

Neural Correlates of Perceived Confidence in a Partial Report Paradigm

Martín Graziano¹, Lucas C. Parra², and Mariano Sigman¹

Abstract

■ Confidence judgments are often severely distorted: People may feel underconfident when responding correctly or, conversely, overconfident in erred responses. Our aim here was to identify the timing of brain processes that lead to variations in objective performance and subjective judgments of confidence. We capitalized on the Partial Report Paradigm [Sperling, G. The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74, 1, 1960], which allowed us to separate experimentally the moment of encoding of information from that of its retrieval [Zylberberg, A., Dehaene, S., Mindlin, G. B., & Sigman, M. Neurophysiological bases of exponential sensory decay and top-down memory retrieval: A model. *Frontiers in Computa-*

tional Neuroscience, 3, 2009]. We observed that the level of subjective confidence is indexed by two very specific evoked potentials at latencies of about 400 and 600 msec during the retrieval stage and by a stationary measure of intensity of the alpha band during the encoding period. When factoring out the effect of confidence, objective performance shows a weak effect during the encoding and retrieval periods. These results have relevant implications for theories of decision-making and confidence, suggesting that confidence is not constructed online as evidence is accumulated toward a decision. Instead, confidence attributions are more consistent with a retrospective mechanism that monitors the entire decision process. ■

INTRODUCTION

Subjective confidence indicates the degree to which a decision-maker considers a choice to be correct. Tasks, which dissociate objective performance and subjectively perceived confidence, present an ideal tool to identify neurophysiological markers of confidence judgments (Maniscalco & Lau, 2012; Graziano & Sigman, 2009; Lau & Passingham, 2006). Different experimental situations may lead to distorted confidence judgments, for instance, when participants respond correctly with low confidence, as in blindsight or in implicit tasks (Merikle, Smilek, & Eastwood, 2001), and also in high confidence errors—for instance, in detection tasks that lead to confounding of targets and distractors (Graziano & Sigman, 2009; Baldassi, Megna, & Burr, 2006). This is not just a reflection of noise in the confidence estimate but instead reflects a robust difference in the signals used for the commitment to a choice and for the construction of subjective confidence (Zylberberg, Barttfeld, & Sigman, 2012).

Several studies point to the potential importance of PFC in regulating the accuracy of subjective confidence estimates (Fleming & Dolan, 2012). Lesion and inactivation (with TMS) of dorsolateral PFC can impair human

ability to estimate the accuracy of choice without affecting task performance (Rounis, Maniscalco, Rothwell, Passingham, & Lau, 2010; Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009). Individuals with frontal lobe lesions are impaired at making confidence judgments despite normal objective performance (Kennedy & Yorkston, 2000; Vilkkki, Surma-aho, & Servo, 1999), and in normal participants, the introspective ability correlates with gray-matter density in rostralateral PFC (BA 10; Fleming, Huijgen, & Dolan, 2012; Fleming, Weil, Nagy, Dolan, & Rees, 2010; Yokoyama et al., 2010). Studies of mnemonic metacognition have consistently shown that high-confidence decisions are associated with greater activity in the medial temporal lobe, the anterior cingulate (AC), and medial prefrontal regions (Chua, Schacter, & Sperling, 2009; Kim & Cabeza, 2007; Chua, Schacter, Rand-Giovannetti, & Sperling, 2006). Animal studies have found a link between firing rate of neurons and the degree of choice certainty in the parietal cortex (Kiani & Shadlen, 2009), OFC (Kepecs, Uchida, Zariwala, & Mainen, 2008), and frontal cortex (specifically, supplementary eye field area; Middlebrooks & Sommer, 2012).

We previously reported that a partial report paradigm can be used to generate strong dissociations between objective responses (the capacity to report the correct item in a multiple-choice task) and subjective responses (the conscious perception of the participant about its performance; Graziano & Sigman, 2009). In this work,

¹University of Buenos Aires and CONICET, Buenos Aires, Argentina, ²The City College of New York

we capitalize on this finding to analyze the dynamics associated with the construction of subjective confidence in the brain. We recorded high-density EEG during a partial report task. The main focus of this experiment was to identify neurophysiological markers that selectively distinguish trials in which participants feel confident about their response, from those in which they feel uncertain. Specifically, in our paradigm, we investigate whether perceived confidence and objective performance modulate brain activity during the sensory encoding stage of the stimulus and/or the retrieval stage and whether both processes (decision-making and metacognition) have different timings in the brain. Prominent psychological theories have characterized “confidence” as a retrospective report (Petrusic & Baranski, 2003; Vickers, 1979) and modeled this as a top-down (or two-stage) process of conscious access (Insabato, Pannunzi, Rolls, & Deco, 2010; Zylberberg, Dehaene, Mindlin, & Sigman, 2009; Dehaene, Sergent, & Changeux, 2003; Chun & Potter, 1995). Consistent with this, we hypothesized that global markers of confidence should be mostly confined to the retrieval process.

METHODS

Participants

Nineteen native Spanish speakers (nine men) with a mean age of 24 ± 4 years (mean formal educational level = 17 ± 4 years) participated in this experiment. All participants were right-handed and reported normal or corrected-to-normal vision. All participants gave written consent to participate in this study.

Visual Stimuli and Procedure

Behavioral experiment was programmed using the Python programming language. In each trial, participants fixated on a cross at the center of the screen for 1000–1500 msec before stimulus presentation (time interval selected at random). A circular array of eight letters was presented during 153 msec (corresponding to 13 frames with a refresh rate of 85 Hz). Stimuli were displayed on a 19-in. screen (with a resolution of 800×600 pixels) placed at a distance of 73 cm. Font type was uppercase Times New Roman with a size of 1.2° . Letters were chosen randomly from the alphabet (26 symbols), without repetition. The eight letters were arranged on a circle, around the fixation point at an eccentricity of 5.2° . After a fixed delay of 753 msec (ISI), a red dot (0.1°) on an array of blue dots (with the same configuration of the letters but at an eccentricity of 5.5°) indicated the position of the target letter. The cue was presented for 153 msec. Participants had to report verbally the target letter after a waiting period of 1000 msec indicated by the occurrence of a short beep (880 Hz). The responses were recorded with the computer. Subsequently, participants had to report the confidence of their response on a visual analog

scale (a horizontal bar placed at the center of the screen composed of 13 marks and two labels: “0% Confidence” and “100% Confidence” [“0% Seguro” and “100% Seguro,” in Spanish]). The participants could move the mouse freely to select the appropriated response. We opted for a long ISI value to allow the analysis of late components elicited at the encoding stage, which could be possibly modulated by the factors analyzed in this work.

Each participant first completed a practice block of 80 trials. The main experiment was divided in six blocks of 80 trials each (total = 480 trials). In each block, all eight positions were randomly and uniformly selected at target positions. Participants were instructed to fixate at the center of the screen during the entire experiment and to report verbally the letter as fast as they could (after the beep), selecting between the 26 letters of the alphabet (forced choice). They were also instructed to stay still as best as they can, indicating also that, after the beep, they could freely move and blink their eyes before starting the next trial. A complete session lasted approximately 60 min (without considering the preparation time for the EEG measurement).

EEG Data Collection and Preprocessing

EEG was recorded at 512 Hz with a 128-electrode Active II Biosemi device (Amsterdam, The Netherlands). After the experiment, all signals were filtered with a finite impulse response low-pass filter with a cutoff frequency of 30 Hz and high-pass filtered with a high-pass cutoff frequency of 1 Hz. All signals were rereferenced to a common mean, and the mean baseline activity during 300 msec before stimulus presentation was subtracted from each trial and each electrode. We rejected trials with voltage exceeding $\pm 150 \mu\text{V}$ and electro-oculogram activity exceeding $\pm 70 \mu\text{V}$.

Data Analysis

Data analysis was performed using Matlab (The MathWorks, Natick, MA), the EEGLAB toolbox (UC San Diego), and the Brainstorm suite toolbox (Tadel et al., 2011). Behavioral data were analyzed given the two main factors used in this study: subjective confidence (the conscious perception of the participant about its response, a continuous measure between 0 and 1, normalized for each participant) and objective performance (the capacity to accurately report the cued letter—a categorical variable).

EEG Data: Between-subject Factorial Analysis

As the distribution of confidence was bimodal for most individual participants, it could be easily partitioned into high-confidence and low-confidence categories by visual inspection of the distribution (mean threshold = 0.49 ± 0.08). A partition of the distribution based on a median

split did not change any of the main results reported in this work. According to this, EEG data of each participant were divided in four conditions (two factors with two levels each): incorrect–low confidence (incorrect and uncertain trials), incorrect–high confidence (incorrect and certain), correct–low confidence (correct and uncertain trials), and correct–high confidence (correct and certain trials). We rejected participants who had fewer than eight trials in each condition ($n = 2$) to reduce the error of the estimated mean and the differences in the variance across conditions (incorrect–low conf.: 220 ± 70 ; incorrect–high conf.: 40 ± 30 ; correct–low conf.: 50 ± 20 ; correct–high conf.: 130 ± 70). In addition, to avoid possible effects of nonhomogenous variance in the data, we performed a nonparametric test (Friedman test with repeated measures) to analyze the statistical difference between levels for each factor (Laganaro, Morand, & Schneider, 2009; Murray, Camen, Andino, Bovet, & Clarke, 2006; Schmid et al., 2006). This test computes the χ^2 Friedman rank sum value for group differences of one factor after possibly adjusting effects of the other factor. To evaluate possible interaction effects between factors and levels, post hoc comparisons were performed using the pairwise Wilcoxon rank-sum tests in case of a significant result of the Friedman test. Multiple comparisons were corrected nonparametrically using shuffle statistic according to Maris and Oostenveld (2007) and Nichols and Holmes (2002). Clusters were defined as two or more spatially contiguous sensors in which the statistic used exceeded a chosen threshold. We selected a threshold of $p = .01$ for the Friedman rank-sum tests and a threshold of $p = .05$ for the pairwise Wilcoxon tests, given the number of comparisons computed in each test (electrodes \times time: 128×1278 for the Friedman Test and 128×101 for the Wilcoxon rank-sum test). We considered electrodes with a distance of 4 cm as neighbors (yielding and average of 9.3 ± 1.7 neighbors per channel), and we assume that a robust cluster should encompass at least four neighboring channels and 26 samples—corresponding to a time window of 50 msec, giving a minimum cluster size of $n = 104$. After that, a cluster-based statistic was calculated computing the sum of each statistic value of the sensors in a cluster. Subsequently, random shuffles of the labels (correct/incorrect, high/low confidence) were used to obtain the null distribution of the maximum cluster-level test statistic. Hence, the entire analysis (comparison test, thresholding, finding clusters, computing cluster-level statistics) was repeated for each randomization trial ($n = 5,000$). For a given cluster, the corrected p value was estimated as the proportion of the elements in the shuffle null distribution exceeding the observed cluster-level test statistic.

Single-trial Level Analysis

Single-trial analysis using linear regression was performed to seek for regions of significant correlation between ERP activity and the confidence level of the participant, using

the continuous measure of confidence (instead of the categorization high/low). We use a repeated sample methodology to improve the estimation of the linear regression, namely, estimating the regression coefficients with all samples within a time window (instead of the mean) and combining all trials. This increases the number of samples available to compute the regression coefficients, which is required here as the number of parameters (128) is large relative to the number of trials (<480). In addition, to disambiguate the effect of confidence and accuracy on ERP activity, we used only correct trials for the linear regression analysis (Parra, Spence, Gerson, & Sajda, 2005). Thus, for each participant, ERP activity at the 128 electrodes among a time window of 50 msec ($n = 26$) was regressed with the reported confidence through a leave-one-out algorithm, iterating all trials but one. The regression coefficients obtained in the train trials were used to compute an estimated confidence value from the activity of each test trial through a purely linear model. This was done independently for each time window. Finally, we computed the Pearson correlation coefficient between the observed and estimated confidence ratings (obtained through the leave-one-out procedure) for each time window and participant, allowing us to estimate the mean correlation coefficient across participants throughout the task. Figure 2D depicts the mean Pearson correlation coefficient $\pm SEM$ ($n = 17$). Statistical significance was estimated through multiple paired t tests against the null hypothesis of zero correlation, and the obtained p values were corrected with false discovery rate (FDR) < 0.05 . The discriminant components for each significant time were obtained through forward modeling as described previously, which can be used to interpret the anatomical origin of the discriminative ability (Haufe et al., 2014; Parra et al., 2005).

Frequency Analysis

After preprocessing, data for each trial were divided in two periods corresponding to encoding and retrieval (0–900 and 900–1800 msec, respectively). In each period and for each trial, data were Fourier transformed after removal of any linear trend in time. Frequency bins corresponding to theta (4–8 Hz), alpha (9–11 Hz), and beta (15–30 Hz) were squared and summed to obtain an estimate of power in each band for each trial and for each period. For each electrode, we averaged estimated band power over correct and incorrect trials as well as high- and low-confidence trials. These averages were then summed across all participants weighting them by the number of trials contributed by each participant. Effects are reported as the ratio of these trial-averaged band powers in decibels. Because the two factors of confidence and correctness are correlated, we repeated the analysis for one factor while controlling for the second. For instance, to contrast correct versus incorrect trials while controlling for confidence, we took the following

contrast: ratio of [average power of correct–high confidence trials + average power of correct–low confidence trials] divided by [the average power of incorrect–high confidence trials + average power of incorrect–low confidence trials]. We chose this nonparametric method to control for one factor because the data were non-Gaussian and ANOVA or repeated-measures ANOVA analyses were not conclusive. As before, these averages were weighted and summed across participants. To determine statistical significance of the results, we repeated the identical procedures after randomly shuffling the labels across trials (correct/incorrect and high/low confidence). For each of the 128 electrodes, we computed the p value as the fraction of permutations (among $n = 10,000$) with power ratios that were more extreme than what were observed with the original labels. All electrodes with a significant difference in band power are marked with a dot in Figure 3 (p values corrected with FDR < 0.05, $n = 128$).

Detection of Response Components

To capture the main components associated with the temporal signal obtained from the ERPs, we averaged the data across participants and conditions and used the absolute mean activity (among electrodes) to decompose the ERP activity in a sequence of response components (Sigman & Dehaene, 2008). Response components were detected through a peak detection analysis, assuming that the points of local maxima at this time series reflect the peak of each response component. Each local maximum had to exceed a minimum threshold voltage of 0.3 μ V to be considered a peak of a response component. Consequently, we obtained the latencies for the local maxima and, with these, the topography of each response component at the electrode space.

Source Localization Analysis

Source localization was done using the Brainstorm suite toolbox (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). Briefly, cortical current density mapping was obtained using a distributed model of 15,000 current dipoles. Dipole locations and orientations were constrained to the cortical mantle of a generic brain model built from the standard brain of the Montreal Neurological Institute (Colin 27 brain with a 1-mm resolution). EEG forward modeling was computed using a Berg three-shell single sphere analytical model (Mosher, Leahy, & Lewis, 1999). Cortical current maps were computed from the EEG time series using a linear inverse estimator (weighted minimum-norm current estimate; Baillet, Mosher, & Leahy, 2001). Paired t tests (p values corrected with FDR < 0.05) were used to analyze differences in the mean of the cortical current density activity at the encoding and the retrieval stage. In addition, several bilateral ROIs were selected from the main areas activated by the late ERP response

components of the retrieval phase: AC areas ($n = 130$), OFC (superior frontal gyrus and middle frontal gyrus [BA 10], $n = 1,170$), superior parietal lobe (precuneus, $N = 969$), middle temporal gyrus ($n = 1,250$), and occipital cortex ($n = 1,109$). Thus, we computed the averaged time series of the current density for each participant and ROI, grouping by subjective condition (high and low confidence). The statistical assessment was performed through a multiple paired t test analysis for each time and ROI, correcting for multiple comparisons through a nonparametric cluster-based randomization test. For each ROI, we clustered the data exceeding a threshold t statistic = 1.73 (two sided, $p < .05$) and considered a minimum cluster size of $n = 26$ (corresponding to a time window of 50 msec), in consonance with the previous analyses described above. The significance of each cluster was assessed computing the null distribution of the cluster-level test statistic through random shuffles ($n = 5,000$), and the corrected p value was estimated as the proportion of the elements in the randomization null distribution exceeding the observed cluster-level test statistic.

RESULTS

Experimental Design and Behavioral Analysis

The partial report task separates the phase of sensory encoding—starting with stimulus presentation—from the retrieval phase that begins after the subsequent presentation of a response cue. Because, from our prior studies, objective performance almost did not vary within an ISI of 500–900 msec (Graziano & Sigman, 2008, 2009) and because we hypothesized that late evoked potentials could be of interest, we selected an ISI of 750 msec to be sure that the evoked responses to the stimulus and to the cue would not overlap.

In each trial, participants saw a circular eight-letter array (STIM), which lasted 153 msec (Figure 1A). At a fixed ISI delay of 753 msec, a small red circle was presented at one of the eight locations of the array (selected at random). This cue indicated the position of the letter to be reported (CUE). The cue was presented for 153 msec and was placed at the same angular location but at a greater eccentricity than the array to minimize the possibility that it may induce masking of the target letter. Fixation had to be maintained on a cross at the center of the display for the duration of the trial. Participants reported the letter aloud after a waiting period of 1000 msec. This pause assured a sufficiently long temporal interval to record the EEG signal without contamination of voice artifacts (Figure 1A). After this initial response, participants reported their level of confidence in their response on a visual analog scale placed at the center of the screen. They were informed that the left extreme of the bar indicated 0% confidence, when they thought they were simply guessing, and the right extreme indicated 100% confidence, when they were completely certain of their

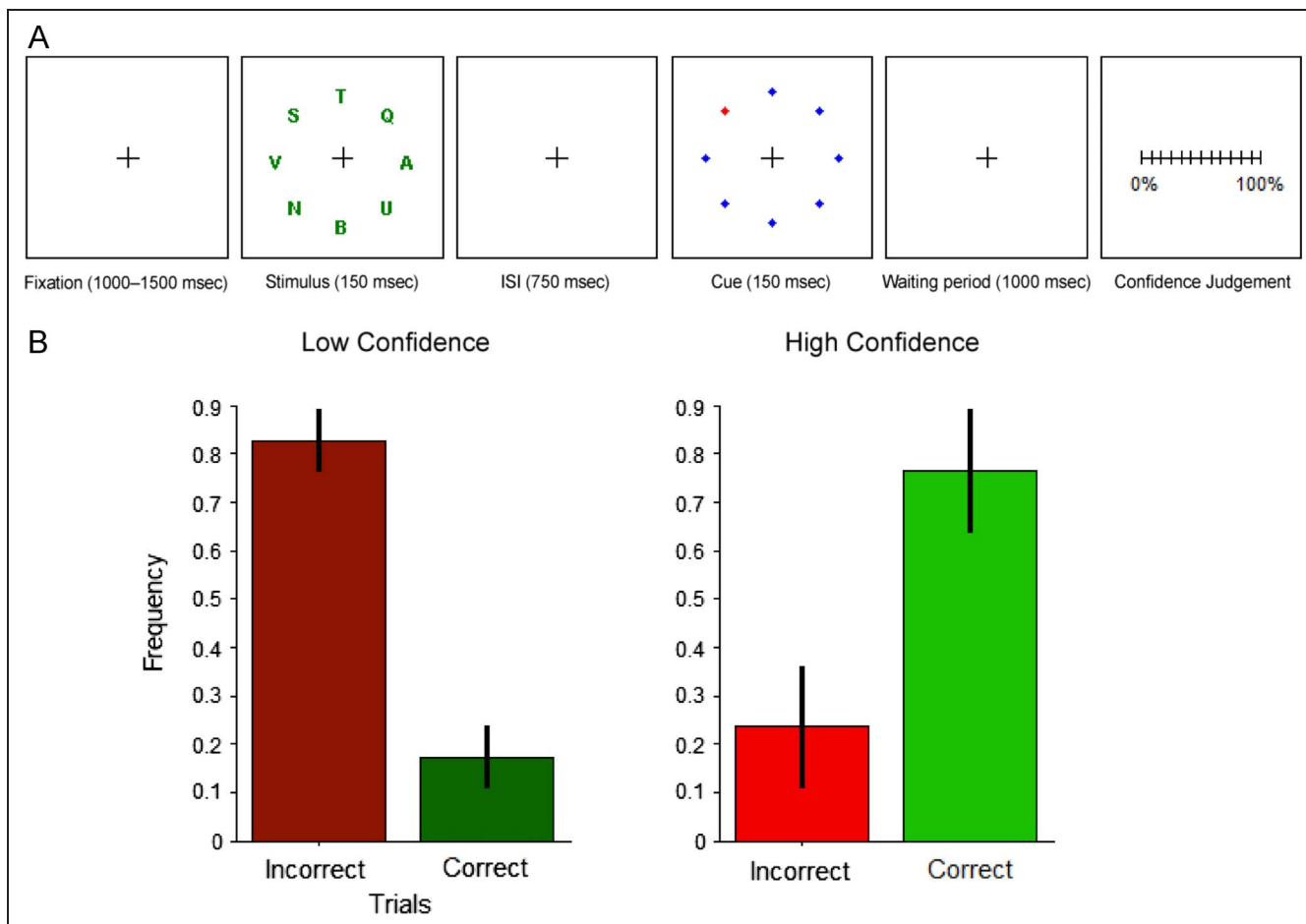


Figure 1. Experimental design and behavioral analysis of the objective and subjective reports. (A) A circular array of eight letters was presented during 153 msec after a waiting period of 1000–1500 msec. Participants fixated in a cross at the center of the array. A small red circle (the cue) was presented for 153 msec after a fixed delay of 750 msec (ISI) in one of the locations of the array indicating the letter that had to be responded. These two events (the stimulus and the cue) demarcate the encoding and retrieval stages for which ERPs were obtained. Participants had to report verbally the responded letter after a 1000-msec waiting period (since the offset of the cue) and the occurrence of a short beep and, subsequently, their confidence level through an ad hoc bar placed at the center of the screen. The response ranged between 0% of confidence (guessing) and 100% (completely certain; for more details, see Methods). (B) Mean frequency of trials across different report conditions. The plot shows the frequency of trials classified by the subjective response (the conscious perception of the participant about its response, divided in low-confidence trials and high-confidence ones) and the objective response (the capacity to report the correct letter). It can be shown that there is a reliable decoupling of the responses in all participants, which was used to infer the subjective confidence from EEG regardless of accuracy.

response. Nineteen participants completed an experimental session with 480 trials each.

The behavioral data obtained confirmed previous results observed with this paradigm (Figure 1B): low levels of objective performance at long ISI (0.4 ± 0.1 proportion of correct responses), corresponding to a memory load of 3.0 ± 0.9 items (Landman, Spekreijse, & Lamme, 2003; Pashler, 1988; false alarm rate was computed as in Graziano and Sigman, 2008) and a strong dissociation of behavioral measures (objective performance and subjective report) although mainly correlated (quantified by measuring the area under the receiver operating characteristic curve: 0.81 ± 0.01 ; Bartfeld et al., 2013; Graziano & Sigman, 2008, 2009). Most participants showed bi-modal distributions of subjective confidence; hence, it could be easily partitioned into high-confidence and

low-confidence categories by visual inspection of the distribution. Given that one of the main objective was obtain decoupled measures of objective and subjective performance, we excluded two participants who had insufficient low-confidence correct trials and high-confidence error trials, allowing a balanced analysis with confidence and accuracy as factors (Figure 1B; see Methods for details).

ERPs Are Modulated by Objective Performance and Subjective Confidence during the Retrieval Stage but not during the Encoding Stage

We locked the EEG signal to the onset of the stimulus and averaged across trials. This makes the resulting ERPs both stimulus-locked as well as cue-locked, because of

the fixed stimulus–cue delay. These ERPs capture thus the unfolding EEG signal after STIM and CUE, indicating the onset of the sensory encoding stage and the retrieval phase of the task, respectively (Figure 2A, dotted lines). The EEG data of each participant were divided in four conditions (two factors with two levels each): incorrect–low confidence (incorrect and uncertain trials), incorrect–high confidence (incorrect and certain), correct–low confidence (correct and uncertain trials), and correct–high confidence (correct and certain trials).

We first performed a nonparametric test (Friedman test) to compare the ERP magnitude (regardless of sign) for every time point and electrode while correcting for multiple comparisons (see Methods), assessing a group-level comparison for one factor after possibly adjusting effects of the other factor. The test revealed a main effect of confidence in two spatiotemporal clusters during the retrieval stage (black dots superimposed to the grand-averaged ERP, corrected cluster p value $< .0002$; Figure 2A). We did not observe significant clusters during the sensory encoding phase (corrected cluster p value $> .1$). Moreover, we did not find significant clusters after correcting for multiple comparisons for the main effect of performance possibly adjusting effects of subjective confidence (corrected cluster p value $> .1$). The clusters showing a significant effect of confidence correspond to a set of central and parietal electrodes at a latency of 1200–1250 msec (300–350 msec after the onset of the cue) and a wider temporal window in a group of frontal electrodes for 1500–1600 msec (600–700 msec from the cue).

The main effect of confidence is clearly seen in the time series of the electrical activity of two representative central and frontal electrodes (Figure 2B). Regions of significance are labeled with gray shadows, with higher ERP amplitude for the high-confidence condition as compared with the low-confidence condition. The principal effect observed here is a modulation of the ERP by the subjective response at the time of the memory retrieval. There is no obvious correlate of confidence or performance at the time of sensory encoding or during the brief buffering of information in iconic memory. In addition, the visual inspection of the time series of several electrodes showed a clear interaction term between correct versus incorrect trials at high confidence levels (see Figure 2B). Motivated by this finding, we performed post hoc comparisons through pairwise Wilcoxon tests for correct versus incorrect trials (only at significant times of the main effect of confidence). We found several clusters defined by sensors with a $p < .05$ (uncorrected) that survive the correction for multiple comparisons (corrected cluster p value $< .0002$) (Figure 2C). The effect is stronger at 600–700 msec after the cue ($n = 19$ electrodes). Trials with hits and false alarms could be reliably differentiated at the ERP level of the retrieval phase.

Next, we investigate the possibility of finding neural markers of the subjective confidence level in a trial-by-trial basis. To this aim, we perform a linear regression

of the EEG data (Figure 2D) to predict on a single-trial basis the reported level of confidence (see Methods for more details). We performed this analysis for each participant only for correct trials to disambiguate the results from the objective response. In concordance with the factor group analysis, we observed a broad time window during the retrieval stage (300–600 msec from cue onset) for which the ERP's amplitude correlates with the subjective response at the single-trial level. The topographies of the forward model computed from the linear regression model showed a similar profile to the ones previously found using the factor group analysis. Similar results were obtained equalizing (subject by subject) the number of trials for which the cue was presented at the right or left visual field. This was done to eliminate any perceptual or attentional effect that could arise from a bias in the spatial position of the cue. Overall, these results show that the ERPs are modulated during the retrieval phase, but not during the encoding phase, by the level of subjective confidence.

Frequency Analysis at the Encoding and Retrieval Stages

To determine the covariation of oscillatory stationary activity with confidence and performance, we performed a frequency analysis on theta, alpha, and beta bands (4- to 8-, 9- to 11-, and 15- to 30-Hz bands, respectively). We analyzed independently the encoding and retrieval phases. When only using information about performance (ignoring confidence), we observed that correct trials, compared with incorrect, had higher alpha power mainly over right central areas during the encoding phase (Figure 3, 28 of 128 electrodes) and lower power in the beta band during the retrieval phase mainly in left-central frontal areas (15 electrodes). Correct trials also had increased power in the theta band at central-frontal electrodes (23 electrodes) during the retrieval period. A similar pattern was observed for the effect of confidence, when performance was not included in the variance. After controlling for the correlation of the main factors, we observed that most of the variance in the alpha band is affected by confidence (seven electrodes), without reaching significance for the beta and theta bands. We observed a small effect of performance in the alpha band during the encoding stage through a planned comparison over right central areas (six electrodes, $p < .05$), but it did not reach significance when controlling for multiple comparisons.

The Topography of Subjective Confidence

High-density ERP recordings can be combined with source models to provide a tentative estimate of the cortical origin of EEG activity. We performed a source modeling analysis of the ERP during the different stages

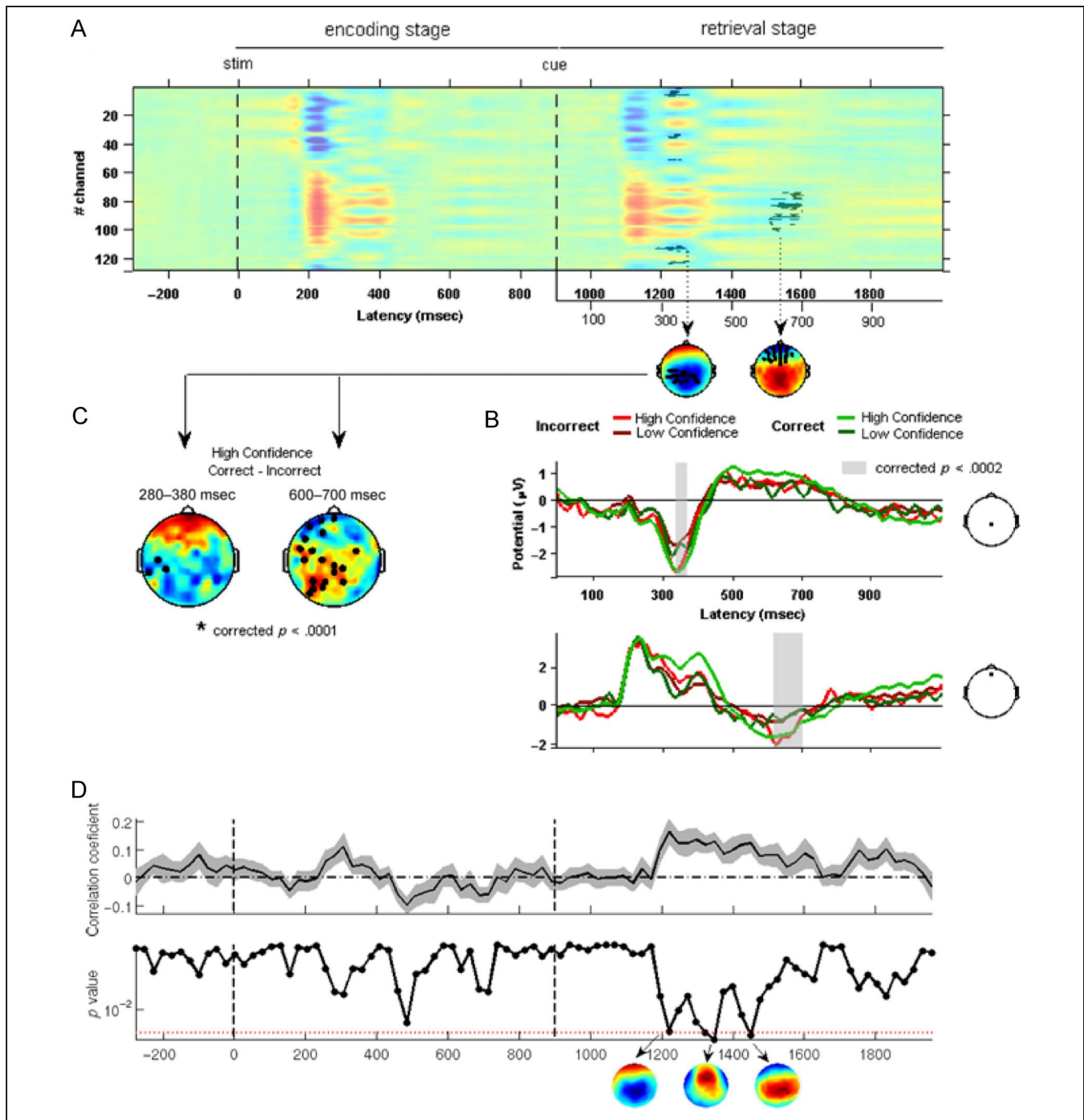
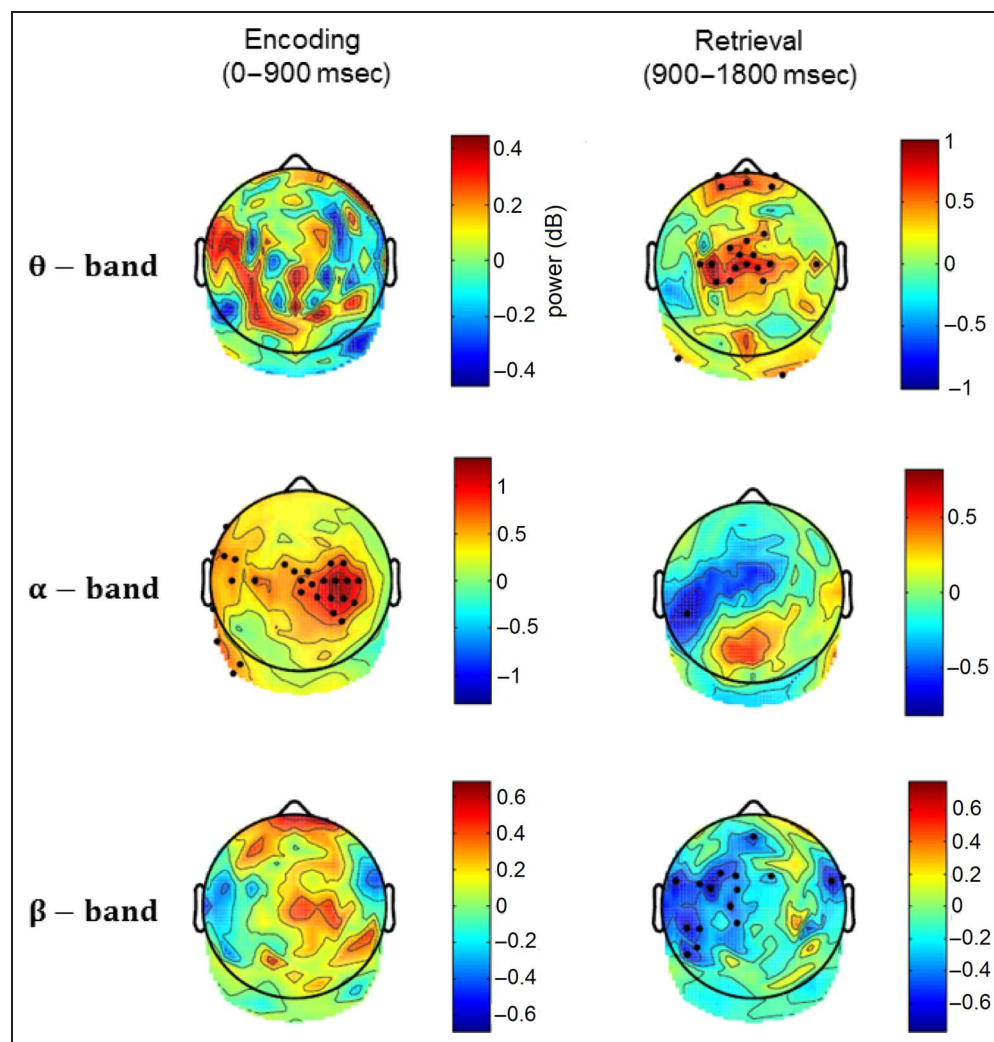


Figure 2. ERP modulations by subjective confidence and objective performance during the encoding and retrieval stages of a partial report paradigm. (A) Raster plot of ERPs averaged across participants and conditions, aligned to the onset of the stimulus (in blue negative potentials, in red positive potential, scale $[-5, +5 \mu V]$) and superimposed to the effect of confidence in the Friedman rank-sum test (black points, cluster-corrected p value $< .0002$). Dashed lines indicate the time of stimulus (0 msec) and cue onset (900 msec after stimulus onset). It is also shown the topographies corresponding to the significant clusters (in black, significant electrodes in the nonparametric analysis). There were no significant differences at any time or electrode for the effect of objective performance (cluster-corrected p value $> .1$). (B) Time series of electric potential for two electrodes across the different conditions of subjective and objective performance. They represent examples of representative electrodes showing a significant effect of confidence. (C) Post hoc comparisons through pairwise Wilcoxon tests for correct versus incorrect trials at significant times of the main effect of confidence. Topographies showed the averaged ERP difference across participants for the correct–incorrect comparison at two different time windows (in blue negative potentials, in red positive potential, scale $[-1, +1 \mu V]$). Asterisk (*) denotes several clusters defined by sensors with a $p < .05$ (uncorrected) that survive the correction for multiple comparisons (corrected cluster p value $< .0002$). (D) Linear decoding analysis of single-trial subjective confidence at correct trials. Upper plot shows the averaged correlation coefficient (SD in shaded gray) between the estimated level and the observed level of subjective confidence for all trials and participants (see Methods for details). Dotted line represents the zero correlation level. Bottom plot shows the p value of the calculated Pearson correlation coefficient for each sample time. Red dotted line indicates the p value threshold corrected with FDR < 0.05 (p value threshold corrected = .002). The topographies correspond to the discriminant components obtained through this analysis, which nicely resemble the topographies of the ERPs elicited at the significant times for the main effect of confidence.

Figure 3. Global correlates of the objective performance at theta/alpha/beta frequency bands. Theta, alpha, and beta power contrasting correct versus incorrect trials. False color indicates power ratio (in dB) over 128 electrodes. Theta/alpha/beta power is computed in the range of 4–8/9–11/15–30 Hz and averaged across all trials and participants. Encoding and retrieval phases were analyzed separately. Black dots indicate a significant difference in power between correct and incorrect trials (FDR < 0.05).



of the task through a model of distributed cortical sources (Baillet et al., 2001) using the ERP data from the entire STIM and CUE periods. Cortical current maps showed greater involvement of occipital areas at the encoding stage than at the retrieval stage and the reverse for frontal and temporal areas (Supplementary Figure 1). EEG inverse modeling should be interpreted cautiously because of its inherent limitations (Michel et al., 2004). Nevertheless, the results are consistent with the general notion that visual processing during the sensory stage should localize in occipital regions, whereas memory retrieval may originate in frontotemporal regions (Pasternak & Greenlee, 2005; Curtis & D'Esposito, 2003; Cohen et al., 1997; Courtney, Ungerleider, Keil, & Haxby, 1997).

To tentatively localize the dynamic involvement of different cortical regions in the construction of subjective confidence, we estimated the electrical sources that generate the different ERP waveforms during the retrieval stage. We first identified the principal response components of the system through local peak analysis (Supplementary Figure 2). As described previously (Sigman &

Dehaene, 2008), response components are data-driven vectors of 128 coordinates corresponding to specific pattern of activity across all electrodes, obtained after averaging the absolute value of the voltages recorded over all electrodes from the grand-averaged ERP (see Methods). We analyzed the cortical current maps for the response components of the retrieval phase. Late response components (C-332ms, C-457ms, C-630ms), which showed central topography and have latencies that coincide with the time windows of the clusters showing a modulation effect of subjective confidence, presented extensive coactivations in OFC (BA 10, middle and superior frontal gyri) and parietal (superior parietal lobe), AC, and temporal (mainly middle temporal gyrus and inferior temporal gyrus) regions (Figure 4A).

From the previous analysis, we identified four ROIs with maximal intensity in these cortical maps and analyzed the temporal dynamics of the current sources, separately for high-confidence versus low-confidence trials (Figure 4B). As a control, we also included an occipital ROI, which was activated during earlier phases but not during late phases of the retrieval period. An analysis of

the amplitude difference of cortical generators between confidence conditions (Figure 4B and C) showed significant differences for AC, orbitofrontal, temporal, and parietal ROIs at latencies between 1180 and 1380 msec

and differences in AC and orbitofrontal activation at latencies between 1500 and 1600 msec, coinciding with the occurrence of the C-332ms and C-630ms components in the voltage scalp maps. In all cases, high-confidence

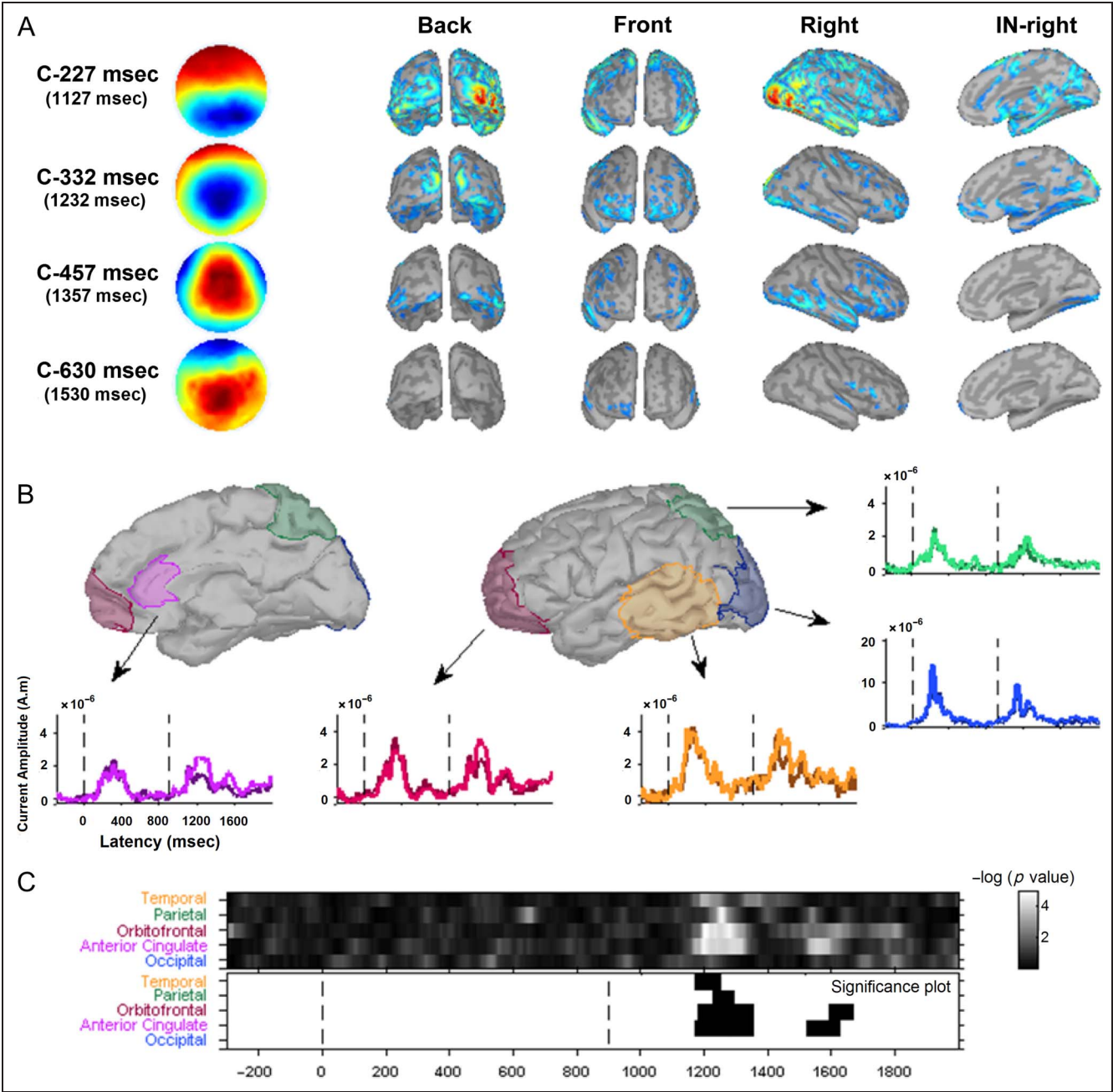


Figure 4. The sources of subjective confidence. (A) Source estimation of each response component of the retrieval stage. Cortical current maps accounting for the observed topography represented on a smoothed version of the standardized cortex. Activations are expressed in terms of dipole current amplitude normalized for each density map (max: red, min: blue). There is an extensive network of coactivations in OFC, parietal, AC, and temporal regions in components modulated by subjective confidence. (B) Neuroanatomical regions selected for time series analysis of current density. We selected four bilateral regions encompassing the network of activations previously found: orbitofrontal, AC, parietal, and temporal regions. The plots show the time series of the current density in these ROIs, classified by subjective condition (low confidence, dark trace; high confidence, light trace). There are several time windows where the current density is higher in the high-confidence condition than in the low-confidence condition. (C) Statistical significance of the time series analysis. Top: Raster plot of p values ($-\log p$) for each ROI, assessed by multiple paired t test. Bottom: Significance plot corrected by multiple comparisons through a cluster randomization approach. Black lines represent significant samples in each ROI (cluster-corrected p value $< .0002$). Results showed a differential activation pattern for AC, orbitofrontal, temporal, and parietal regions at latencies between 1180 and 1380 msec and differences in AC and orbitofrontal activation at latencies between 1500 and 1600 msec, corresponding with the occurrence of C-332ms and C-630ms response components in the voltage scalp maps.

trials show higher current density estimates than low-confidence trials.

DISCUSSION

Our main aim here was to use EEG to reveal the dynamics of brain processes indexing objective performance and subjective judgments of confidence. We implemented a version of the partial report paradigm (Sperling, 1960), which separates the time of encoding and retrieval of information (Zylberberg et al., 2009). Our results, obtained through several complementary analyses, identified more prominent effects of confidence during the retrieval stage. This is in part expected by the design of the task, as only when the cue is presented, participants know exactly the letter to which they have to respond. However, during the encoding period, participants may also have an estimate of the general knowledge that they have (i.e., the precision of the iconic memory of the array), which may be expressed in confidence and performance. Indeed, we observe that the power of the alpha band in the ERPs is indicative of the confidence expressed retrospectively. Overall, these results suggest that confidence is partly determined by a broad state (indexed by the power in oscillatory activity in the EEG) during the encoding period, which may reflect participant's perception of the quality and precision of the iconic memory of the array. Then, during the retrieval period, when the full specification of the task is determined by providing the specific cue, we observe that confidence is indexed by relatively late potentials at latencies of approximately 400 and 600 msec. This indicates that confidences are biased by specific brain states during early sensory acquisition but that, subsequently, it is not constructed online as evidence is accumulated toward a decision. Instead, confidence attributions are more consistent with a retrospective mechanism that monitors the entire decision process.

In our study, we measured in each trial confidence and performance. These two measures are of course correlated but with sufficient dispersion to perform a decoupled analysis on both factors. If we only take into account the objective response (without considering our measure of confidence), we clearly observe main effects of performance in brain activity. These results would have been interpreted as a pure effect of performance in our study if we had not also collected estimates of confidence. After controlling for the covariation between factors, our results indicate that most of the variance in the EEG signal is accounted by explicit reports of confidence. Additional variability in performance, which is not accessible to the confidence report system, shows a weak overall effect on the EEG signal.

Previous work on classical visual working memory (VWM) or visual STM (VSTM) has found several neural correlates of VWM maintenance (Roux & Uhlhaas, 2014; Luck & Vogel, 2013; Klimesch, Freunberger, Sauseng, &

Gruber, 2008). Imaging studies (EEG, fMRI) in humans have found that a neural activity contralateral to the cue hemisphere exhibits a sustained activity during the encoding/retention period (Todd & Marois, 2004, 2005; Vogel, McCollough, & Machizawa, 2005; Vogel & Machizawa, 2004). Recently, studies on a weak intermediate form of VSTM showed that this high-capacity store is represented in visual parts of the brain (i.e., V4; Sligte, Scholte, & Lamme, 2009). The fact that we did not observe global precursors at the ERP level of accuracy (and confidence) during the encoding period is in agreement with these findings, given the spatial specificity of the effect. Hence, we cannot conclude that fluctuations in brain activity during encoding do not affect subjective confidence or objective performance.

Oscillatory processes are also important in VSTM and VWM: Alpha and theta band amplitudes are enhanced during STM and working memory (WM) retention period and are suppressed thereafter (Busch & Herrmann, 2003; Jensen, Gelfand, Kounios, & Lisman, 2002; Jensen & Tesche, 2002; Raghavachari et al., 2001). In turn, WM load was predicted by a variety of markers, among contralateral delay activity, increased theta and alpha band frequencies, and theta–gamma phase coupling (Roux & Uhlhaas, 2014; Lisman & Jensen, 2013; Luck & Vogel, 2013; Sauseng, Griesmayr, Freunberger, & Klimesch, 2010). In our actual experiment, we can assume that WM load is mainly constant across the entire experiment (there are no changes in set size, and participants might be holding in WM as much elements as they could), and these results are in line with our finding that theta band power is not increased comparing correct versus incorrect responses during the encoding period. Conversely, we found a robust alpha-band amplitude increase during the encoding period for correct responses and a decrease during the retrieval period, reflecting the fact that this paradigm relies on strong attentional processes (selection of information from iconic memory, active maintenance in WM, re-allocation of attention by the retrocue; Klimesch, 2012; Zylberberg et al., 2009; Graziano & Sigman, 2008; Lepsien & Nobre, 2007). A similar pattern of activation/deactivation was found in memory encoding and retrieval by other authors (Sauseng et al., 2005; Busch & Herrmann, 2003; Jensen et al., 2002). Thus, the increased alpha-band power can be interpreted as reflecting the suppression of irrelevant information (or blocking the arrival of new information that overwrites the weak VSTM; Klimesch, 2012; Sauseng et al., 2009, 2010; Klimesch, 1999), producing a positive effect on the perceived level of confidence by the participant. Besides, increased alpha-band power resembles the well-known effects of meditation on alpha power (Travis & Shear, 2010; Lagopoulos et al., 2009; Cahn & Polich, 2006; Takahashi et al., 2005), suggesting that, for a confident response, participants have to avoid focusing attention on a specific item—like what actually happens in open mind meditation—providing a more efficient allocation of resources (Raffone & Srinivasan, 2010; Travis & Shear, 2010; Lutz et al., 2009; Slagter et al.,

2007; Tang et al., 2007; Cahn & Polich, 2006; Valentine & Sweet, 1999).

In our study, we also find a modulation of beta power over frontal and central electrodes predominantly on the left hemisphere. Modulation of beta power has been observed in many motor tasks, in particular, those involving inhibition of movement (reviewed in Aron, Robbins, & Poldrack, 2014). Increased beta activity particularly in the right inferior frontal cortex has been observed when inhibiting motor response in a decision task (Wessel, Conner, Aron, & Tandon, 2013; Swann et al., 2011). It is interesting to note that, in our study, participants are asked to inhibit their verbal response during the retrieval period, which would be consistent with the left lateralization of speech function and the increased beta activity observed of correct responses. Perhaps, equally relevant is the observation that beta power is modulated by attentional manipulations in sensory cortex in a similar fashion as alpha power (with higher power corresponding to effort-full attention; Jones et al., 2010). Increased beta power is also observed bilaterally in inferior frontal gyrus in top-down attentional control (Green & McDonald, 2008). Our results are in line with these findings and suggest that modulation of successful responses by attention is reflected in a decreased beta power at the retrieval stage.

Our work can be seen as continuing the path set by Hulme, Friston, and Zeki (2009) to experimentally factorize processing, decision, and motor response in the construction of conscious experience. Measuring fMRI responses in a partial report paradigm, they showed a decoupling between stimulus properties indicated by early visual cortex activation and a subsequent parieto/temporal cluster of activation, which is locked to conscious experience of decision-making (Hulme et al., 2009). Our results are in line with these findings, extending to the domain of confidence and revealing a major weight of late components during retrieval in the construction of subjective confidence. Another fMRI study using a postcued report procedure also shows the influence of top-down modulations over conscious visual perception and early sensory processing (Sergent, Ruff, Barbot, Driver, & Rees, 2011).

We identified that, within the retrieval stage, several late response components are modulated by subjective confidence (Figure 4). These response components spread over a space of central topography (N400-like) and are reminiscent of classical N400 and P3 components associated to information processing in different contexts (Kutas & Federmeier, 2011; Polich, 2007; Sergent, Baillet, & Dehaene, 2005). We showed that the N400 component, generally thought to reflect the integration of information (semantic or more general; Kutas & Federmeier, 2011), is modulated by subjective confidence. The P3 (subdivided in P3a and P3b) has been found to be involved in a multitude of tasks, mostly related to the awareness of an event that leads to the activation of

these components (Polich, 2007). The topography of these components projects to a frontotemporal/parietal network that involves the differential activation of ACC, including frontal and prefrontal areas but also parietal and temporal areas related to higher visual processing and attention (Lau, Phillips, & Poeppel, 2008; Polich, 2007; Soltani & Knight, 2000). This is reminiscent of many studies showing a distributed network and long distance connections mediating conscious access (Del Cul, Baillet, & Dehaene, 2007; Sergent et al., 2005) and is in agreement with views of metacognition (including confidence) as a component of the cognitive control of the brain (Barttfeld et al., 2013; McCurdy et al., 2013; Fleming & Dolan, 2012; Yeung & Summerfield, 2012; Persaud et al., 2011). However, as mentioned above, the main contribution of this study is on the timing of brain processes mediating subjective confidence. Other experimental designs (based on focused decisions to attributes better localized in the brain) with higher spatial resolution instruments should be specifically conceived to understand the brain networks involved in this process. In related work developed independently Gherman and Philastides (2014) show that activity related to confidence has a similar spatial distribution and temporal latency as the activity reported here.

Finally, in relation to the dynamics of processes indicative of subjective confidence, our results have three relevant implications for theories of decision-making and confidence.

First, temporal integration in decision tasks is inhomogeneous such that earlier moments of exposure to a stimulus have a stronger influence on choice (Ludwig, Gilchrist, McSorley, & Baddeley, 2005) and confidence (Zylberberg et al., 2012). The fact that confidence is captured late in this study suggests that earlier moments of stimulus presentation (the ones that contribute to choice) are accessed late (when confidence is being evaluated) and shows a temporal decoupling between the moments of relevant stimulus information and the time in which they are accessed by introspection, in line with two-route models of consciousness (Lau & Rosenthal, 2011; Changeux & Dehaene, 2008).

Second, several models suggest that, when confidence report is required, sensory evidence continues to accumulate after the commitment to a choice (Pleskac & Busemeyer, 2010). The late committing to a judgment in confidence in the decision process permits one to use stimuli even after a choice, which is consistent with our results, although this may depend on the specific task (Zylberberg et al., 2012).

Third, a very influential theoretical view is that confidence is a function of decision time (Audley, 1960). For decisions to a fixed threshold, the time passed is a measure of the integrated variance of an accumulator (Kiani & Shadlen, 2009; Audley, 1960), and decisions based on longer RTs ought to be more unreliable. We previously suggested that the estimation of time and,

therefore, of confidence may rely on an integration of activity of neurons, which in turn integrate sensory evidence (Zylberberg et al., 2012). This process would literally instantiate the notion that confidence judgments result from a decision about a decision, that is, a hierarchical cascade of canonical circuits implementing decisions with different levels of abstraction relative to the external world (McClelland, 1979). Our finding of a committal late construction of confidence during the retrieval process is hence coherent with this view, indicating that confidence is not a direct readout of the sensory signal but, instead, of the decision process itself.

Acknowledgments

The authors thank one anonymous reviewer for helpful insight on the interpretation of the results. We also thank Stephanie R. Jones for her insight with regard to beta band results. The research was supported by grants of the University of Buenos Aires, SECyT, and the Human Frontier Science Program.

Reprint requests should be sent to Martín Graziano, Laboratorio de Neurociencia Integrativa, Departamento de Física, FCEyN-UBA and IFIBA, CONICET, Buenos Aires, Argentina, 1428, or via e-mail: martinymartin@gmail.com.

REFERENCES

- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8, 170–177.
- Audley, R. J. (1960). A stochastic model for individual choice behavior. *Psychological Review*, 67, 1.
- Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18, 14–30.
- Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual clutter causes high-magnitude errors. *PLoS Biology*, 4, e56.
- Barttfeld, P., Wicker, B., McAleer, P., Belin, P., Cojan, Y., Graziano, M., et al. (2013). Distinct patterns of functional brain connectivity correlate with objective performance and subjective beliefs. *Proceedings of the National Academy of Sciences, U.S.A.*, 110, 11577–11582.
- Busch, N. A., & Herrmann, C. S. (2003). Object-load and feature-load modulate EEG in a short-term memory task. *NeuroReport*, 14, 1721–1724.
- Cahn, B. R., & Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychological Bulletin*, 132, 180.
- Changeux, J. P., & Dehaene, S. (2008). The neuronal workspace model: Conscious processing and learning. *Learning Theory and Behavior*, 1, 729–758.
- Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: Neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage*, 29, 1150–1160.
- Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, 21, 1751–1765.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 109.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386, 604–608.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, 386, 608–611.
- Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7, 415–423.
- Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences, U.S.A.*, 100, 8520–8525.
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, 5, e260.
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, 132, 2531–2540.
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*, 367, 1338–1349.
- Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *The Journal of Neuroscience*, 32, 6117–6125.
- Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, 329, 1541–1543.
- Gherman, S., & Philastides, M. G. (2014). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage*. doi: 10.1016/j.neuroimage.2014.11.036.
- Graziano, M., & Sigman, M. (2008). The dynamics of sensory buffers: Geometric, spatial, and experience-dependent shaping of iconic memory. *Journal of Vision*, 8, 9.
- Graziano, M., & Sigman, M. (2009). The spatial and temporal construction of confidence in the visual scene. *PLoS One*, 4, e4909.
- Green, J. J., & McDonald, J. J. (2008). Electrical neuroimaging reveals timing of attentional control activity in human brain. *PLoS Biology*, 6, e81.
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87, 96–110.
- Hulme, O. J., Friston, K. F., & Zeki, S. (2009). Neural correlates of stimulus reportability. *Journal of Cognitive Neuroscience*, 21, 1602–1610.
- Insabato, A., Pannunzi, M., Rolls, E. T., & Deco, G. (2010). Confidence-related decision making. *Journal of Neurophysiology*, 104, 539–547.
- Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex*, 12, 877–882.
- Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, 15, 1395–1399.
- Jones, S. R., Kerr, C. E., Wan, Q., Pritchett, D. L., Hämäläinen, M., & Moore, C. I. (2010). Cued spatial attention drives functionally relevant modulation of the mu rhythm in primary somatosensory cortex. *The Journal of Neuroscience*, 30, 13760–13765.
- Kennedy, M. R., & Yorkston, K. M. (2000). Accuracy of metamemory after traumatic brain injury: Predictions during

- verbal learning. *Journal of Speech, Language, and Hearing Research*, 43, 1072.
- Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455, 227–231.
- Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324, 759–764.
- Kim, H., & Cabeza, R. (2007). Trusting our memories: Dissociating the neural correlates of confidence in veridical versus illusory memories. *The Journal of Neuroscience*, 27, 12190–12197.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29, 169–195.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16, 606–617.
- Klimesch, W., Freunberger, R., Sauseng, P., & Gruber, W. (2008). A short review of slow phase synchronization and memory: Evidence for control processes in different memory systems? *Brain Research*, 1235, 31–44.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Laganaro, M., Morand, S., & Schnider, A. (2009). Time course of evoked-potential changes in different forms of anomia in aphasia. *Journal of Cognitive Neuroscience*, 21, 1499–1510.
- Lagopoulos, J., Xu, J., Rasmussen, I., Vik, A., Malhi, G. S., Eliassen, C. F., et al. (2009). Increased theta and alpha EEG activity during nondirective meditation. *The Journal of Alternative and Complementary Medicine*, 15, 1187–1192.
- Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43, 149–164.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De) constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15, 365–373.
- Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences, U.S.A.*, 103, 18763–18768.
- Lepsien, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cerebral Cortex*, 17, 2072–2083.
- Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, 77, 1002–1016.
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17, 391–400.
- Ludwig, C. J., Gilchrist, I. D., McSorley, E., & Baddeley, R. J. (2005). The temporal impulse response underlying saccadic decisions. *The Journal of Neuroscience*, 25, 9907–9912.
- Lutz, A., Slagter, H. A., Rawlings, N. B., Francis, A. D., Greischar, L. L., & Davidson, R. J. (2009). Mental training enhances attentional stability: Neural and behavioral evidence. *The Journal of Neuroscience*, 29, 13418–13427.
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21, 422–430.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287.
- McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *The Journal of Neuroscience*, 33, 1897–1906.
- Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition*, 79, 115–134.
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, 115, 2195–2222.
- Middlebrooks, P. G., & Sommer, M. A. (2012). Neuronal correlates of metacognition in primate frontal cortex. *Neuron*, 75, 517–530.
- Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46, 245–259.
- Murray, M. M., Camen, C., Andino, S. L. G., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *The Journal of Neuroscience*, 26, 1293–1302.
- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15, 1–25.
- Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, 28, 326–341.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44, 369–378.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6, 97–107.
- Persaud, N., Davidson, M., Maniscalco, B., Mobbs, D., Passingham, R. E., Cowey, A., et al. (2011). Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *Neuroimage*, 58, 605–611.
- Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review*, 10, 177–183.
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117, 864.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148.
- Raffone, A., & Srinivasan, N. (2010). The exploration of meditation in the neuroscience of attention and consciousness. *Cognitive Processing*, 11, 1–7.
- Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., et al. (2001). Gating of human theta oscillations by a working memory task. *The Journal of Neuroscience*, 21, 3175–3183.
- Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, 1, 165–175.
- Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, 18, 16–25.
- Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience & Biobehavioral Reviews*, 34, 1015–1022.

- Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., et al. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Current Biology*, *19*, 1846–1852.
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, *22*, 2917–2926.
- Schmid, D. A., Wichniak, A., Uhr, M., Ising, M., Brunner, H., Held, K., et al. (2006). Changes of sleep architecture, spectral composition of sleep EEG, the nocturnal secretion of cortisol, ACTH, GH, prolactin, melatonin, ghrelin, and leptin, and the DEX-CRH test in depressed patients during treatment with mirtazapine. *Neuropsychopharmacology*, *31*, 832–844.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*, 1391–1400.
- Sergent, C., Ruff, C. C., Barbot, A., Driver, J., & Rees, G. (2011). Top-down modulation of human early visual cortex after stimulus offset supports successful postcued report. *Journal of Cognitive Neuroscience*, *23*, 1921–1934.
- Sigman, M., & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *The Journal of Neuroscience*, *28*, 7585–7598.
- Slagter, H. A., Lutz, A., Greischar, L. L., Francis, A. D., Nieuwenhuis, S., Davis, J. M., et al. (2007). Mental training affects distribution of limited brain resources. *PLoS Biology*, *5*, e138.
- Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2009). V4 activity predicts the strength of visual short-term memory representations. *The Journal of Neuroscience*, *29*, 7432–7438.
- Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology*, *14*, 199–224.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*, 1.
- Swann, N., Poizner, H., Houser, M., Gould, S., Greenhouse, I., Cai, W., et al. (2011). Deep brain stimulation of the subthalamic nucleus alters the cortical profile of response inhibition in the beta frequency band: A scalp EEG study in Parkinson's disease. *The Journal of Neuroscience*, *31*, 5721–5729.
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*, 8.
- Takahashi, T., Murata, T., Hamada, T., Otori, M., Kosaka, H., Kikuchi, M., et al. (2005). Changes in EEG and autonomic nervous activity during meditation and their association with personality traits. *International Journal of Psychophysiology*, *55*, 199–207.
- Tang, Y. Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., et al. (2007). Short-term meditation training improves attention and self-regulation. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 17152–17156.
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754.
- Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 144–155.
- Travis, F., & Shear, J. (2010). Focused attention, open monitoring and automatic self-transcending: Categories to organize meditations from Vedic, Buddhist and Chinese traditions. *Consciousness and Cognition*, *19*, 1110–1118.
- Valentine, E. R., & Sweet, P. L. (1999). Meditation and attention: A comparison of the effects of concentrative and mindfulness meditation on sustained attention. *Mental Health, Religion & Culture*, *2*, 59–70.
- Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
- Vilkki, J., Surma-aho, O., & Servo, A. (1999). Inaccurate prediction of retrieval in a face matrix learning task after right frontal lobe lesions. *Neuropsychology*, *13*, 298–305.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.
- Wessel, J. R., Conner, C. R., Aron, A. R., & Tandon, N. (2013). Chronometric electrical stimulation of right inferior frontal cortex increases motor braking. *The Journal of Neuroscience*, *33*, 19611–19619.
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1310–1321.
- Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., et al. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neuroscience Research*, *68*, 199–206.
- Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79.
- Zylberberg, A., Dehaene, S., Mindlin, G. B., & Sigman, M. (2009). Neurophysiological bases of exponential sensory decay and top-down memory retrieval: A model. *Frontiers in Computational Neuroscience*, *3*, 4.