{meta-}d'$ was 0.18, lower than the SDT-expected value of 1. In Experiment 3, $r = -0.22$, and in Experiment 4, $r = -0.08$; we estimate that the empirically observed values of $r = -0.22$ and $r = -0.08$ correspond to one-tailed p values of 0.001 and 0.004 in Experiments 3

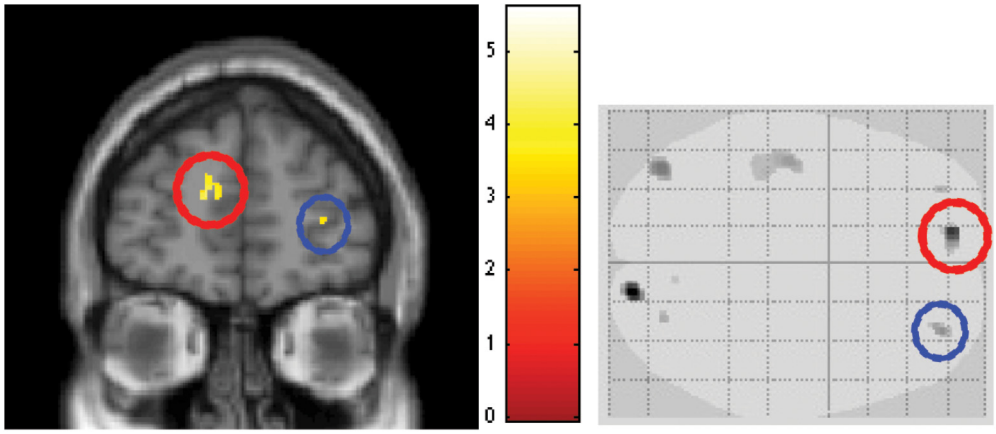

Figure 3. Brain regions selected for voxel-based morphometry analysis. Two regions of interest in aPFC were selected for analysis on the basis of positive correlations with metacognitive efficiency (meta-d'/d'). These regions were identified in a previous analysis of the data, conducted by McCurdy et al. (2013), and are consistent with previous findings relating metacognitive sensitivity to aPFC gray matter volume (Fleming et al., 2010). To obtain the most robust estimate of aPFC volume, we combined both aPFC clusters to produce an average volume, as described by McCurdy et al. (2013). The peak voxel coordinate for left aPFC is [−12, 54, 16]. The peak voxel coordinate for right aPFC is [32, 50, 7]. Both survived cluster familywise error correction. This figure is adapted from McCurdy et al. (2013).

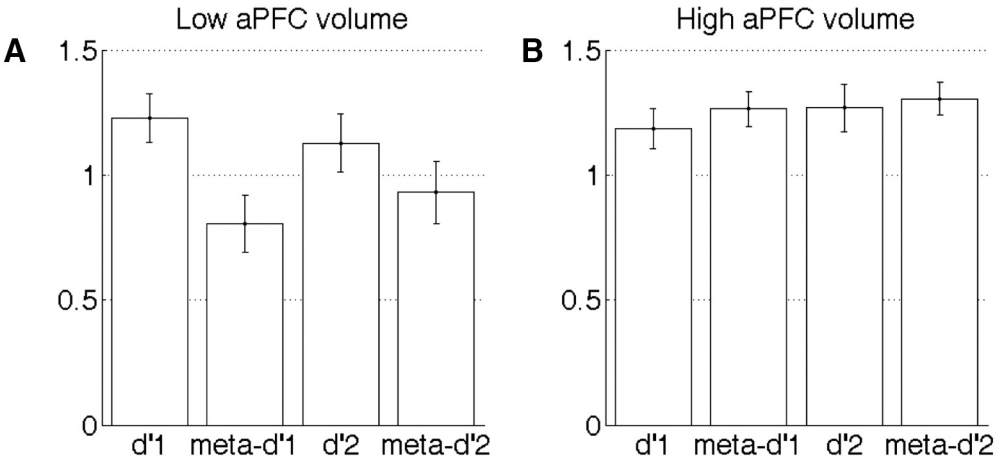

Figure 4. Perception and metacognition as a function of aPFC volume. A median split analysis revealed that subjects with lower aPFC volume (***A***) tended to experience decreases in d' and increases in meta-d' (task type × time × aPFC volume, $p = 0.03$), contrary to the pattern of subjects with higher aPFC volume (***B***). This suggests that the between-subject inverse relationship between changes in d' and meta-d' may be partially accounted for by individual differences in aPFC volume. Error bars represent within-subjects SEs (Morey, 2008).

and 4, respectively (see Fig. 7*C*). The Deming regression slopes relating $\Delta d'$ and Δmeta-d' were −6.22 and −2.29 in these two experiments, lower than the SDT-expected value of 1.

Differences in aPFC volume can explain the trade-off effect between $\Delta d'$ and Δmeta-d'

We also investigated whether the observed variability in metacognitive efficiency was correlated with interindividual differences in brain structure. In a previous study (McCurdy et al., 2013), we defined a measure of metacognitive efficiency on the visual behavioral task (Fig. 1*A*) as the ratio meta-d'/d'; for SDT-ideal observers, this ratio should equal 1, and for metacognitively suboptimal observers, it should be <1. Voxel-based morphometry analysis revealed that metacognitive efficiency was positively correlated with gray matter volume in regions in aPFC (Fig. 3; adapted from McCurdy et al., 2013). In the present study, we focused on the two regions in the aPFC identified by McCurdy et al. (2013) as regions of interest (ROIs; peak voxel coordinate for left aPFC was [−12, 54, 16]; peak voxel coordinate for right aPFC was [32, 50, 7]; both survived cluster familywise error correction.) The two clusters were used to define ROIs using the Marsbar toolbox (Brett et al., 2002), and gray matter volume in the aPFC clusters was calculated. To obtain the most robust estimate of aPFC volume, we combined both aPFC clusters in the region to produce an average volume, as described by McCurdy et al. (2013); all subsequent analyses refer to this combined data as aPFC.

To assess how aPFC volume influenced the trade-off effect, we performed a median split on aPFC volume and calculated d' and meta-d' over time for subjects with low and high aPFC volume (Fig. 4). A 2 (task type, type 1/type 2) × 2 (time, first block half/second block half) × 2 (aPFC volume, low/high) ANOVA revealed a significant task type × aPFC interaction ($p = 0.002$) and a significant task type × time × aPFC interaction ($p = 0.03$). On average, subjects with high aPFC volume did not exhibit decreases in d' or meta-d' over time (task type × time, $p = 0.7$) and were also metacognitively optimal in the sense that meta-d' was not significantly different from d' (task type, $p = 0.4$). By contrast, subjects with low aPFC volume were metacognitively suboptimal overall in the sense that meta-d' was significantly lower than d' (task type, $p = 0.002$). Crucially, low aPFC subjects also exhibited a numerical decrease in d' over time as well as an in-

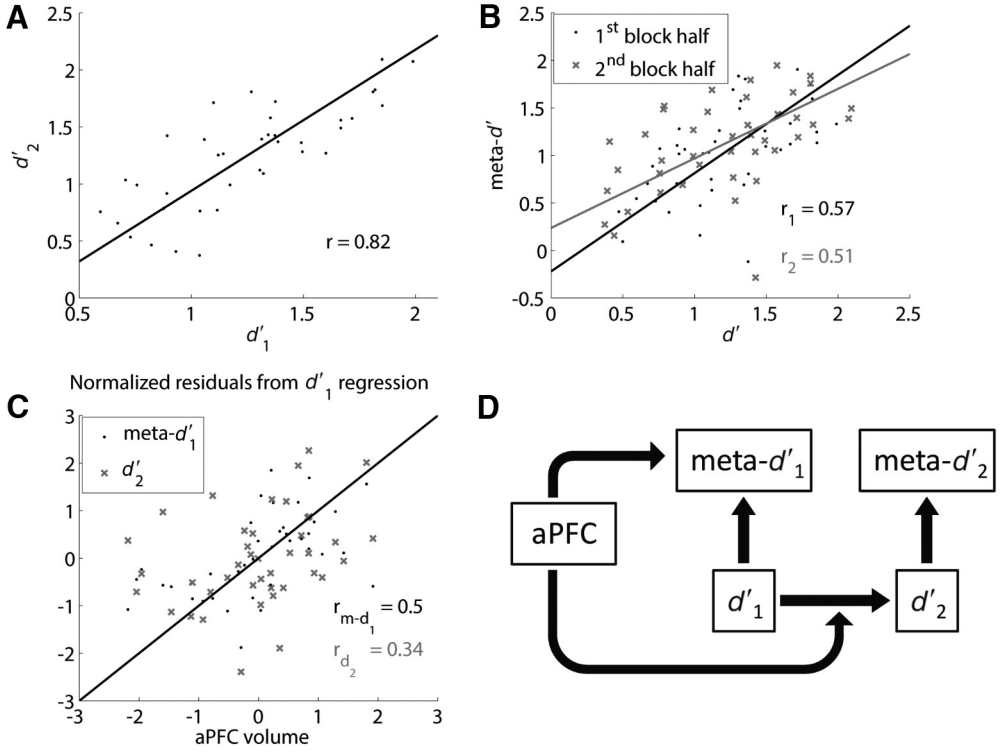

Figure 5. Model of the relationship between aPFC volume and changes in perceptual and metacognitive performance. ***A–C***, Correlation analyses from Experiment 2 reveal significant positive correlations between d' across block halves ($p < 0.001$) (***A***), meta-d' and d' within block halves ($p < 0.001$) (***B***), and aPFC volume with first-half meta-d' ($p = 0.001$) and second-half d' ($p = 0.03$) (***C***), after removing variation attributable to first-half d'. (Lines of best fit for both correlations overlap.) ***D***, A schematic representation based on the correlations exhibited in ***A–C***.

crease in meta-d', such that the interaction was significant (task type × time, $p = 0.01$). This pattern of changes in d' and meta-d' having the opposite sign for low aPFC subjects mirrors the trade-off effect exhibited in Experiments 1–4 (Fig. 2). Thus, individual differences in aPFC volume are a candidate mechanism to explain the observed trade-off effect between $\Delta d'$ and Δmeta-d'.

Additional correlations reveal the relationships between aPFC volume, d', and meta-d'

We further explored the relationship of aPFC volume to changes over time in d' and meta-d' by analyzing the patterns of correlation between d_1', meta-d_1', d_2', meta-d_2', and aPFC volume, using data from Experiment 2. As expected, d_1' and d_2' positively correlated (Pearson's $r = 0.82$, $p < 0.001$; Fig. 5*A*). Consistent with SDT expectation, meta-d' positively correlated with d' in each block half ($r = 0.57, 0.51$; $p < 0.001$; Fig. 5*B*). aPFC volume did not correlate with either d_1' ($p = 0.8$) or d_2' ($p = 0.2$), but a partial correlation between aPFC volume and d_2', controlling for d_1', was significant ($r = 0.33$, $p = 0.03$; Fig. 5*C*). Thus, larger aPFC volume was associated with better perceptual vigilance (higher $\Delta d'$).

aPFC volume was significantly correlated with meta-d_1' ($r = 0.43$, $p = 0.005$), and this correlation remained significant when controlling for d_1' ($r = 0.50$, $p = .001$; Fig. 5*C*). Although aPFC volume also correlated with meta-d_2' ($r = 0.33$, $p = 0.04$), this correlation did not remain significant when controlling for d_2' ($r = 0.26$, $p = 0.1$) or meta-d_1' ($r = -0.02$, $p = 0.9$). Indeed, although aPFC regions were selected on the basis of their correlation with overall meta-d'/d' ($r = 0.34$, $p = 0.03$), aPFC volume correlated with meta-d_1'/d_1' ($r = 0.51$, $p = .0006$) but not meta-d_2'/d_2' ($r = 0.1$, $p = 0.5$). Thus, aPFC volume robustly correlated with metacognitive sensitivity only in the first block half. The significant correlation between aPFC volume and meta-d_2' appears to be attributable to the fact that aPFC volume correlates with d_2', which in turn correlates with meta-d_2'. Because larger aPFC volume was associated with higher initial metacognitive sensitivity only, the sign of the correlation between aPFC volume and Δmeta-d' was negative (although nonsignificant; $r = -0.15$, $p = 0.3$). We also note that an evaluation of whether our two averaged aPFC ROIs independently predict overall metacognition (meta-d'/d') yielded the following results: left aPFC, $r = 0.39$, $p = 0.01$; right aPFC, $r = 0.27$, $p = 0.09$.

In Figure 5*D*, we present a simple schematic account to summarize these patterns of correlations. On this account, d' in the second block half depends heavily on initial d', and meta-d' in each block half is primarily a consequence of d'. Without further components, this account would be consistent with SDT expectation. However, there is an additional component corresponding to aPFC volume, and this factor contributes both to better initial metacognition and to better maintenance of perceptual performance over time. Larger aPFC is associated with larger meta-d_1' and therefore with smaller Δmeta-d'. Larger aPFC is also associated with larger $\Delta d'$. Since larger aPFC is associated with positive values for $\Delta d'$ and negative values for Δmeta-d', the contributions of aPFC appear to drive the deviation from SDT expectation encapsulated in the trade-off relationship between $\Delta d'$ and Δmeta-d'. (Also see below, *SDT simulations better characterize the data when taking into account the aPFC model*).

On this account, aPFC could be considered as a flexible cognitive resource that can contribute to both metacognitive monitoring and top-down control of perceptual task performance. To provide an additional test of this account, in Experiments 3 and 4 we included conditions where subjects did not have to provide metacognitive judgments in the first half of some experimental blocks. On this "resource" account, we might expect that when

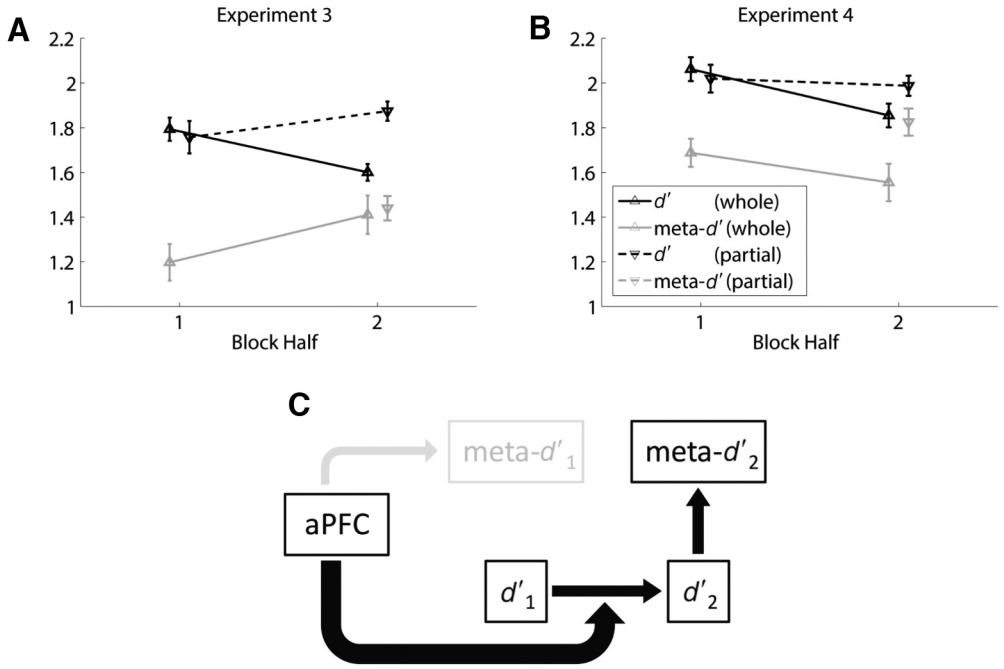

Figure 6. Results for Experiments 3 and 4. ***A***, ***B***, Mean perceptual (d') and metacognitive (meta-d') performance over time. When subjects were not required to place metacognitive judgments in the first block half (partial type 2 blocks), perceptual vigilance increased (block type × time interaction, $p = 0.002$), but metacognition in the second block half, as measured by meta-d_2'/d_2', was not affected (block type, $p > 0.4$). Error bars represent within-subjects SEs (Morey, 2008). ***C***, Resource account of findings. The results of Experiments 3 and 4 can be understood in terms of the model derived from Experiment 2. By relieving subjects of the requirement to place metacognitive judgments in the first block half, aPFC resources normally dedicated to initial metacognitive performance may have been spared for the separate task of maintaining perceptual vigilance.

subjects do not have the initial cognitive burden of placing metacognitive judgments, the resources shared by perceptual and metacognitive processes can be better applied to the task of maintaining perceptual vigilance.

In Experiment 3, we used a design similar to Experiment 1, with the primary difference that in even-numbered blocks, subjects were not asked to provide confidence ratings in the first half of each block (Fig. 1*B*). We call these blocks partial type 2 blocks, as opposed to the blocks in which metacognitive judgments are required on every trial, which we call whole type 2 blocks. According to a resource interpretation of the aPFC schematic (Fig. 5*D*), in the absence of the need to "boost" metacognitive performance, subjects should be better at maintaining perceptual performance over time in partial than in whole type 2 blocks (Fig. 6*C*). Experiment 4 was similar to Experiment 3 but used a point-wagering system with feedback on perceptual and metacognitive performance after each block. In both experiments, trial length was a constant 2.533 s, yielding blocks of a 253.3 s duration.

We tested whether the manipulation on task demand yielded the expected effect on perceptual performance over time. A 2 (block type, partial type 2/whole type 2) × 2 (time, first block half/second block half) × 2 (experiment, 3/4) mixed-design ANOVA on d' revealed a significant block type × time interaction ($p = 0.002$). The interaction is driven by the fact that $\Delta d'$ is smaller for whole type 2 blocks (mean, −0.20) than for partial type 2 blocks (mean, 0.04; Fig. 6*A*,*B*).

The block type × time × experiment interaction was not significant ($p = 0.4$), suggesting that the difference in $\Delta d'$ for whole and partial type 2 blocks is robust across Experiment 3 (where participants made metacognitive judgments by rating confidence) and Experiment 4 (where participants made metacognitive judgments by wagering points, were instructed to maximize points earned, and received performance feedback after each block). Thus, the observed decrement in perceptual performance is not attributable to lack of motivation or lack of a clear objective for how to perform the metacognitive task.

A 2 (block type, whole/partial) × 2 (experiment, 3/4) ANOVA yielded a nearly significant main effect of block type on meta-d_2' ($p = 0.053$), such that meta-d_2' was higher for partial type 2 blocks. However, the same ANOVA design shows that d_2' was also higher for partial type 2 blocks ($p < 0.001$), and so the larger value for meta-d_2' in partial type 2 blocks was likely mediated by the larger d_2' value. Indeed, the same ANOVA, when applied to the ratio meta-d_2'/d_2', did not reveal a main effect of block type ($p > 0.4$). Thus, the experimental manipulation on initial metacognitive demand did not influence metacognitive sensitivity in the second block half.

SDT simulations better characterize the data when taking into account the aPFC model

Finally, we performed additional Monte Carlo SDT simulations to computationally assess the empirical results in light of SDT expectation and to investigate whether the SDT model could yield a closer fit to the empirical data when taking into account the relationship between aPFC volume and task performance (Fig. 5*D*). (For additional details, see *Monte Carlo SDT simulations*.)

For each simulated subject, we defined the parameters of an SDT model specifying performance in the first and second block halves of a binary decision task with confidence ratings. SDT model parameters were sampled from distributions closely reflecting the statistical patterns in Experiments 1–4. Random samples were then drawn from the SDT models to generate a simulated value for $\Delta d'$ and Δmeta-d'. In all, we simulated 2000 experiments, each containing 30 simulated subjects. Consistent with strict SDT expectation, these simulations yielded a strong

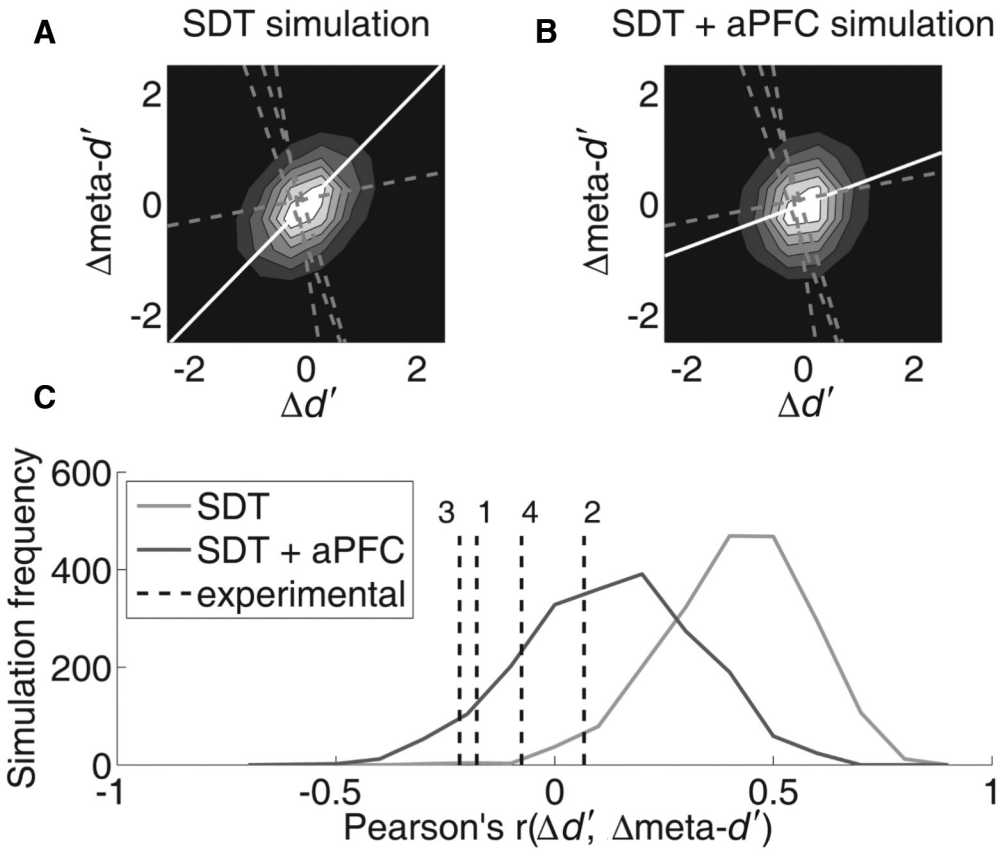

Figure 7. Signal detection theory simulations of the relationship between changes in perceptual and metacognitive sensitivity. ***A***, Basic SDT model. In a series of SDT simulations closely matching the properties of Experiments 1–4, changes in d' and meta-d' across block half are strongly positively related. Displayed is a contour plot based on the two-dimensional histogram of Δmeta-d' versus $\Delta d'$ for all simulated subjects in all simulated experiments. The white line is the line of best fit to simulated data; gray dashed lines are lines of best fit from data in Experiments 1–4. ***B***, SDT model with aPFC adjustment. We adjusted the outcomes of the initial SDT simulation so as to conform to the empirically observed relationships between aPFC volume, $\Delta d'$, and meta-d_1'/d_1' in Experiment 2 (see Materials and Methods for details). This substantially weakened the relationship between $\Delta d'$ and Δmeta-d' in the simulated data, as demonstrated by a more circular contour plot and smaller slope for the line of best fit. ***C***, Distributions of correlation coefficients for $\Delta d'$ and Δmeta-d'. Across 2000 simulated experiments, the basic SDT model yielded correlation values consistently higher than those observed in Experiments 1–4 (one-tailed p values of 0.002, 0.026, 0.001, and 0.004). The adjusted model incorporating the aPFC findings from Experiment 2 yielded a distribution of correlations more closely in line with the data (one-tailed p values of 0.067, 0.383, 0.043, and 0.161).

positive correlation between $\Delta d'$ and Δmeta-d' (Fig. 7*A*, white line). Next, we adjusted the initial simulation values for meta-d_1' on the basis of regression-estimated relationships between $\Delta d'$, meta-d_1', and aPFC volume in Experiment 2. This adjustment significantly attenuated the positive correlation between simulated values for $\Delta d'$ and Δmeta-d' (Fig. 7*B*, white line).

For each simulated experiment, we computed the Pearson's r correlation for $\Delta d'$ and Δmeta-d', yielding 2000 r values. The distribution of simulated r values under the SDT and SDT + aPFC models is displayed in Figure 7*C* alongside the empirically observed r values from Experiments 1–4. Under the SDT model, the distribution's mean value is 0.412 and only 0.9% of all values are lower than zero. Under the SDT + aPFC model, the mean shifts to 0.127 and 25.7% of all values are lower than zero, which is in better agreement with the data. For each empirical r value, we can compute a corresponding one-tailed p value using the r distribution for the SDT and SDT + aPFC models. The empirical r values from Experiments 1–4 are −0.18, 0.07, −0.22, and −0.08. Under the strict SDT model, these correspond to p values of 0.002, 0.026, 0.001, and 0.004. Under the SDT + aPFC model, these p values increase, on average, by a factor of about 30 to 0.067, 0.383, 0.043, and 0.161. Thus, the SDT + aPFC model is considerably better in accommodating the observed patterns of correlation between $\Delta d'$ and Δmeta-d' than is the standard SDT model.

Discussion

In summary, across four experiments, we find a robust trade-off effect whereby changes in perceptual and metacognitive sensitivity within a block of trials are negatively or weakly correlated. This finding contradicts the strong positive relationship predicted by single-process SDT and instead provides evidence for a dual-process model. Voxel-based morphometry analysis suggests that this trade-off effect may be explained by the contribution of neural resources in aPFC. Consistent with this account, perceptual vigilance decrements are alleviated when subjects are not required to provide metacognitive judgments in the first half of a block of trials.

The trade-off relationship between perceptual and metacognitive vigilance

The classical SDT model, which has enjoyed considerable success in modeling two-choice decision paradigms with confidence ratings (Macmillan and Creelman, 2005), predicts a strong, positive relationship between primary task performance and metacognitive performance (Galvin et al., 2003; Maniscalco and Lau, 2012, 2014). Thus, when considering the Pearson's correlation between $\Delta d'$ and Δmeta-d', we used SDT as the null hypothesis describing the expected distribution of correlation coefficients. We used Monte Carlo SDT simulations to construct the SDT-expected distribution of r values for $\Delta d'$ and Δmeta-d', which yielded a

distribution with a mean r value of 0.41 (see Materials and Methods, *Monte Carlo SDT simulations*, as well as Figs. 2 and 7*C*).

We found that most changes in d' and meta-d' across block halves failed to exhibit positive correlations, contradicting SDT expectation. Importantly, although the correlation coefficients for $\Delta d'$ and Δmeta-d' were small in magnitude, the relevant point of comparison is not with a distribution whose mean $r = 0$, but rather with the SDT distribution whose mean $r > 0$. The correlations in Experiments 1–4 significantly deviated from this SDT expectation. Thus, relative to the SDT-expected positive relationship, perceptual and metacognitive vigilance appeared to "trade off," such that improvement in one precluded comparable improvement in the other.

Interpreting the trade-off relationship

Research on the perceptual vigilance decrement has suggested that the decrement is caused by the depletion of limited cognitive resources (Grier et al., 2003; Helton et al., 2005; Helton and Warm, 2008; Warm et al., 2008). Experiment 2 of the present study suggests that regions of aPFC whose anatomical structure has previously been associated with metacognitive sensitivity in visual tasks (Fleming et al., 2010; McCurdy et al., 2013) may partially instantiate the resources supporting perceptual vigilance, since larger gray matter volume in these regions is associated with smaller declines in perceptual sensitivity.

The gray matter volume of aPFC was also associated with better metacognitive sensitivity during the first, but not second, half of each block (Fig. 5*C*). This may have driven a negative relationship between aPFC volume and Δmeta-d' in two ways. First, higher values for meta-d_1' would directly lead to lower values for Δmeta-d'. Second, according to SDT, meta-d' is theoretically constrained to be less than or equal to d' (Maniscalco and Lau, 2012, 2014). Therefore, all else being equal, better meta-d_1' leaves less room for meta-d_2' to improve, entailing a smaller maximum possible value for Δmeta-d'.

Thus, aPFC simultaneously exhibited a positive association with $\Delta d'$ and a negative association with Δmeta-d'. Subjects with larger aPFC exhibited strong perceptual vigilance (higher $\Delta d'$) as well as SDT-ideal metacognitive performance (meta-$d' = d'$; Fig. 4*B*). Conversely, subjects with smaller aPFC exhibited poorer perceptual vigilance (lower $\Delta d'$) and poorer initial metacognition (contributing to higher Δmeta-d'; Fig. 4*A*). In this way, individual differences in aPFC volume could produce the trade-off effect whereby $\Delta d'$ and Δmeta-d' failed to positively correlate (Fig. 2*A*,*C*,*D*).

One way of interpreting these findings is that perception and metacognition are subserved by separate processes that can independently tap into a common cognitive resource housed in aPFC. Presumably, as a block of trials wears on, resources would be increasingly allocated to the perceptual process (and thus away from the metacognitive process) to counteract the perceptual vigilance decrement (Fig. 5*D*). This account views perception and metacognition as separate processes that can draw on a common set of limited cognitive resources in a flexible manner, creating the potential for interference and competition for resources when both tasks are performed concurrently (Kahneman, 1973; Matthews et al., 2000; Wickens, 2002). More generally, this interpretation is consistent with accounts ascribing a broadly domain-general functionality to prefrontal cortex in guiding behavior (Koechlin and Summerfield, 2007; Badre, 2008; Passingham and Wise, 2012).

An alternative account is that since larger aPFC is associated with superior visual metacognition, the positive association between aPFC volume and perceptual vigilance could be mediated by superior metacognitive monitoring. Higher metacognitive sensitivity entails better ability to gauge ongoing perceptual performance, which could enable better ongoing regulation of task performance. On this account, aPFC is not a domain-general resource, but rather serves a specifically metacognitive function.

However, if better metacognitive monitoring directly contributes to superior perceptual vigilance, we might expect that perceptual vigilance should decrease when subjects are not required to engage in metacognitive monitoring. The resource account makes the opposite prediction; relieving the burden of placing confidence ratings should free up resources to support perceptual vigilance. In Experiments 3 and 4, we found that subjects were indeed more perceptually vigilant when not required to place confidence ratings in the first half of a block, more in line with the resource account than the metacognitive monitoring account. However, we take this result to be suggestive rather than decisive. Ultimately, these hypotheses will need to be further explored in future research. Additionally, we note that while we can only speculate as to what comprises the common resource affecting both measures, it seems plausible to hypothesize that mechanisms related to attention may underlie some of the effects observed in the four experiments.

Implications for models of metacognition

An active area of research concerns the relationship between perceptual and metacognitive processing. According to some accounts, seemingly complex and high-level functions such as metacognition and awareness actually bear simple and direct relationships to basic perceptual processing (Kepecs et al., 2008; Kiani and Shadlen, 2009; Kepecs and Mainen, 2012). The intuition behind these models is captured well by the conventional SDT model of confidence ratings, which characterizes perceptual judgments and confidence ratings as originating from the comparison of the same sensory information to different decision criteria (Macmillan and Creelman, 2005). Crucially, if perceptual and metacognitive judgments are different evaluations of the same underlying sensory information, then they should have similar informational content (formally, d' = meta-d'; Galvin et al., 2003; Maniscalco and Lau, 2012, 2014).

However, whereas the SDT model predicts a strong positive relationship between perceptual and metacognitive vigilance, we consistently observed this relationship to be neutral or negative. In our SDT-based simulations, we found that the empirical correlations between $\Delta d'$ and Δmeta-d' could not plausibly be accounted for by sampling variation under the SDT model (Fig. 7*A*,*C*). However, adjusting the simulation outcomes to reflect the mediating effect of aPFC volume on the behavioral measures entailed a theoretical outcome more in line with the data (Fig. 7*B*,*C*). In turn, the fact that aPFC volume had an opposite direction of association with $\Delta d'$ and Δmeta-d' suggests that perception and metacognition are separate processes with dissociable levels of sensitivity.

Why do we give subjects short breaks in perceptual experiments?

Although originally found in the context of long task durations (30+ min), the vigilance decrement has been shown to arise as early as the first 5–10 min of task performance (Nuechterlein et al., 1983; Temple et al., 2000) and to be dependent on factors such as overall perceptual sensitivity, rate of stimulus presentation, type of stimuli used, and memory load (See et al., 1995). Vigilance decrements are further associated with subjective effects such as

reduced arousal and elevated feelings of stress (Helton and Warm, 2008; Warm et al., 2008). Thus, a wide range of experimental tasks may be subjectively fatiguing and induce relatively rapid decrements in task performance.

In the current work, we found that perceptual (Experiments 3 and 4) and metacognitive (Experiment 1) vigilance decrements can occur even in experimental blocks of ~4–5 min in a fairly simple and standard visual discrimination task. Because we analyzed performance as a function of time across repeated blocks of trials, rather than analyzing the dynamics of task performance across a single prolonged block of trials, these results suggest a systematic pattern of performance decrements occurring within repeated blocks of trials that are nonetheless alleviated by regular intervals of rest.

What cognitive mechanisms benefit from the regular intervals of rest commonly used in perceptual experiments? The trade-off between perceptual and metacognitive vigilance found in Experiments 1–4, combined with the elevation of perceptual vigilance solely by relieving metacognitive task demand in Experiments 3 and 4, suggests the workings of a higher-level cognitive resource. The results of Experiment 2 identify aPFC as a contributor to this resource. Thus, our results suggest that rest primarily refreshes high-level cognitive resources, located at least partially in aPFC, rather than lower-level sensory mechanisms.

References

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. CrossRef Medline

Badre D (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn Sci 12:193–200. CrossRef Medline

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis10:433–436. CrossRef Medline

Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage 16:S497.

Davies DR, Parasuraman R (1982) The psychology of vigilance. London: Academic.

Deming WE (1943) Statistical adjustment of data. New York: Wiley.

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G (2010) Relating introspective accuracy to individual differences in brain structure. Science 329: 1541–1543. CrossRef Medline

Fleming SM, Ryu J, Golfinos JG, Blackmon KE (2014) Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. Brain 137:2811–2822. CrossRef Medline

Galvin SJ, Podd JV, Drga V, Whitmore J (2003) Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. Psychon Bull Rev 10:843–876. CrossRef Medline

Grier RA, Warm JS, Dember WN, Matthews G, Galinsky TL, Parasuraman R (2003) The vigilance decrement reflects limitations in effortful attention, not mindlessness. Hum Factors 45:349–359. CrossRef Medline

Helton WS, Warm JS (2008) Signal salience and the mindlessness theory of vigilance. Acta Psychol 129:18–25. CrossRef Medline

Helton WS, Hollander TD, Warm JS, Matthews G, Dember WN, Wallaart M, Beauchamp G, Parasuraman R, Hancock PA (2005) Signal regularity and the mindlessness model of vigilance. Br J Psychol 96:249–261. CrossRef Medline

Kahneman D (1973) Attention and effort. Upper Saddle River, NJ: Prentice-Hall.

Kepecs A, Mainen ZF (2012) A computational framework for the study of confidence in humans and animals. Philos Trans R Soc Lond B Biol Sci 367:1322–1337. CrossRef Medline

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455: 227–231. CrossRef Medline

Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324:759–764. CrossRef Medline

Koechlin E, Summerfield C (2007) An information theoretical approach to prefrontal executive function. Trends Cogn Sci 11:229–235. CrossRef Medline

Macmillan NA, Creelman CD (2005) Detection theory: a user's guide, Ed 2. New York: Cambridge UP.

Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious Cogn 21:422–430. CrossRef Medline

Maniscalco B, Lau H (2014) Signal detection theory analysis of type 1 and type 2 data: meta-d′, response-specific meta-d′, and the unequal variance SDT model. In: The cognitive neuroscience of metacognition (Fleming SM, Frith CD, eds), pp 25–66. Berlin: Springer.

Matthews G, Davies DR, Westerman SJ, Stammers RB (2000) Human performance: cognition, stress, and individual differences. Psychology. Philadelphia: Taylor and Francis.

McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, De Lange FP, Lau H (2013) Anatomical coupling between distinct metacognitive systems for memory and visual perception. J Neurosci 33:1897–1906. CrossRef Medline

Morey RD (2008) Confidence intervals from normalized data: a correction to Cousineau (2005). Tutorials Quant Methods Psychol 4:61–64. CrossRef

Nuechterlein KH, Parasuraman R, Jiang Q (1983) Visual sustained attention: image degradation produces rapid sensitivity decrement over time. Science 220:327–329. CrossRef

Passingham RE, Wise SP (2012) The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight. Oxford: Oxford UP.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442. CrossRef Medline

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

Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cogn Neurosci 1:165–175. CrossRef Medline

See JE, Howe SR, Warm JS, Dember WN (1995) Meta-analysis of the sensitivity decrement in vigilance. Psychol Bull 117:230–249. CrossRef

Temple JG, Warm JS, Dember WN, Jones KS, LaGrange CM, Matthews G (2000) The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. Hum Factors 42:183–194. CrossRef Medline

Warm JS (1984) An introduction to vigilance. In: Sustained attention in human performance (Warm JS, ed), pp 1–14. Chichester, UK: Wiley.

Warm JS, Parasuraman R, Matthews G (2008) Vigilance requires hard mental work and is stressful. Hum Factors 50:433–441. CrossRef Medline

Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. Percept Psychophys 33:113–120. CrossRef Medline

Wickens CD (2002) Multiple resources and performance prediction. Theoret Issues Ergon Science 3:159–177. CrossRef