Distinct brain mechanisms for conscious versus subliminal error detection

Lucie Charlesa, Philip van Opstala, Sébastien Martia, Stanislas Dehaene

Published in final edited form as:
Neuroimage. 2013 June ; 73: 80–94. doi:10.1016/j.neuroimage.2013.01.054.

Distinct brain mechanisms for conscious versus subliminal error detection

Lucie Charles[a,b,c,*], **Filip Van Opstal**[a,b,c,d], **Sébastien Marti**[a,b,c], and **Stanislas Dehaene**[a,b,c,e]

[a]INSERM, U992, Cognitive Neuroimaging Unit, CEA/SAC/DSV/DRM/NeuroSpin, Bât 145, Point Courrier 156 F-91191 Gif/Yvette, France

[b]CEA, DSV/I2BM, NeuroSpin Center, Bât 145, Point Courrier 156 F-91191 Gif/Yvette, France

[c]Univ Paris-Sud, Cognitive Neuroimaging Unit, Bât. 300-91405 Orsay cedex

[d]Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium

[e]Collège de France, 11, place Marcelin Berthelot, 75231 Paris Cedex 05, France

Abstract

Metacognition, the ability to monitor one's own cognitive processes, is frequently assumed to be univocally associated with conscious processing. However, some monitoring processes, such as those associated with the evaluation of one's own performance, may conceivably be sufficiently automatized to be deployed non-consciously. Here, we used simultaneous electro- and magneto-encephalography (EEG/MEG) to investigate how error detection is modulated by perceptual awareness of a masked target digit. The Error-Related Negativity (ERN), an EEG component occurring ~ 100 ms after an erroneous response, was exclusively observed on conscious trials: regardless of masking strength, the amplitude of the ERN showed a step-like increase when the stimulus became visible. Nevertheless, even in the absence of an ERN, participants still managed to detect their errors at above-chance levels under subliminal conditions. Error detection on conscious trials originated from the posterior cingulate cortex, while a small response to non-conscious errors was seen in dorsal anterior cingulate. We propose the existence of two distinct brain mechanisms for metacognitive judgements: a conscious all-or-none process of single-trial response evaluation, and a non-conscious statistical assessment of confidence.

Keywords

Error-related negativity; Consciousness; MEG; EEG

Introduction

What are the limits of non-conscious processing? In the past twenty years, evidence has accrued in favor of deep processing of subliminal stimuli (i.e., stimuli presented below the threshold of subjective visibility). Not only can early visual processing be preserved under

*Corresponding author at: INSERM-CEA Cognitive Neuroimaging unit CEA/SAC/DSV/DRM/NeuroSpin Bât 145, Point Courrier 156 F-91191 Gif/Yvette, France. Fax: +33 1 69 08 79 73. lucie.charles.ens@googlemail.com (L. Charles).

Europe PMC Funders Author Manuscripts

masking conditions (Del Cul et al., 2007; Melloni et al., 2007), but subliminal primes can modulate visual (Dehaene et al., 2001), semantic (Van den Bussche et al., 2009) and motor stages (Dehaene et al., 1998; for a review, see Kouider and Dehaene, 2007). Even executive processes, once considered the hallmark of the conscious mind, can be partially influenced by non-conscious signals related to motivation (Pessiglione et al., 2007), task switching (Lau and Passingham, 2007) and inhibitory processes (Van Gaal et al., 2008). These findings raise the issue of whether subliminal stimuli could affect *any* cognitive process, or whether certain processes depend on an all-or-none conscious ignition (Del Cul et al., 2007).

Here, we investigate meta-cognition — the ability to reflect on oneself and on one's own cognitive processes. Intuitively, introspective reflection is virtually indistinguishable from conscious processing: it is hard to envisage introspection without consciousness. This intuition has served as a basis for the frequent identification of consciousness with self-oriented, metacognitive or “second-order” cognition: any information that can enter into a higher-order thought process would be conscious by definition (Kunimoto et al., 2001; Lau and Rosenthal, 2011; Persaud et al., 2007). However, this conclusion may also be disputed. Some metacognitive monitoring processes, such as those associated with the evaluation of one's performance (Logan and Crump, 2010) or the subsequent correction of one's errors (Endrass et al., 2007; Nieuwenhuis et al., 2001; Wessel et al., 2011) are conceivably sufficiently simple and automatized to be deployed non-consciously. Thus, whether metacognitive processing implies conscious processing can and should be tested empirically.

To investigate how performance monitoring relates to conscious perception, the present experiments concentrate on the error-related negativity (ERN), a key marker of error processing. The ERN is an event-related potential that peaks on fronto-central electrodes 50 to 100 ms after making an erroneous response; it is easily observed in EEG recordings (Dehaene et al., 1994; Falkenstein et al., 2000; Gehring et al., 1993), and a similar, though harder to detect MEG component has been reported (Keil et al., 2010; Miltner et al., 2003). The ERN is assumed to originate in the cingulate cortex (Agam et al., 2011; Debener et al., 2005) and its role in cognitive control has been related to error detection (Gehring and Fencsik, 2001; Nieuwenhuis et al., 2001), reinforcement learning (Holroyd and Coles, 2002) and conflict processing (Botvinick et al., 2001; Veen and Carter, 2002).

The debated issue that we address here is whether the ERN indexes a process which is automatic enough to be deployed unconsciously. In relating this issue to the existing literature, it is crucial to keep in mind that an error can fail to be consciously detected for several reasons. A distinction must be made between errors that remain unnoticed (1) because the erroneous action itself is not detected (for instance because it consists in a fast key press or eye-movement (Endrass et al., 2007; Nieuwenhuis et al., 2007; Logan and Crump, 2010; Hughes and Yeung, 2011)), (2) because the subject cannot determine which response is the correct one (e.g. when responding to a visible but confusing stimulus or instruction), or (3) because the subject is completely unaware of the stimulus and therefore of the correct response (e.g. when responding to a stimulus made invisible by masking).

Initially, the relationship between consciousness and the ERN was explored in the context of case (1), i.e. unaware actions (Nieuwenhuis et al., 2001). It suggested that the ERN may

remain present even when participants are unaware of having made a partially erroneous eye-movement (Endrass et al., 2007; Nieuwenhuis et al., 2001; but see Wessel et al., 2011). In these studies, crucially, subjects performed a difficult antisaccade task and were sometimes unaware of their erroneous glances in the pro-saccade direction. These results were further extended to case (2) (i.e., confusion about which response is the correct one), in paradigms where undetected errors were induced by conflicting stimuli evoking two contradictory responses (Dhar et al., 2011; Hughes and Yeung, 2011; O'Connell et al., 2007 but see Maier et al., 2008; Steinhauser and Yeung, 2010). These studies have typically used the Eriksen flanker task, in which the presence of multiple conflicting letters may purposely confuse the participant as to the nature of the correct response.

Here, however, we aimed at testing the third case, i.e. whether an ERN can be elicited by an unseen masked stimulus. Our main motivation was to extend the existing literature on the depth of subliminal processing of masked words and digits (Kouider and Dehaene, 2007). In masking experiments, it is well known that participants may deny seeing the stimuli, yet still perform above chance level in a broad range of categorization task, such as deciding whether a digit is larger or smaller than 5 (Dehaene et al., 1998; Del Cul et al., 2007). As an extreme case, in blindsight, a patient may deny any conscious experience, while remaining able to perform way above chance in simple tasks on stimuli presented in their blind hemi-field (Kentridge and Heywood, 1999; Weiskrantz, 1996).

The specific question for the present research is whether, in subliminal conditions induced by masking, the error detection system may also be triggered non-consciously. We evaluate this question both by monitoring the presence of the ERN, as well as by asking the participants for a second-order behavioral response. On each trial, the participant first makes a forced-choice number comparison, and is then asked to decide whether he made an error or not. The finding of either an unconscious ERN, or of an above-chance second-order metacognitive performance on subliminal trials, would expand the range of unconscious operations. Corroborating recent evidence that even executive processes of task switching and response inhibition may be partially initiated non-consciously (Lau and Passingham, 2007; Van Gaal et al., 2008), it would indicate that an unseen masked stimulus is capable of progressing through a hierarchy of successive processing stages, all the way up to a level of metacognitive monitoring. A negative answer, on the other hand, would support the view that there are sharp limits to unconscious processing, and that some cognitive operations only proceed once the stimulus has crossed an all-or-none threshold for conscious access (Aly and Yonelinas, 2012; Dehaene and Changeux, 2011; Province and Rouder, 2012; Sergent and Dehaene, 2004a).

Only two studies (Pavone et al., 2009; Woodman, 2010) investigated the existence of an ERN on subliminal trials, yet they obtained contradictory results: Woodman (2010) found that the ERN was absent for masked stimuli, while Pavone et al. (2009) found that it could still be detected. Crucially, in order to contrast conscious versus non-conscious processing, both studies manipulated parameters of contrast or duration. Such sensory manipulations *per se* can have a large impact on the amount of information available on subliminal trials compared to conscious trials. Their findings may therefore result in a large part from this objective change in stimulus strength. One of our aims was therefore to determine if changes

in subjective perception alone, in the presence of a constant stimulus, would modulate the ERN and metacognitive performance. To this end, we measured error responses to visual stimuli of variable masking strength, ranging from fully visible to fully invisible (Fig. 1). Such design allowed us to determine how subjective perception of a stimulus, by itself, affects performance-monitoring processes, as assessed by behavioral and error-related MEEG brain measures.

In two masking experiments, participants performed a number comparison task on a masked digit, while perceptual evidence was systematically manipulated by varying the target-mask Stimulus Onset Asynchrony (SOA; Del Cul et al., 2007). To maximize the number of errors, a strong pressure to respond fast was imposed in experiment 1. The main results were replicated in a second experiment in which this pressure was reduced. Crucially, subjective perception was assessed on a trial by trial basis by asking participants to report their visibility of the target (*Seen/Unseen*) as well as their perceived performance (*Error/Correct*) in the number comparison task. Given that subjective reports vary spontaneously across trials, this approach allowed us to study how the ERN and error-detection performance were modulated by subjective perception of the stimulus (subliminal/subjectively *unseen* trials versus conscious/*seen* trials), independently of the objective variation in masking strength.

Materials & methods

Participants

In the first experiment, seventeen volunteers were tested (5 women and 12 men; mean age 23.8 years). Because our experimental conditions were partially determined by subjective reports, four participants were discarded for having insufficient numbers of trials in some of the conditions. Specifically, we removed participants with false-alarm rate superior to 10% in the mask-only condition, or with less than 15% of *seen* trials in the 50 ms SOA condition. In the second experiment, sixteen participants were tested (6 women and 10 men; mean age 23.2 years). Two had to be discarded due to technical problems during MEG recording. One participant was discarded using the same behavioral criteria as in the first experiment. In the end, each experiment comprised data from 13 participants. All participants had normal or corrected-to-normal vision.

Design & procedure

A masking paradigm similar to Del Cul et al. (2007) was used in this experiment. The target-stimuli (the digits 1, 4, 6, or 9) were presented on a white background screen using E-Prime software. The trial started with a small increase in the size of the fixation cross (100 ms duration) signalling the beginning of the trial. Then the target stimulus appeared for 16 ms at one of two positions (top or bottom, 2.29° from fixation), with a 50% probability. After a variable delay, a mask appeared at the target location for 250 ms. The mask was composed of four letters (two E's and two M's, see Fig. 1) tightly surrounding the target stimulus without superimposing or touching it. The stimulus-onset asynchrony (SOA) between the onset of the target and the onset of the mask was varied across trials. Five SOAs were randomly intermixed: 16, 33, 50, 66 and 100 ms. The foreperiod duration was manipulated so that the mask always appeared 800 ms after the signal of the beginning of the trial. In one

sixth of the trials, the target number was replaced by a blank screen with the same duration of 16 ms (mask-only condition), allowing us to study visibility ratings when no target was presented.

Participants primarily had to perform a forced-choice task of comparing the target number to the number 5. Responses were collected within 1000 ms (experiment 1) or 2000 ms (experiment 2) after target onset with two buttons using the index of each hand (left button press = smaller-than-5; right button-press = larger-than-5 response). To induce errors, participants were instructed to respond as fast as they could just after the appearance of the target. In experiment 1, time pressure was increased by presenting an unpleasant sound (mean pitch: 136.2 Hz, 215 ms duration) 1000 ms after target presentation whenever response time exceeded 550 ms. In experiment 2, no further time pressure was imposed.

At the end of each trial, after another delay of 500 ms, participants were requested to provide two subjective answers with no time–pressure. The first answer was related to the subjective visibility of the target number. In this visibility task, participants had to indicate if they saw a target number or not. The second answer concerned the participants' knowledge of their performance. Here, they had to indicate whether they thought they had made an error or not in the number comparison task (performance evaluation task). Instructions were clearly stated to ensure that participants understood that the performance evaluation task was directed to the number comparison task and not the visibility judgment. Furthermore, participants were informed that, even when they had not seen the stimulus and thought that they responded randomly, they still had a 50% chance of having made a correct response. Therefore, they were told to hazard a guess on their performance, even when they did not see the stimulus. For both subjective responses, words corresponding to the two responses (*seen*/*unseen* and *error/correct*) were displayed on the screen and participants had to use the corresponding-side buttons to answer. The words were presented at randomized left and right locations (2.3° from fixation) to ensure that participants didn't use automatized button-press strategy.

The experiment was divided in blocks of 96 trials. Each block contained 16 trials for every SOA condition, with each digit presented at the two possible target locations (top/bottom). Participants performed 6 or 7 blocks during EEG/MEG recording. For Experiment 1, in order to achieve fast responses, participants were given a training session before the actual recording. They first received 5 min of training where the target stimulus was not masked. Next, participants performed 3 pre-recording blocks of the actual experiment in order to check that overall performance was suitable for MEG/EEG recording. In Experiment 2, where fast responding was not required, only ten trials of the experiment were given as training before starting the actual recording.

Simultaneous EEG and MEG recordings

Simultaneous recording of MEG and EEG data was performed. The MEG system (the Elekta-Neuromag) comprised 306 sensors: 102 Magnetometers and 204 orthogonal planar gradiometers (pairs of sensors measuring the longitudinal and latitudinal derivatives of the magnetic field). The EEG system consisted of a cap of 60 electrodes with reference on the

nose and ground on the clavicle bone. Six additional electrodes were used to record electrocardiographic (ECG) and electro-oculographic (vertical and horizontal EOG) signals.

A 3-dimensional Fastrak digitizer (Polhemus, USA) was used to digitize the position of three fiducial head landmarks (Nasion and Pre-auricular points) and four coils used as indicators of head position in the MEG helmet, for further alignment with MRI data. Sampling rate was set at 1000 Hz with a hardware band-pass filter from 0.1 to 330 Hz.

SDT analysis

To obtain an unbiased measure of visibility and performance, we used Signal Detection Theory (SDT) to compute d' = z(HIT) − z(FA) for the target-detection task (*detection-d'*, where HIT = proportion of trials with target present and response *seen*, and FA = proportion of trials with target absent and response *seen*) and the number comparison task (where HIT = proportion of trials with target smaller than 5 and a left response, and FA = proportion of trials with target larger than 5 and a left response).

The *meta-d'* measure was computed according to Maniscalco and Lau (2012). Briefly, classic SDT can be extended to predict what should be the theoretical performance in meta-cognitive judgements where one must evaluate one's own primary performance, such as confidence ratings or error detection. The theory assumes that both primary and meta-cognitive judgements have access to the same stimulus sample on the same continuum. First-order judgments are performed by setting a first criterion in the middle of the continuum. Meta-cognitive judgements are performed by setting two additional criteria surrounding the first-order one, and responding “error” if the sample falls between these two criteria, or “correct” if the sample falls beyond them (i.e. a sample distant enough from the first-order criterion signals high confidence in the primary response). From this ideal-observer theory, precise mathematical relations linking performance and meta-performance can be deduced (Galvin et al., 2003) and it is possible to compute a second-order measure of meta-performance by classifying meta-cognitive responses as second-order hits and false alarm. However, the traditional measure of d' does not directly apply to a second-order task because it is not unbiased (second-order d' systematically depends on the first-order criterion) and the assumption of normality of the distributions is violated. In order to obtain a valid measure of meta-performance, unbiased and comparable to the first-order d', Maniscalco et al. (http://www.columbia.edu/~bsm2105/type2sdt/) proposed an alternative solution, *meta-d'*. Their proposal consists in bringing both first and second-order performance to the same scale, by determining what should have been the d' in the first-order task given the observed second-order (meta) performance, under the assumption that the subject used exactly the same information in both cases. Since *meta-d'* is expressed in the same scale as d', the two can be compared directly. When *meta-d'* $< d'$, it means that the subject did worse in the performance evaluation task than expected according to his actual d' value. On the opposite, if the *meta-d'* $> d'$, it means that more information was available for subjective performance evaluation than for the primary objective decision.

Meta-d' was estimated by fitting the parameters of a type-I SDT model so that the predicted type-II hits and false-alarm rates were fitted to the actual type-II data. Therefore, *meta-d'*

corresponds to the d' that maximizes the likelihood of the observed type performance, assuming the same bias of response as the one observed in the data.

MEG/EEG data analysis

MEG data were first processed with MaxFilter™ software using the Signal Space Separation algorithm. Bad MEG channels were detected automatically and manually, and interpolated. Head position information recorded at the beginning of each block was used to realign head position across runs and transform the signal to a standard head position framework.

To remove the remaining noise, Principal Component Analysis (PCA) was used. Artifacts were detected on the electro-occulogram (EOG) and electro-cardiogram. Data were averaged on the onset of each blinks and heart beats separately and PCA was performed separately for each type of sensor. Then, one to three of the first components characterizing the artifact were selected by mean of visual inspection to be further removed.

Data were then entered into Matlab software and processed with Fieldtrip software (http://fieldtrip.fcdonders.nl/). For the first experiment, an automatic rejection of trials based on signal discontinuities (all signal above 30 and 25 standard deviations in 110–140 Hz frequency range) was performed. However, less than 1% of the trials removed, and therefore this step was omitted in experiment 2, where the number of error trials was smaller. A low-pass filter at 30 Hz was then applied as well as a baseline correction from 300 ms to 200 ms before target onset.

Data were then realigned on response onset to be further averaged by subject and conditions. To obtain grand-average evoked response data, we first averaged individual data for each SOA separately, then averaged across SOAs and then across participants. For the first experiment only, response times were equalized across error and correct trials (see Supplementary Methods). Without such a correction, the slower RTs on *seen* correct trials caused artifactual differences due to non-aligned sensory-evoked components on response-locked averages (Fig. S4). This RT correction was not needed in experiment 2 where RTs were longer and response-locked ERPs were therefore uncontaminated by sensory-evoked components. An additional baseline correction was simply performed from 200 to 50 ms before motor response. We verified that these small differences in procedure did not affect the main results, and in particular the same dependency of ERN on visibility was observed when no RT correction was applied to experiment 1 (See Supplementary Results).

Combined EEG/MEG source reconstruction

Brainstorm software was used to derive current estimate from correct and error MEEG waveforms, for each condition of visibility and each subject separately. Cortical surfaces of 22 participants (2 participants were discarded in each experiment as no MRI data could be obtained) were reconstructed from individual MRI with FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) for cortex surface (gray-white matter boundary) and Brainvisa (http://brainvisa.info/) for scalp surface. Inner skull and outer-skull surfaces were estimated by Brainstorm, in order to compute accurate forward model using a three-compartment boundary-element method (OpenMeeg toolbox; http://www-sop.inria.fr/athena/software/OpenMEEG/). Sources were computed with weighted minimum-norm method and dSPM

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

(depth-weighting factor of 0.8, loosing factor of 0.2 for dipole orientation). Individual source estimate data were then projected on a template cortical surface, in order to be averaged across participants, separately for each experiment. Mean power (i.e. square of the t-values) of regions of interest was computed to present time-courses of brain activity.

Statistical analysis

Behavioral data analysis—All behavioral data analyses were performed with Matlab software with the help of the Statistics toolbox using repeated-measures analysis. Reaction-time analysis was performed on the median RT of each condition.

MEG data analysis—To detect significance differences between error and correct conditions for each type of sensor, we used a cluster-based non-parametric t-test with Monte Carlo randomization provided in the Fieldtrip software (Maris and Oostenveld, 2007). This method identifies clusters of nearby sensors presenting a significant difference between two conditions for a sufficient duration while correcting for multiple comparisons. For each sample, t-values and associated p-value were first computed by means of a non-parametric Monte-Carlo randomization test. Clusters were then identified by taking all samples adjacent in space or in time (minimum of 2 sensors per cluster, 4.3 average spatial neighbors per EEG electrode and 8.2 per MEG channel) with $p < 0.05$. The final significance of the cluster was found by computing the sum of t-values of the entire cluster, and comparing with the results of Monte-Carlo permutations (1500 permutation). Clusters were considered significant at corrected $p < 0.05$ if the probability computed with the Monte-Carlo method was inferior to 2.5% (two-tailed test). Time-windows of interest were chosen for each experiment on the basis of the EEG results for *seen* trials to optimize cluster detectability. The ERN is usually observed in a 100 ms time-window after button press (Dehaene et al., 1994). As the onset of the difference was observed slightly later in experiment 1 than experiment 2, search for clusters was performed respectively on a 30–100 ms time-window after motor response for experiment 1 and 0–100 ms in experiment 2.

For statistical analysis on a-priori clusters, average voltage over central electrodes (FC1, FC2, C1, Cz, C2) were computed over the same time-window as for the cluster analysis (30–100 ms and 0–100 ms after motor response respectively for experiment 1 and 2, analysis of later time windows is reported in Supplementary Results). Analysis was performed in Matlab using repeated-measures t-tests (two-tailed) and ANOVA with visibility and performance as within-subjects factors. Analysis by SOA required more sophisticated statistical analysis as trial rejection and factorial analysis (SOA*Visibility*Performance) led to unequal number of participants in each combination of condition. Therefore, analysis of variance was performed in R software using a linear mixed-effects model ((Baayen et al., 2008) R package lme4) which allowed us to include all data available (unbalanced design) and still encompass repeated-measures. The functions used yield t statistic and, as degrees of freedom cannot be computed for this kind of analysis, p-values were derived from a Markov Chain Monte Carlo (MCMC) method.

Results

Subjectivity visibility is reliably affected by masking

Subjective visibility, as measured by the percentage of *seen* responses, increased in a non-linear sigmoid manner with SOA ($F_{5,55} = 316.7$, $p < 10^{-4}$, see Supplementary result), replicating earlier results (Del Cul et al., 2007). Stimuli that were masked after a short latency (SOA < ~ 50 ms) were almost always judged as invisible, while visibility rose very rapidly after this point (Fig. 2). Visibility was slightly higher in experiment 1 compared to experiment 2 (two way ANOVA with factor experiment and SOA, $F_{1,55} = 3.371$, $p = 0.094$), probably because participants underwent more training in experiment 1 than in experiment 2. However, the main effect of SOA was highly significant in both cases, and no interaction was found between SOA and experiment ($F_{5,55} = 1.77$, $p = 0.135$).

Raw visibility reports (*Seen*, *Unseen*) can be criticized as subjective and potentially biased measures. We therefore transformed them into an objective index of target detection sensitivity and bias, using classical signal detection theory. To this end, at each SOA level, visibility ratings (percent *Seen* responses) were compared against those in the mask-only condition, and converted to *detection-d'* and bias values (see Materials & methods). For the shortest SOA condition (SOA = 16 ms), participants were at chance to detect the presence of the target, as the *detection-d'* did not differ significantly from 0 (Exp1: average $d' = 0.15$, $t_{12} = 0.98$, $p = 0.34$, Exp2: average $d' = 0.01$, $t_{12} = 0.07$, $p = 0.94$). Furthermore, participants adopted a conservative criterion (bias > 0, $t_{12} = 14.6$, $p < 10^{-4}$, $t_{12} = 17$, $p < 10^{-4}$), reflecting the frequent use of the *unseen* response on both target-present and mask-only trials, and therefore confirming the invisibility of the targets at this SOA. As SOA increased, *detection-d'* increased ($F_{4,44} = 220.7$, $p < 10^{-4}$) while response-bias toward the *unseen* response decreased ($F_{4,44} = 221$, $p < 10^{-4}$), confirming that visibility improved with SOA. Finally, on mask-only trials, false-positives were very rare (exp 1: 3% erroneous *seen* responses; exp 2: 4%). Overall, these observations confirm that subjective visibility reports were reliable and that masking at short SOA induced a subjective state of invisibility on a large proportion of trials.

Cognitive and metacognitive performance are affected by masking

We then looked at the variations in performance and meta-performance as a function of SOA (see Fig. 2; Response times are reported in Supplementary material).

Objective performance in the number comparison task increased with SOA ($F_{4,44} = 318.89$, $p < 10^{-4}$), with a non-linear profile virtually parallel to subjective visibility (Figs. 2C-D). As intended, in the first experiment where strong time pressure was imposed, participant's performance did not reach ceiling even for the largest SOA (SOA 100 ms, Fig. 2C). Thus, experiment 1 achieved its goal of generating a minimum of ~20% errors at each SOA, allowing us to explore the mechanisms of error detection. In the second experiment, where time pressure was relaxed, performance at the longest SOA reached 95% correct (Fig. 2D), thus resulting in a much smaller number of analyzable errors. This pattern resulted in a significant SOA by experiment interaction ($F_{4,44} = 19.49$, $p < 10^{-4}$).

Next, we investigated meta-cognitive performance as a function of SOA. Our procedure allowed us to compare, on each trial, the subject's objective accuracy with his evaluation of his performance. Trials were classified as "*meta-correct*" if they were error trials perceived as errors, or correct trials perceived as correct. Otherwise they were labelled as "*meta-incorrect*". Meta-cognitive performance (i.e. percentage of meta-correct trials) increased with SOA ($F_{4,44} = 165.83$, $p < 10^{-4}$), reaching 97% meta-correct trials in both experiments. As seen on Figs. 2C–D, both types of meta-incorrect responses (undetected errors as well as correct trials misperceived as errors) progressively vanished with increasing SOA, in tight parallel with increasing target visibility.

Overall, these results indicate that the SOA manipulation successfully modulated, in tight parallel, the performance of our three tasks: objective number comparison, metacognitive evaluation, and visibility judgment. In the next section, we show how visibility, independently of SOA, indexes a major switch in the performance of the other two tasks.

Cognitive and metacognitive performance are affected by visibility

To better characterize how behavior changed on conscious and non-conscious trials, the data were then split by visibility (*Seen* vs *Unseen*). As visibility increased in a non-linear way with SOA, many participants had fewer than 5 trials in one of the visibility condition for extreme SOA values. Therefore, we removed these trials from the analysis and from the figures, keeping for *seen* trials only trials corresponding to SOA larger than 33 ms and for *unseen* trials those corresponding to SOA smaller than 50 ms.

As can be seen in Figs. 3A–B, participants performed way above chance both in the number comparison task and in the performance evaluation task when they could see the target number, independently of the SOA condition (for experiments and all SOA, performance and meta-performance > 50%, $p < 0.005$). When averaging together all SOAs or when considering only intermediate SOAs (33 and 50 ms) for which we had approximately as many *seen* and *unseen* trials, both performance and meta-performance were significantly superior on *seen* compared to *unseen* trials (for both experiments, all $p < 0.01$). This finding was similar in both experiments, with a small difference: for the *seen* trials, at the longest SOA (100 ms), performance was lower in experiment 1 compared to experiment 2 (80% versus 96%), again because of the strong time pressure imposed in experiment 1.

To obtain a clearer view of the relative sensitivity of the subject in the second-order performance evaluation task compared to the primary task, performance was converted to d' and *meta-*d' values (Figs. 3C–D). As described by second-order Signal Detection Theory (Galvin et al., 2003; Maniscalco and Lau, 2012; Rounis et al., 2010) (SDT), d' and *meta-*d' give an unbiased estimate of performance, respectively for first-order task (here, number comparison) and second-order task (error detection). Since these two measures are on the same scale, they allow us to compare what the first-order performance actually was to what it should have been, given second-order error detection accuracy (Galvin et al., 2003; Maniscalco and Lau, 2012; Rounis et al., 2010).

This analysis confirmed that even for equal SOA, both performance and meta-performance showed a sudden jump with visibility (see Figs. 3C–D; statistics in Table 1). Thus, visibility

judgment, although a subjective task, also indexes a large change in objective performance: *seen* and *unseen* trials differ massively in the quantity of usable information for both primary and secondary judgments (Del Cul et al., 2007, 2009).

For *seen* trials (Figs. 3C-D, solid lines), performance and meta-performance (d' and *meta-d′*) increased significantly with SOA in both experiments (see Table 2). *Meta-d′* always significantly exceeded d', in particular in Experiment 1 with time pressure ($F_{1,12} = 167.3$, $p < 10^{-4}$), but also in Experiment 2 ($F_{1,12} = 9.93$, $p = 0.008$). This finding indicates that some of the primary responses were errors that could be detected prior to second-order judgment, resulting in "change-of-mind" (Resulaj et al., 2009). In sum, on *seen* trials, participants managed to perform the metacognitive task with very high accuracy.

Cognitive and metacognitive performance are above chance on unseen trials

We next performed similar analyses of cognitive and metacognitive performance restricted to the *unseen* trials.

For first-order performance, performance remained at chance level on *unseen* trials in experiment 1 (%correct = 50%, for all SOA, $p > 0.30$, Fig. 3A), presumably due to the pressure on speed. In experiment 2, when time pressure was relaxed, performance slightly surpassed 50% (%correct > 50%, for all SOA, $p < 0.05$, Fig. 3B).

These results were confirmed by an analysis of first-order d' values. In experiment 1, performance was at chance for all SOAs ($d' = 0$, all $p > 0.10$, Fig. 3C), but once speed pressure was relaxed in experiment 2 (Fig. 3D), objective performance increased with SOA ($F_{2,24} = 10.589$, $p = 0.0005$) and differed from chance for SOA 33 ms ($t_{12} = 2.99$, $p = 0.011$) and 50 ms ($t_{12} = 3.97$, $p = 0.002$). Experiment 2 thus demonstrates a classical subliminal effect (Persaud et al., 2007; Pessiglione et al., 2007), i.e. a partial accumulation of evidence about the *unseen* targets.

Most importantly, second-order performance in the error detection task (i.e. meta-performance) was significantly above chance in both experiments for intermediate SOAs (SOA 33 and 50 ms, meta-performance > 50%, all $p < 0.005$). Indeed, as shown in Figs. 3A–B, when pooling these two intermediate SOAs, a large number of correct trials were correctly classified as such (exp 1: 65.8%; exp 2: 72.9%). Again, SDT analysis confirmed this result, as *meta-d′* was significantly superior to 0 (chance level) on *unseen* trials, both in experiment 1 (SOA 16 ms: $t_{12} = 2.42$, $p = 0.032$, SOA 33 ms: $t_{12} = 2.26$, $p = 0.043$ and SOA 50 ms: $t_{12} = 3.79$, $p = 0.003$) and in experiment 2 (SOA 33 ms: $t_{12} = 3.27$, $p = 0.007$ and SOA 50 ms: $t_{12} = 4.52$, $p = 0.0007$) and seem to increase with SOA (Exp1: $F_{2,24} = 2.65$, $p = 0.091$; $F_{2,24} = 8.50$, $p = 0.002$).

Direct comparison of d' and *meta-d′* showed that, for both experiments, meta-cognitive performance exceeded primary task performance on *unseen* trials. This was true over all *unseen* trials (SOA 16–50 ms, Exp1: $F_{1,60} = 11.48$, $p = 0.005$; Exp 2: $F_{1,60} = 13.2$, $p = 0.003$), at intermediate SOAs 33 ms (Exp1: $t_{12} = -1.89$, $p = 0.041$; Exp2: $t_{12} = -1.97$, $p = 0.036$) and at SOA 50 ms (Exp1: $t_{12} = -3.28$, $p = 0.003$; Exp2: $t_{12} = -2.09$, $p = 0.023$). Even

in subliminal conditions, once a primary response is emitted, participants can categorize it as correct or incorrect with better-than-chance performance.

To summarize, we found that in both experiments, participants were above chance in judging their own errors, even on trials classified as *unseen*. Most remarkably, for subliminal stimuli in experiment 1, participants were at chance for the objective task, presumably due to time pressure, and yet they were still able to evaluate their accuracy better than chance. In experiment 2, they were above chance for both cognitive and metacognitive tasks, a result that may relate to the reduced time pressure compared to experiment 1.

The error-related negativity is present only on seen trials

We then turned to EEG recordings, in order to probe whether metacognitive performance was accompanied by an ERN, even under subliminal conditions (Fig. 4).

Starting with the *seen* trials, a significant ERN, manifested by more negative central voltages on error than on correct trials, was found in both experiments (Figs. 4A-B, Exp. 1: $t_{12} = -3.39$, $p = 0.0053$; Experiment 2: $t_{12} = -3.42$, $p = 0.0051$). Importantly, no significant difference was detectable on *unseen* trials in experiment 1 ($t_{12} = -0.55$ $p = 0.59$), suggesting that the ERN was absent under subliminal conditions. In this experiment, the number-comparison task was strongly speeded, leaving open the possibility that the results might be an artefact of time–pressure, with the response being emitted too fast to observe an ERN. However, this interpretation was rejected by experiment 2, where a similar result was observed ($t_{12} = 0.02$, $p = 0.98$) although time–pressure was relaxed and response-time was longer (see Supplementary material).

The variation of the ERN with subjective report was confirmed by a significant interaction between visibility (*seen* or *unseen*) and performance (*error* or *correct*) on central voltages in the time window of the ERN (Exp 1 $F_{1,36} = 8.62$, $p = 0.012$; Exp 2 $F_{1,36} = 10.46$, $p = 0.0072$, see Materials & methods). The ERN remained undetectable on *unseen* trials, even when we restricted the analysis to trials in which metacognitive performance was correct (see Supplementary Results) and therefore a maximal amount of stimulus information was accumulated. The absence of the ERN on these trials suggests that above-chance metacognitive performance on subliminal trials was not mediated by the ERN, which was simply absent or drastically reduced under subliminal conditions.

The ERN depends on visibility, not SOA

The above *seen*/*unseen* comparison is partially confounded with differences in SOA, as the majority of *seen* trials comes from trials with long SOAs. It could therefore be argued that the presence of the ERN on *seen* trials has nothing to do with subjective visibility, but is simply due to the additional information made available by the longer SOA (indeed, a similar confound applies to previous research by Pavone et al. (2009) and Woodman (2010). However, because we collected visibility information on every trial, our design allowed bypassing this limitation. We sorted the trials as a function of both SOA and trial-by-trial judgement of visibility, taking advantage of spontaneous fluctuations in visibility for a fixed SOA. This analysis could only be performed in experiment 1 as too few error trials occurred in experiment 2.

On *unseen* trials, a general linear model (see Materials & methods) with SOA (16, 33 or 50 ms) and performance (*correct* or *error*) as within-subject factors confirmed the absence of a difference between error and correct trials (no ERN, $p = 0.91$, Fig. 5F) and no interaction with SOA ($p = 0.76$). Indeed, none of the SOAs showed a significant ERN (all $p > 0.25$). For *seen* trials, conversely, a similar ANOVA over SOAs 33, 50, 66 and 100 ms revealed a main difference between error and correct trials ($p < 10^{-4}$, Fig. 5E). Furthermore, an interaction with SOA ($p = 0.04$) indicated that the ERN increased with SOA.

Most crucially, for SOA 50 ms, the voltage difference between correct and error trials varied drastically with visibility. No ERN was observed for *unseen* trials ($t_{10} = 0.58$, $p = 0.29$, Fig. 5F) while a clear ERN was present for *seen* trials ($t_{11} = 2.48$ $p = 0.015$, Fig. 5E). Thus, subjective visibility, over and above objective variations in SOA, determined the presence or absence of an ERN. For SOA 33 ms, the difference between error and correct trials did not reach significance neither for the *unseen* ($t_{12} = -0.23$, $p = 0.59$), nor for the *seen* trials ($t_8 = 1.16$, $p = 0.14$) probably due to the small number of participants having enough data points in this condition. Fig. 5E suggests that at this SOA, the ERN was present but temporally spread out, which we verified by observing significantly more negative voltages for errors than for correct trials once averaging over the interval 50-200 ms ($t_8 = 2.53$, $p = 0.018$). Within the *seen* trials, the error-correct difference reached significance for all other SOAs (SOA 66 ms: $t_{11} = 3.02$, $p = 0.006$; SOA 100 ms: $t_{11} = 3.37$, $p = 0.003$).

In summary, at any SOA, the ERN was present if and only if participants reported seeing the target.

MEG detects signatures of conscious and non-conscious errors

To identify the cerebral signatures of error processing, cluster analysis was applied to MEG and EEG data in order to identify any cluster of sensors showing a difference between error and correct trials. To take advantage of the possible differences in sensitivity between sensors, we analyzed separately each type of sensor (electrodes, magnetometers, longitudinal and latitudinal gradiometers) for *seen* and *unseen* trials. For EEG, cluster analysis essentially replicated the above ERN analysis. On *seen* trials, a significant cluster, with more negative voltages on error trials, was found on fronto-central electrodes in EEG, for both experiment 1 ($p = 0.0067$, Fig. 6A) and 2 ($p = 0.0013$, Fig. 6C). The cluster began at motor onset in experiment 2, and continued for 100 ms, while it started at 50 ms after the response in experiment 1. In *unseen* trials, no significant EEG cluster was detected.

For MEG, in experiment 1, significant clusters were found for two of the three types of channels in the *seen* trials (Fig. 6A, latitudinal gradiometers cluster: left fronto-lateral region, 25–70 ms after response, $p = 0.015$; magnetometers cluster: right parieto-central region, 65–90 ms, $p = 0.023$), suggesting different sensitivity to error-related signals across sensor types. Again however, no significant cluster was found for the *unseen* trials (Fig. 6B).

As time–pressure induced speeded responses in experiment 1, we then turned to experiment 2, in which more evidence should be available at response onset and error-related processes should have full ability to develop. Indeed, MEG sensors revealed a different pattern of activity for this experiment. For *seen* trials, only magnetometers (Fig. 6C) showed error-

related activity (orbito to dorso-frontal regions, 5–55 ms). More surprisingly, even for *unseen* trials, significant differences were observed in two clusters of sensors (Fig. 6D; longitudinal gradiometers, 0–65 ms, p = 0.002; magnetometers, 0–45 ms, p = 0.007), none of them resembling however with those found for the *seen* trials. These results suggest that MEG sensors may provide a more sensitive and comprehensive view of error-processes than EEG, a result that is coherent with recent studies showing accrued sensitivity of MEG sensors to sources located in the cingulate gyrus, where the generators of the ERN are thought to be located (Irimia et al., 2011). Furthermore, this analysis confirms that these error-processes are modulated by consciousness but also by time–pressure as different results were obtained in the two experiments.

Conscious error detection originates from posterior cingulate cortex

To shed more light on the cerebral generators of these error responses observed at the sensor level, we applied distributed source estimation on error and correct MEEG signals. For *seen* trials in experiment 1, the main source of the difference between error and correct trials was found bilaterally in the anterior part of the Posterior Cingulate Cortex (PCC, Fig. 7A). Its time course matched the dynamics of the ERN (Fig. 7E), and its peak coordinates (Talairach coordinates x = −6 y = −22 z = 33) felt close to a recently published MEEG and fMRI study (Agam et al., 2011). In the *unseen* condition, this activity was drastically reduced, in accordance with the absence of a significant effect at the sensor level. Lowering the threshold only revealed weak and inconsistent differences in the most posterior part of the cingulate cortex (Fig. 7C).

In experiment 2, the involvement of PCC on conscious errors was replicated (Talairach coordinates x = −9y = −23 z = 31), but additional error-related activity was also observed in dorsal anterior cingulate (dACC, Talairach peak at coordinates x = 7 y = 2 z = 27, Figs. 7B and F), explaining the observed differences in MEG sensor-level topographies in experiments 1 versus 2. Again, activation in these regions was drastically reduced for *unseen* trials. Nevertheless, small patches in dACC (Fig. 7D) remained active in the *unseen* condition, compatible with the small but significant effect detected at the sensor level in MEG data.

When further restricting the analysis to *unseen* meta-correct trials, in which performance was correctly evaluated (see Supplementary Results), time-courses indeed revealed a short-lived response (Fig. S5) in dACC coinciding with the early part of the error-related activation observed on *seen* trials. Thus, this transient dACC activation might be one of the substrates for above-chance metacognitive performance.

Discussion

In this study we explored whether the meta-cognitive process of error detection in a simple response-time decision task requires conscious perception of the stimulus in order to be deployed. We recorded brain responses in a masking paradigm with variable time–pressure and masking strength, and evaluated the relation between first-order performance, meta-cognition, and subjective visibility. Our findings indicate that two types of metacognitive processes have to be distinguished: (1) The likelihood of having made an error can be

estimated above chance level, in a statistical manner, even when making a forced-choice response to a subliminal stimulus; (2) the ERN, which reflects the detection of whether an error was made on a given trial, indexes another process that is only deployed on trials where the stimulus is consciously perceived.

Metacognition without consciousness

Behaviorally, we compared performance in the number comparison task and in the meta-performance task of detecting one's own errors. For the latter, following Maniscalco and Lau (2012), we used a *meta-d′* measure that evaluates what should have been the performance in the first-order task given the performance observed in the second order task. This method allowed us to compare, on the same scale, performance in the number comparison task (d') and performance in error detection (*meta-d′*).

In two distinct experiments, we found that participants were able to do better than chance in detecting their own performance under conscious, but also under non-conscious conditions. In Experiment 1, meta-performance in error detection exceeded performance in the first-order task, presumably because, under time–pressure, the primary response was emitted too early, and participants later revised their judgments using a more complete accumulation of evidence on the stimulus (Resulaj et al., 2009). This interpretation was supported by Experiment 2: when time–pressure was weakened, both performance and meta-performance reached above-chance levels and evolved in close parallel as a function of SOA (Fig. 3).

Crucially, participants performed above chance in detecting their own errors even on *unseen* trials. In both experiments, meta-cognitive performance on *unseen* trials increased with SOA, suggesting that longer SOAs allowed increasing amounts of evidence to be accumulated, as previously demonstrated for subliminal visual and motor processing (Del Cul et al., 2007; Vorberg and Mattler, 2003).

Our findings therefore suggest that meta-cognition should be added to the list of processes that can be partially deployed non-consciously. Such a result is in line with a previous report showing a higher-than-chance performance in metacognitive judgments of confidence under conditions of invisibility due to inattention (Kanai et al., 2010). Similarly, another study showed that a blindsight patient was able to perform above chance-level in his second-order confidence judgments, even when the stimulus was presented in his blind hemi-field (Evans and Azzopardi, 2007). Such findings contradict the view that under conditions of subjective invisibility, participants are not able to predict their accuracy in detecting a masked target. Indeed, measurement of post-error slowing suggests that participants are able to monitor their performance non-consciously, and are sensitive to their objective errors even when the experimental paradigm misleads them into thinking that their performance was correct (Logan and Crump, 2010).

These findings conflict with the common intuition according to which self-oriented monitoring processes are tightly linked to consciousness (Kunimoto et al., 2001; Lau and Passingham, 2006; Persaud et al., 2007). In particular, our finding that above-chance metacognitive judgments do not necessarily indicate conscious perception of the stimulus seems incompatible with the use of wagering or confidence as an index of consciousness

(Kunimoto et al., 2001; Persaud et al., 2007). Nonetheless, such a critique must be qualified, as above-chance subliminal metacognition is probably limited to experimental circumstances where a forced-choice judgment is imposed. Furthermore, in the present study, participants had to be explicitly informed that even when responding randomly they still had a 50% chance of being correct. Therefore they should venture "error" and "correct" responses on approximately half of trials. Prior to this instruction, a pilot study showed that most of them spontaneously responded with the "error" key on all unseen trials, suggesting a total lack of confidence in their capacity to make both first- and second-error judgments. In the same manner, blindsight patients may first have to gain an explicit awareness that their performance largely exceeds chance level before performing a second-order metacognitive task (Evans and Azzopardi, 2007). It remains unclear whether above-chance subliminal metacognitive abilities would be observed without this prior knowledge of first-order accuracy. In that sense, wagering and confidence judgments may vary more tightly with subjective reports of visibility in some contexts than others. Altogether however, these findings confirm that, as any other decision processes, second-order judgments are subject to response biases (Evans and Azzopardi, 2007; Fleming and Dolan, 2010) and should therefore be analyzed carefully to disentangle the effect of criterion setting from the true level of "meta-evidence" available about a given cognitive process.

Second-order signal detection theory (SDT) offers a theoretical framework within which to analyze such measures, and is capable of explaining both first- and second-order non-conscious performance. According to classical SDT, an observer receives a sensory sample on a continuum, and the first-order response is selected by deciding on which side of a decision boundary it falls. Second-order SDT points out that information on the distance of the sensory evidence from the decision boundary can be used to partially predict response accuracy, thus supporting a second-order judgement (Galvin et al., 2003). Intuitively, sensory evidence that falls very close to the decision boundary is highly ambiguous and will therefore likely lead to an error. In contrast, sensory evidence that falls far from the boundary is (statistically) more indicative of a correct response. According to this model, decision and confidence are therefore computed simultaneously from the same data. Previous behavioral and neural evidence (Kepecs et al., 2008; Kiani and Shadlen, 2009; Resulaj et al., 2009) supports this view. Furthermore, the theory can explain the gist of our present results: since first-order evidence towards a decision can be accumulated from *unseen* stimuli, resulting in above-chance first-order performance (Vorberg and Mattler, 2003), it follows from the theory that it should also be possible for the same system to compute second-order confidence information non-consciously — as demonstrated here.

However, the data of Experiment 1 impose a small revision on the second-order SDT mechanism proposed by Galvin et al. (2003). This theory supposes that a single sample of sensory evidence is used for both first-order and second-order tasks, predicting that meta-performance cannot exceed performance (Galvin et al., 2003). However, in Experiment 1, under strong time pressure, primary judgment was at chance while second-order performance was above chance. In that respect, our findings are reminiscent of the observation of "changes-of-mind" in a sensori-motor task, i.e. accurate corrective movements performed after the first response was launched even though no additional sensory data was provided (Resulaj et al., 2009). Both findings can be accounted for by

supposing that early responses do not fully make use of the available sensory evidence and that, with additional time, participants can accumulate additional evidence in order to ultimately revise their judgments. Indeed, when we removed time pressure in Experiment 2, both performance and meta-performance became aligned with each other (d' and *meta-d′* did not differ).

The SDT framework can be modified to take into account such dynamics of decision making (Resulaj et al., 2009). Indeed, the recently introduced Two-Stage Dynamic Signal Detection Theory (Pleskac and Busemeyer, 2010) integrates these two elements into a framework that accurately predicts both the dynamics of decision-making and subsequent confidence judgments. This model allows additional processing of the stimulus to take place even after an initial decision has been made. Such feature results in confidence judgments that can potentially rely on more information than primary choices, especially when speed is emphasized over accuracy, exactly as observed in our study.

All-or-none error detection and conscious perception

The SDT framework for metacognition is, however, inherently limited. It is continuous and statistical in nature, and cannot label, with near-certainty, whether a given trial was correct or erroneous. Rather, it merely achieves above-chance meta-performance on average. While such a statistical mechanism adequately accounts for the observed metacognitive performance on subliminal trials, it seems insufficient to explain error detection on conscious trials. When participants reported seeing the stimuli, they were often highly confident in the detection of their errors, and accurately categorized their performance on each trial in the absence of any feedback (Fig. 3). A distinct mechanism therefore seems needed to account for the capacity to label specific trials as erroneous, which only occurred on conscious trials. Indeed, EEG and MEG recordings gave evidence that a distinct performance monitoring mechanism, indexed by the ERN, was deployed exclusively on conscious trials.

In Experiment 1, the ERN was detectable on conscious trials but was drastically reduced to undetectable levels when participants reported not seeing the target. This result was confirmed by an analysis of the neural generators of the ERN, whose activation showed a step-like increase with visibility. Even for identical masking strength, the ERN was observed on *seen* trials but not on *unseen* trials. This result was replicated in Experiment 2 where the pressure to respond quickly was removed, showing that the absence of a subliminal ERN was not caused by a lack of processing time.

Our results replicate and extend prior research using a 4-dot masking task (Woodman, 2010). In this task, Woodman observed an ERN when the target was consciously perceived, but not when it was masked and became invisible. In this study, however, visibility was confounded with a physical change in the display (delayed mask offset). Our study goes beyond their finding by taking advantage of the spontaneous fluctuations in visibility that occur for a fixed stimulus. We demonstrate that the ERN is modulated purely as function of subjective reportability without any objective change in the stimulus. Our study also shows that the absence of the ERN needs not be accompanied by a lack of meta-cognitive

performance, and provides information as to the generators of these two error monitoring devices.

In contrast to the results of Woodman (2010), Pavone et al. (2009) reported the detection of a significant ERN on both unaware and aware errors, compared to correct trials. A close examination of their graphs, however, suggests that their difference might be related to pre-response baseline shifts, possibly due to the fact that response times were not equalized. Note that in our experiment, we only examined the ERPs to error and correct trials that were carefully equalized to have equal distributions of responses times (see Materials & methods). A failure to do so may result in the emergence of artifactual differences in the time course of the ERPs which are unrelated to errors themselves, but simply reflect variations in response speed between correct and error trials. If a baseline correction was applied to Pavone et al.'s results, their graphs suggest that an identical negativity would be seen on correct and erroneous subliminal trials — i.e. an absence of a subliminal ERN, similar to what we observed.

Some studies aimed at manipulating more directly the awareness of making an error which, as we noted in the Introduction, constitutes a different question. In antisaccade studies (Endrass et al., 2007; Nieuwenhuis et al., 2001; Wessel et al., 2011) an ERN has been observed when participants made eye-movement errors that were not consciously detected. The apparent conflict with our work is only superficial as in these studies the target was always consciously visible and a conscious motor intention could always be prepared. The only aspect of which participants remained unaware was the deviation of their actual movements from the intended trajectory. Their results therefore suggest that the ERN may remain present when the action itself is non-conscious. In contrast, our results suggest that the ERN vanishes when the target, and therefore the correct response, cannot be consciously represented.

Other studies (Dhar et al., 2011; Hughes and Yeung, 2011; O'Connell et al., 2007), focused exclusively on error awareness in experimental paradigms where conflicting stimulus–response rules induced confusions on the nature of the correct response. Again, they found that the ERN was present even for errors that were undetected. However it remains unclear in such paradigms whether participants were unaware of their errors because of an erroneous representation of the correct response, or because of a failure in the error-detection process itself. In either case, such results do not conflict with our finding as these studies did not manipulate awareness of the stimulus itself but rather introduced confusion on the stimulus–response mapping.

A converging finding of these studies, confirmed by others (Hewig et al., 2011; Hughes and Yeung, 2011; Steinhauser and Yeung, 2010), is that the ERN does not necessarily signal a consciously perceived error. Again, this conclusion is not incompatible with our result: while the ERN is evoked only when a conscious target is present, it may not yet reflect the conscious detection of the error. Rather, it may just index an intermediate process on the way to conscious error detection. Indeed, several recent articles suggest that error awareness might be related to the error positivity (Pe) (Dhar et al., 2011; Endrass et al., 2007; Hewig et al., 2011; Hughes and Yeung, 2011; Nieuwenhuis et al., 2001; O'Connell et al., 2007;

Steinhauser and Yeung, 2010) which follows the ERN. In that sense, the Pe may be analogous to the sensory P3 potential observed in many experiments where conscious and unconscious sensory trials are contrasted (Dehaene and Changeux, 2011). A detailed analysis of the behavior of the Pe in our two experiments, confirming the dissociation between ERN and Pe and partially supporting the above hypotheses, may be found in Supplementary materials (see also Fig. 4).

The present results further clarify the types of brain events that occur when a sensory stimulus becomes conscious and crosses the threshold for reportability. The Global Neuronal Workspace (GNW) model proposes that conscious access is associated with a sharp non-linear transition in brain activity (Dehaene and Changeux, 2011), leading to an all-or-none change in subjective reports and late brain activity on *seen* compared to *unseen* trials (Del Cul et al., 2007; Quiroga et al., 2008; Sergent and Dehaene, 2004b; Sergent et al., 2005). However, this all-or-none view has been challenged on the grounds that behavioral measures, priming, and brain activation often show a continuous rather than discontinuous reduction on subliminal relative to supraliminal trials (Dehaene et al., 1998; Overgaard et al., 2006; Van Gaal et al., 2008; Vorberg and Mattler, 2003). The present results on the ERN speak in favor of a non-linear transition between subjectively *seen* and *unseen* trials: while subliminal performance in both first- and second-order tasks increased smoothly with the target-mask delay (SOA), the ERN did not vary continuously with SOA. Instead, it jumped suddenly as a sole function of subjective visibility showing that the error-detection system reflected by the ERN was strongly impeded for subjectively invisible trials. The crossing of the subjective threshold for conscious reportability was accompanied by a step-like improvement in the availability of information and, more crucially, by the sudden emergence of the ERN. Importantly, the ERN strictly followed the subjective reports of visibility, above and beyond objective variation in stimulation.

These results were obtained by asking participants to subjectively label the trial into two categories, "*seen*" and "*unseen*". This binary visibility judgment was motivated by previous reports showing that in masking paradigms, participants focus their responses on the extreme points of a continuous scale when they are asked to report prime visibility (Sergent and Dehaene, 2004a). Our approach was also adopted for simplicity. Participants already performed no less than three responses on each trial. Requiring them to perform a more complicated visibility rating task would have lengthened the experiment even further. In the future, it might useful to examine if the present findings replicate with a more continuous estimate of visibility (Overgaard et al., 2006; Sergent and Dehaene, 2004a; Sergent et al., 2005; Seth and Dienes, 2008), thus improving our ability to detect whether the ERN follow an all-or-none pattern.

One may raise the critique that subjective reports of visibility are potentially biased and do not accurately reflect the conscious content of the subjects (Persaud et al., 2007). While the issue of finding an appropriate measure of perceptual consciousness remains debated (Lau, 2008; Overgaard et al., 2010; Persaud et al., 2007; Seth et al., 2006) and is not the subject of this study, our results argue that subjective reports provide valid data inasmuch as they correlate strongly with objective changes in behavior and brain activity. Confirming previous results (Del Cul et al., 2007, 2009), we found that visibility reports present a tight correlation

with objective performance in the number-comparison task, suggesting that participants are accurately able to monitor and report the state of their perception. Furthermore, our results suggest that subjective reports of visibility reliably index a large objective change in brain activity, namely the ERN. Even when considering only near-threshold stimuli (intermediate SOA), the ERN switched on or off in tight correlation with subjective reports of visibility or invisibility.

Our results probably go beyond what could have been found using objective measures of visibility alone. Our shortest SOA conditions correspond to fully subliminal trials (Dehaene et al., 2006), since both objective detection and task d' are indistinguishable from zero. We found that these trials are characterized by an absence of ERN and a lack of metacognitive ability. As interesting as such a result might be, it may not be unexpected, considering how much the available sensory evidence is reduced on such heavily masked trials. To determine whether the ERN can be deployed non-consciously, it is therefore crucial to focus on more lightly masked trials, where a longer SOA provides greater sensory evidence for error detection. Unfortunately, such trials provide a challenge for purely objective approaches to consciousness, as their detection d-prime is way above chance. Nevertheless, by sorting trials as a function of whether they fall above or below the threshold for conscious perception, a purely subjective criterion, we found that *unseen* trials are also characterized by an absence of ERN, while at the same time subjects remain better than chance in the metacognitive task of detecting their errors. Interestingly, we show here a complete dissociation between the continuously increasing estimation of error likelihood on *unseen* trials, and the all-or-none detection of errors reflected by the ERN on subjectively *seen* trials.

Computational models of the ERN

How do the brain generators of the ERN compute whether the response is correct or erroneous or a given trial in the absence of any experimenter feedback? Some models of the ERN postulate that it reflects a comparison (Bernstein et al., 1995; Falkenstein et al., 2000) or conflict (Veen and Carter, 2002; Yeung et al., 2004) between the actual and the intended response. How can one integrate awareness in such models? The dual-route model proposed by Del Cul et al. (2009) provides a model of how conscious and non-conscious decisions are made, and how they might be compared to yield an error signal. According to this model, two parallel routes accumulate sensory evidence towards a categorical decision on the same input stimulus. Each route has different noise levels and thresholds: One is a fast, non-conscious sensori-motor route, and one is a slower conscious decision route. A motor response is emitted by the route that first reaches its decision threshold. In the case where time–pressure in emphasized over accuracy, the response is emitted mainly via the fast and noisy motor route which is subject to non-conscious influences (Dehaene et al., 1998; Vorberg and Mattler, 2003). On such trials, the “conscious route” slowly computes the intended response (Del Cul et al., 2009). Any discrepancy between these two responses would then result in an ERN — a difference between intended and executed action. By its very nature, the model generates an ERN only when a conscious intention exists, i.e. when the second route has crossed its threshold. Thus, the model can explain the correlation between conscious perception and the presence of the ERN.

This model is compatible both with the view of the ERN as a conflict monitoring system (Veen and Carter, 2002; Yeung et al., 2004) or a comparison process (Bernstein et al., 1995; Falkenstein et al., 2000). In a similar vein, others have proposed that the ERN is a "prediction-error" signal that indexes the difference between a prediction and an observed outcome: either an ongoing response that departs from the one intended given the perceived stimulus (Alexander and Brown, 2011), or an anticipated reward that departs from the usual one expected when the response is correct (Holroyd and Coles, 2002). Assuming that such expectations are derived from a conscious-level representation of the correct intended response, these mechanisms explain why the ERN is seen only when the stimulus is consciously perceived. On *unseen* trials, no conscious intention or expectation can be computed. Accordingly, the difference process putatively indexed by the ERN is impeded, and cannot distinguish between correct and erroneous responses.

These models also predict that the ERN should vary with the amount of evidence in favor of the correct response and the confidence in the correctness of that response. Indeed, several studies demonstrated a tight correlation between subjective ratings of confidence in one's response, and the size of the ERN (Scheffers and Coles, 2000; Shalgi and Deouell, 2012; Wessel et al., 2011). Scheffers and Coles (2000) showed that for errors due to data limitation, the amplitude of the ERN was identical on correct and error trials. Even within objectively correct responses, the ERN varied massively as a function of whether subjects *believed* that they made an error. Similarly, Shalgi and Deouell (2012) found that for objective errors for which participants were highly confident in their performance rating, the ERN amplitude was predictive of whether the participant thought he had made an error or not. In particular, the ERN vanished when the participant thought he responded correctly, even though the objective performance did not change.

Apparently contradicting the finding, other studies found that it was only a later event-related potential, the Pe, which showed a systematic trial-by-trial correlation with confidence and error awareness. (Dhar et al., 2011; Hughes and Yeung, 2011; O'Connell et al., 2007). Steinhauser and Yeung (2010) demonstrated that financial rewards could shift the participants' threshold for reporting having made an error or a correct response, but that this criterion shift had no impact on the ERN itself. Hughes and Yeung (2011) also found that, while the ERN was reduced in masking conditions, the Pe was the most predictive component of error awareness. In both cases, the ERN remained invariant to changes in error awareness or in error signaling.

Taken together these findings suggest an interesting dissociation between these two components in the global system of performance monitoring. While the ERN seems to reflect a comparison or difference of intended and executed actions (Carbonnell and Falkenstein, 2006) and thus, as we suggest here, varies continuously as a function of intention strength, the Pe seems to be directly linked to the awareness of making an error (Hughes and Yeung, 2011; Nieuwenhuis et al., 2001) and its subsequent signalling (Steinhauser and Yeung, 2010). Such a model predicts that both ERN and Pe should be affected when manipulating the amount of evidence concerning the correct response (Hughes and Yeung, 2011; Maier et al., 2008; Scheffers and Coles, 2000; Shalgi and Deouell, 2012 but see Steinhauser and Yeung, 2012). However, as found by Steinhauser and

Yeung (2010), only the Pe should be changed when considering error awareness and subsequent error reportability (Hughes and Yeung, 2011; Nieuwenhuis et al., 2001; Steinhauser and Yeung, 2010). Further analysis of our data on the Pe time-window tended to confirm this hypothesis. While such a model remains speculative and will require further studies to be validated, the present findings provide converging evidence on the role of the ERN in the hierarchy of processes leading to error detection.

Brain regions involved in error monitoring

What brain mechanisms underlie conscious versus non-conscious metacognitive computations? Our results show that error detection is independent of the ERN on *unseen* trials. In both experiments, no ERN was present on *unseen* trials, even when participants correctly evaluated their own performance. In fact, we observed a double dissociation between the ERN and behavioral error detection: no ERN was observed when meta-performance exceeded performance in non-conscious trials (Experiment 1) while the ERN was present even though meta-performance was aligned on performance in conscious trials (Experiment 2). Source reconstruction of the MEEG signal confirmed that activity in one of the main generators of the ERN, the posterior cingulate cortex (PCC) (Agam et al., 2011; Dhar et al., 2011; Schie et al., 2004), was drastically reduced in the *unseen* condition.

However, on *unseen* trials, brain activity correlating with performance was observed for some of the MEG sensors. Source analysis revealed that this signal originated from the dorsal anterior cingulate cortex (dACC), a region also known to activate after errors (Debener et al., 2005; Dehaene et al., 1994; Keil et al., 2010). Importantly, this activation was present only when time–pressure was relaxed (Experiment 2) and response-times longer, highlighting its sensitivity to evidence accumulation. Activity in this region might thus convey some non-conscious information on the level of confidence in the current response, possibly explaining the participants' subliminal meta-cognitive ability. Note that this brain signal is short-lived and thus may not be sufficient to fully explain above-chance metacognitive responses occurring several hundreds of milliseconds later. However, this activity might be the input to other brain processes that compute the final judgment of confidence in one's response. Brodmann's area 10 is a plausible candidate, as several imaging studies associate it with confidence judgments (Fleming et al., 2010; Rolls et al., 2010; Yokoyama et al., 2010).

Although dACC has long been proposed to be the sole generator of the ERN (Debener et al., 2005; Dehaene et al., 1994; Emeric et al., 2008), our results are compatible with recent evidence suggesting that PCC might be another plausible source for the ERN (Agam et al., 2011; Munro et al., 2007; Vlamings, 2008). Both PCC and dACC have been shown to be active in several error-processing studies (Fassbender et al., 2004; Wittfoth et al., 2008). However it has been suggested that dACC could not only reflect error detection process but might be related to behavioral adjustment such as error avoidance (Magno et al., 2006), mapping between stimulus and response (Williams et al., 2004) and reward prediction-error (Kennerley et al., 2011). Furthermore, dACC has been shown to be activated on conflict trials independently of objective accuracy (Ullsperger and Von Cramon, 2001). Because functional connectivity analyses show that both PCC and dACC are part of a larger

functional network (Agam et al., 2011) and share direct anatomical connections (Vogt et al., 2006), it is therefore likely that these regions are both active when an error is made, as suggested by the present MEEG source modelling of experiment 2. Nonetheless, they might have different roles in performance monitoring. A possible framework to explain our data could be that, while PCC directly detects the commission of an error (Agam et al., 2011; Munro et al., 2007; Vlamings, 2008), dACC integrates this information to implement corrective behavior (Modirrousta and Fellows, 2008) and further monitoring processes. While more studies will be needed to pinpoint the functional architecture of cingulate cortex, the present results suggest an interesting difference in sensitivity to conscious versus non-conscious choices for posterior versus anterior cingulate cortex, in keeping with speculations as to the role of the PCC as a crucial node for conscious awareness (Immordino-Yang et al., 2009; Vogt and Laureys, 2009).

Conclusion

Our study suggests the existence of at least two meta-cognitive systems for performance monitoring. One of them is capable of being deployed non-consciously, but it only provides statistical information on the likelihood of having made an error. The other, associated with the ERN, shows an all-or-none signal specifically on error trials where the target was consciously perceived, making it possible for participants to realize their error. By demonstrating the co-existence of these two mechanisms, we provide new evidence on the global architecture of cognitive control and its link to consciousness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the NeuroSpin infrastructure groups, in particular to the doctors Ghislaine Dehaene-Lambertz, Andreas Kleinschmidt, Caroline Huron, Lucie Hertz-Pannier and the nurses Véronique Joly-Testault and Laurence Laurier, for their support in participant recruitment and testing; Virginie van Wassenhove, Marco Buiatti, Leila Rogeau, Etienne Labyt and all the MEG team for their help on technical difficulties; Lauri Parkkonen, Alexandre Gramfort and François Tadel for their assistance on MEEG analysis and source reconstruction; Aaron Schurger and Christophe Pallier for their advice statistical issues; Moti Salti and Simon van Gaal for useful discussions.

This project was supported by a PhD grant of the Direction Générale de l'Armement (DGA, Didier Bazalgette) and a senior grant of the European Research Council to S.D. (NeuroConsc program), as part of a general research program on functional neuroimaging of the human brain (Denis Le Bihan). The NeuroSpin MEG facility was sponsored by grants from INSERM, CEA, the Fondation pour la Recherche Médicale, the Bettencourt-Schueller foundation, and the Région île-de-France. F.V.O. is a Postdoctoral Fellow of the Research Foundation — Flanders (FWO-Vlaanderen). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abbreviations

ERN	Error-Related Negativity
ERP	event-related potential
ERF	event-related field
SDT	Signal Detection Theory

SOA	stimulus onset asynchrony
MEEG	simultaneous magneto- and electroencephalography

References

Agam Y, Hamalainen M, Lee ACH, Dyckman KA, Friedman JS, Isom M, Makris N, Manoach DS. Multimodal neuroimaging dissociates hemodynamic and electrophysiological correlates of error processing. Proc Natl Acad Sci. 2011; 108:17556–17561. [PubMed: 21969565]

Alexander WH, Brown JW. Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci. 2011; 14:1338–1344. [PubMed: 21926982]

Aly M, Yonelinas AP. Bridging consciousness and cognition in memory and perception: evidence for both state and strength processes. PLoS One. 2012; 7:e30231. [PubMed: 22272314]

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang. 2008; 59:390–412.

Bernstein PS, Scheffers MK, Coles MGH. "Where did I go wrong?" A psychophysiological analysis of error detection. J Exp Psychol Hum Percept Perform. 1995; 21:1312–1322. [PubMed: 7490583]

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. Psychol Rev. 2001; 108:624–652. [PubMed: 11488380]

Carbonnell L, Falkenstein M. Does the error negativity reflect the degree of response conflict? Brain Res. 2006; 1095:124–130. [PubMed: 16712810]

Debener S, Ullsperger M, Siegel M, Fiehler K, Von Cramon DY, Engel AK, Von Cramon DY. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci. 2005; 25:11730–11737. [PubMed: 16354931]

Dehaene S, Changeux J-P. Experimental and theoretical approaches to conscious processing. Neuron. 2011; 70:200–227. [PubMed: 21521609]

Dehaene S, Posner MI, Tucker DM. Localization of a neural system for error detection and compensation. Psychol Sci. 1994; 5:303–305.

Dehaene S, Naccache L, Le Clec'H G, Koechlin E, Mueller M, Dehaene-Lambertz G, Van de Moortele PF, Le Bihan D. Imaging unconscious semantic priming. Nature. 1998; 395:597–600. [PubMed: 9783584]

Dehaene S, Naccache L, Cohen L, Le Bihan D, Mangin JF, Poline J-B, Riviere D. Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci. 2001; 4:752–758. [PubMed: 11426233]

Dehaene S, Changeux J-P, Naccache L, Sackur J, Sergent C. Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends Cogn Sci. 2006; 10:204–211. [PubMed: 16603406]

Del Cul A, Baillet S, Dehaene S. Brain dynamics underlying the nonlinear threshold for access to consciousness. PLoS Biol. 2007; 5:2408–2423.

Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A. Causal role of prefrontal cortex in the threshold for access to consciousness. Brain. 2009; 132:2531–2540. [PubMed: 19433438]

Dhar M, Wiersema JR, Pourtois G. Cascade of neural events leading from error commission to subsequent awareness revealed using EEG source imaging. PLoS One. 2011; 6:e19578. [PubMed: 21573173]

Emeric EE, Brown JW, Leslie M, Pouget P, Stuphorn V, Schall JD. Performance monitoring local field potentials in the medial frontal cortex of primates: anterior cingulate cortex. J Neurophysiol. 2008; 99:759–772. [PubMed: 18077665]

Endrass T, Reuter B, Kathmann N. ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. Eur J Neurosci. 2007; 26:1714–1720. [PubMed: 17880402]

Evans S, Azzopardi P. Evaluation of a 'bias-free' measure of awareness. Spat Vis. 2007; 20(20):61–77. [PubMed: 17357716]

Falkenstein M, Hoormann J, Christ S, Hohnsbein J. ERP components on reaction errors and their functional significance: a tutorial. Biol Psychol. 2000; 51:87–107. [PubMed: 10686361]

Europe PMC Funders Author Manuscripts

Fassbender C, Murphy K, Foxe JJ, Wylie GR, Javitt DC, Robertson IH, Garavan H. A topography of executive functions and their interactions revealed by functional magnetic resonance imaging. Brain Res Cogn Brain Res. 2004; 20:132–143. [PubMed: 15183386]

Fleming SM, Dolan RJ. Effects of loss aversion on post-decision wagering: implications for measures of awareness. Conscious Cogn. 2010; 19:352–363. [PubMed: 20005133]

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. Relating introspective accuracy to individual differences in brain structure. Science. 2010; 329:1541–1543. [PubMed: 20847276]

Galvin SJ, Podd JV, Drga V, Whitmore J. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. Psychon Bull Rev. 2003; 10:843–876. [PubMed: 15000533]

Gehring WJ, Fencsik D. Functions of the medial frontal cortex in the processing of conflict and errors. J Neurosci. 2001; 21:9430. [PubMed: 11717376]

Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E. A neural system for error detection and compensation. Psychol Sci. 1993; 4:385–390.

Hewig J, Coles M, Trippe R. Dissociation of Pe and ERN/Ne in the conscious recognition of an error. Psychophysiology. 2011:1–7.

Holroyd CB, Coles MGH. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev. 2002; 109:679–709. [PubMed: 12374324]

Hughes G, Yeung N. Dissociable correlates of response conflict and error awareness in error-related brain activity. Neuropsychologia. 2011; 49:405–415. [PubMed: 21130788]

Immordino-Yang MH, McColl A, Damasio H, Damasio A. Neural correlates of admiration and compassion. Proc Natl Acad Sci U S A. 2009; 106:8021–8026. [PubMed: 19414310]

Irimia A, Van Horn JD, Halgren E. Source cancellation profiles of electroen-cephalography and magnetoencephalography. NeuroImage. 2011; 59:2464–2474. [PubMed: 21959078]

Kanai R, Walsh V, Tseng C-H. Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. Conscious Cogn. 2010; 19:1045–1057. [PubMed: 20598906]

Keil J, Weisz N, Paul-Jordanov I, Wienbruch C. Localization of the magnetic equivalent of the ERN and induced oscillatory brain activity. NeuroImage. 2010; 51:404–411. [PubMed: 20149884]

Kennerley SW, Behrens TEJ, Wallis JD. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci. 2011; 14:1581–1589. [PubMed: 22037498]

Kentridge R, Heywood C. The status of blindsight: near-threshold vision, islands of cortex and the Riddoch phenomenon. J Conscious Stud. 1999; 6:3–11.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. Nature. 2008; 455:227–231. [PubMed: 18690210]

Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. Science. 2009; 324:759–764. [PubMed: 19423820]

Kouider S, Dehaene S. Levels of processing during non-conscious perception: a critical review of visual masking. Philos Trans R Soc Lond B: Biol Sci. 2007; 362:857–875. [PubMed: 17403642]

Kunimoto C, Miller J, Pashler HE. Confidence and accuracy of near-threshold discrimination responses. Conscious Cogn. 2001; 10:294–340. [PubMed: 11697867]

Lau HC. A higher order Bayesian decision theory of consciousness. Prog Brain Res. 2008; 168:35–48. [PubMed: 18166384]

Lau HC, Passingham RE. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proc Natl Acad Sci U S A. 2006; 103:18763–18768. [PubMed: 17124173]

Lau HC, Passingham RE. Unconscious activation of the cognitive control system in the human prefrontal cortex. J Neurosci. 2007; 27:5805–5811. [PubMed: 17522324]

Lau HC, Rosenthal D. Empirical support for higher-order theories of conscious awareness. Trends Cogn Sci. 2011; 15:365–373. [PubMed: 21737339]

Logan GD, Crump MJC. Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. Science. 2010; 330:683. [PubMed: 21030660]

Magno E, Foxe JJ, Molholm S, Robertson IH, Garavan H. The anterior cingulate and error avoidance. J Neurosci. 2006; 26:4769–4773. [PubMed: 16672649]

Maier ME, Steinhauser M, Hubner R. Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. J Cogn Neurosci. 2008; 20:2263–2273. [PubMed: 18457501]

Maniscalco B, Lau HC. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious Cogn. 2012; 21:422–430. [PubMed: 22071269]

Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007; 164:177–190. [PubMed: 17517438]

Melloni L, Molina C, Pena M, Torres D, Singer W, Rodriguez E. Synchronization of neural activity across cortical areas correlates with conscious perception. J Neurosci. 2007; 27:2858–2865. [PubMed: 17360907]

Miltner WH, Lemke U, Weiss T, Holroyd CB, Scheffers MK, Coles MGH. Implementation of error-processing in the human anterior cingulate cortex: a source analysis of the magnetic equivalent of the error-related negativity. Biol Psychol. 2003; 64:157–166. [PubMed: 14602360]

Modirrousta M, Fellows LK. Dorsal medial prefrontal cortex plays a necessary role in rapid error prediction in humans. J Neurosci. 2008; 28:14000–14005. [PubMed: 19091989]

Munro GES, Dywan J, Harris GT, McKee S, Unsal A, Segalowitz SJ. ERN varies with degree of psychopathy in an emotion discrimination task. Biol Psychol. 2007; 76:31–42. [PubMed: 17604898]

Nieuwenhuis S, Ridderinkhof KR, Blom JH, Band GPH, Kok A. Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. Psychophysiology. 2001; 38:752–760. [PubMed: 11577898]

Nieuwenhuis S, Schweizer TS, Mars RB, Botvinick MM, Hajcak G. Error-likelihood prediction in the medial frontal cortex: a critical evaluation. Cereb Cortex. 2007; 17:1570–1581. [PubMed: 16956979]

O'Connell RG, Dockree PM, Bellgrove MA, Kelly SP, Hester R, Garavan H, Robertson IH, Foxe JJ. The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. Eur J Neurosci. 2007; 25:2571–2579. [PubMed: 17445253]

Overgaard M, Rote J, Mouridsen K, Ramsøy TZ. Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. Conscious Cogn. 2006; 15:700–708. [PubMed: 16725347]

Overgaard M, Timmermans B, Sandberg K, Cleeremans A. Optimizing subjective measures of consciousness. Conscious Cogn. 2010; 19:682–686. [PubMed: 20097582]

Pavone EFEF, Marzi CA, Girelli M. Does subliminal visual perception have an error-monitoring system? Eur J Neurosci. 2009; 30:1424–1431. [PubMed: 19788580]

Persaud N, McLeod P, Cowey A. Post-decision wagering objectively measures awareness. Nat Neurosci. 2007; 10:257–261. [PubMed: 17237774]

Pessiglione M, Schmidt L, Draganski B, Kalisch R, Lau HC, Dolan RJ, Frith CD. How the brain translates money into force: a neuroimaging study of subliminal motivation. Science. 2007; 316:904. [PubMed: 17431137]

Pleskac TJ, Busemeyer JR. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol Rev. 2010; 117:864–901. [PubMed: 20658856]

Province JM, Rouder JN. Evidence for discrete-state processing in recognition memory. Proc Natl Acad Sci U S A. 2012; 109:14357–14362. [PubMed: 22908285]

Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I. Human single-neuron responses at the threshold of conscious recognition. Proc Natl Acad Sci U S A. 2008; 105:3599–3604. [PubMed: 18299568]

Resulaj A, Kiani R, Wolpert DM, Shadlen MN. Changes of mind in decision-making. Nature. 2009; 461:263–266. [PubMed: 19693010]

Rolls ET, Grabenhorst F, Deco G. Choice, difficulty, and confidence in the brain. NeuroImage. 2010; 53:694–706. [PubMed: 20615471]

Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau HC. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cogn Neurosci. 2010; 1:165–175. [PubMed: 24168333]

Scheffers MK, Coles MGH. Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. J Exp Psychol Hum Percept Perform. 2000; 26:141–151. [PubMed: 10696610]

Van Schie HT, Mars RB, Coles MGH, Bekkering H, Van Schie HT. Modulation of activity in medial frontal and motor cortices during error observation. Nat Neurosci. 2004; 7:549–554. [PubMed: 15107858]

Sergent C, Dehaene S. Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. Psychol Sci. 2004a; 15:720–728. [PubMed: 15482443]

Sergent C, Dehaene S. Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. J Physiol Paris. 2004b; 98:374–384. [PubMed: 16293402]

Sergent C, Baillet S, Dehaene S. Timing of the brain events underlying access to consciousness during the attentional blink. Nat Neurosci. 2005; 8:1391–1400. [PubMed: 16158062]

Seth AK, Dienes Z. Measuring consciousness: relating behavioural and neurophysiological approaches. Trends Cogn Sci. 2008; 12:314–321. [PubMed: 18606562]

Seth AK, Izhikevich E, Reeke GN, Edelman GM. Theories and measures of consciousness: an extended framework. Proc Natl Acad Sci U S A. 2006; 103:10799–10804. [PubMed: 16818879]

Shalgi S, Deouell LY. Is any awareness necessary for an Ne? Front Hum Neurosci. 2012; 6:1–15. [PubMed: 22279433]

Steinhauser M, Yeung N. Decision processes in human performance monitoring. J Neurosci. 2010; 30:15643–15653. [PubMed: 21084620]

Steinhauser M, Yeung N. Error awareness as evidence accumulation: effects of speed-accuracy trade-off on error signaling. Front Hum Neurosci. 2012; 6:240. [PubMed: 22905027]

Ullsperger M, Von Cramon DY. Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. NeuroImage. 2001; 14:1387–1401. [PubMed: 11707094]

Van den Bussche E, Notebaert K, Reynvoet B. Masked primes can be genuinely semantically processed: a picture prime study. Exp Psychol. 2009; 56:295–300. [PubMed: 19447745]

Van Gaal S, Ridderinkhof KR, Fahrenfort JJ, Scholte HS, Lamme VAF. Frontal cortex mediates unconsciously triggered inhibitory control. J Neurosci. 2008; 28:8053–8062. [PubMed: 18685030]

Van Veen V, Carter CS. The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiol Behav. 2002; 77:477–482. [PubMed: 12526986]

Vlamings P. Reduced error monitoring in children with autism spectrum disorder: an ERP study. Eur J Neurosci. 2008; 28:399–406. [PubMed: 18702711]

Vogt BA, Laureys S. Posterior cingulate, precuneal & retrosplenial cortices: cytology & components of the neural network correlates of consciousness. Brain. 2009; 6123:205–217.

Vogt B, Vogt L, Laureys S. Cytology and functionally correlated circuits of human posterior cingulate areas. NeuroImage. 2006; 29:452–466. [PubMed: 16140550]

Vorberg D, Mattler U. Different time courses for visual perception and action priming. Proc Natl Acad Sci. 2003; 100:6275–6280. [PubMed: 12719543]

Weiskrantz L. Blindsight revisited. Curr Opin Neurobiol. 1996; 6:215–220. [PubMed: 8725963]

Wessel J, Danielmeier C, Ullsperger M. Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. J Cogn Neurosci. 2011; 23:3021–3036. [PubMed: 21268673]

Williams ZM, Bush G, Rauch SL, Cosgrove GR, Eskandar EN. Human anterior cingulate neurons and the integration of monetary reward with motor responses. Nat Neurosci. 2004; 7:1370–1375. [PubMed: 15558064]

Wittfoth M, Küstermann E, Fahle M, Herrmann M. The influence of response conflict on error processing: evidence from event-related fMRI. Brain Res. 2008; 1194:118–129. [PubMed: 18177843]

Woodman GFF. Masked targets trigger event-related potentials indexing shifts of attention but not error detection. Psychophysiology. 2010; 47:410–414. [PubMed: 20070578]

Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev. 2004; 111:931–959. [PubMed: 15482068]

Yokoyama O, Miura N, Watanabe J, Takemoto A, Uchida S, Sugiura M, Horie K, Sato S, Kawashima R, Nakamura K. Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. Neurosci Res. 2010; 68:199–206. [PubMed: 20688112]

Europe PMC Funders Author Manuscripts

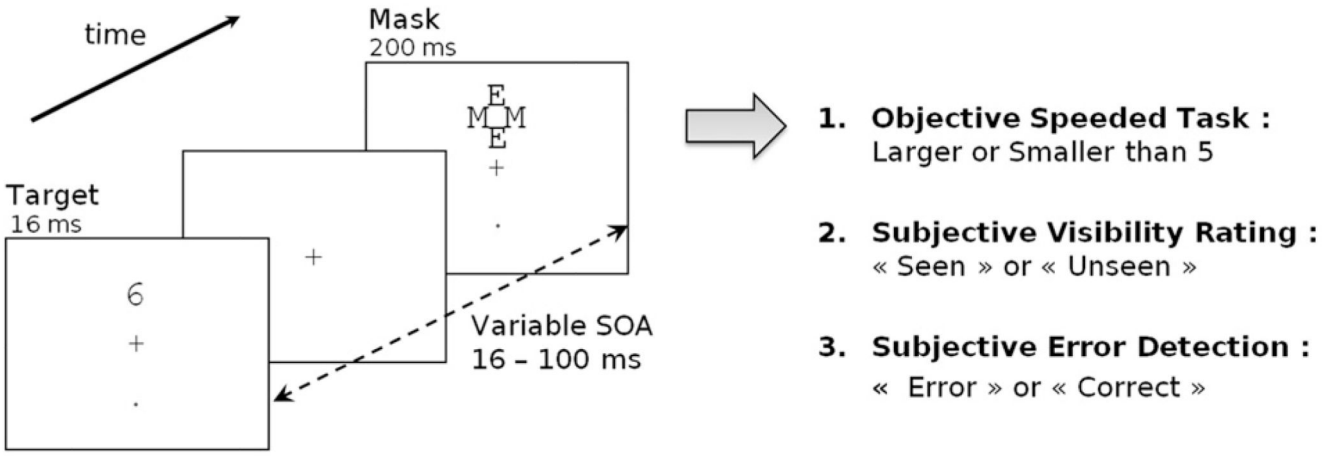

Fig. 1.

Experimental design: On each trial, a number was presented for 16 ms at one of two possible locations (top or bottom). It was followed by a mask composed of a fixed array of letters centered on the target location. The delay between target onset and mask onset (SOA) varied randomly across trials (16, 33, 50, 66 or 100 ms). In one sixth of the trials, the mask was presented alone (mask only condition). Participants first performed an objective forced-choice number comparison task where they decided whether the number was smaller or larger than 5. In experiment 1, the response had to be made in less than 550 ms, otherwise a negative sound was emitted. In experiment 2, participants were simply instructed to respond as fast as they could while maintaining accuracy. Then, on each trial, participants performed two subjective tasks. First they evaluated the subjective visibility of the target by choosing between the words "*Seen*" and "*Unseen*", displayed randomly either left or right of fixation. Second, they evaluated their own performance in the primary number comparison task by choosing between the words "Correct" and "Error", again displayed randomly either left or right.

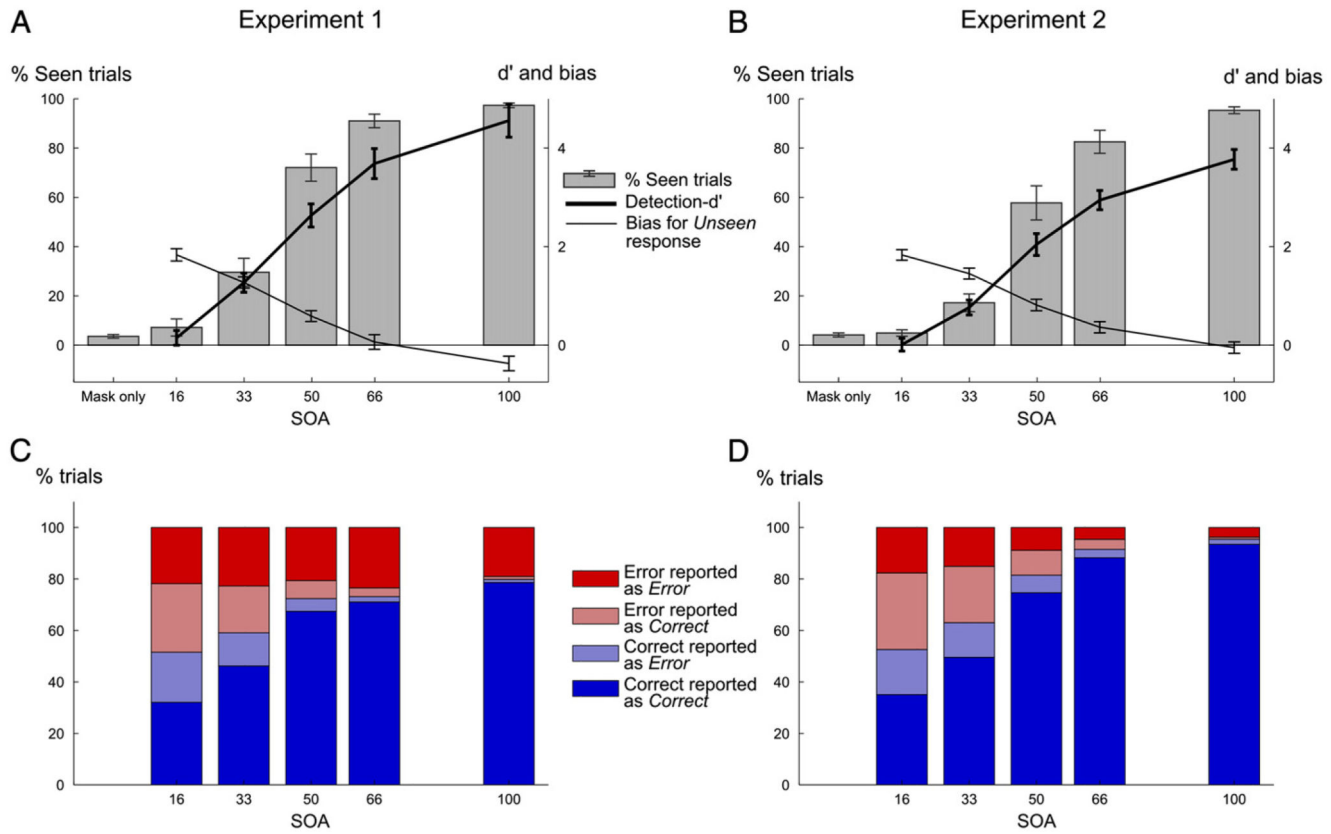

Fig. 2.

Visibility and performance results according to SOA for experiment 1 (left column) and 2 (right column). (A–B) Visibility ratings, expressed as the proportion of *seen* responses (left axis ranging from 0 to 100%) as a function of SOA. The thick line represents *detection-d'* values (right axis, ranging from 0 to 4) while the thin line represents response bias towards *unseen* response (same scale as *detection-d'*), for each SOA. (C-D) Percentage of each category of trials according to actual objective performance and subjective report of performance (Error trials correctly classified as Error in dark red, Correct trials correctly classified as Correct in dark blue, Error trials incorrectly classified as Correct in light red and Correct trials incorrectly classified as Error in light blue), for each SOA.

Europe PMC Funders Author Manuscripts

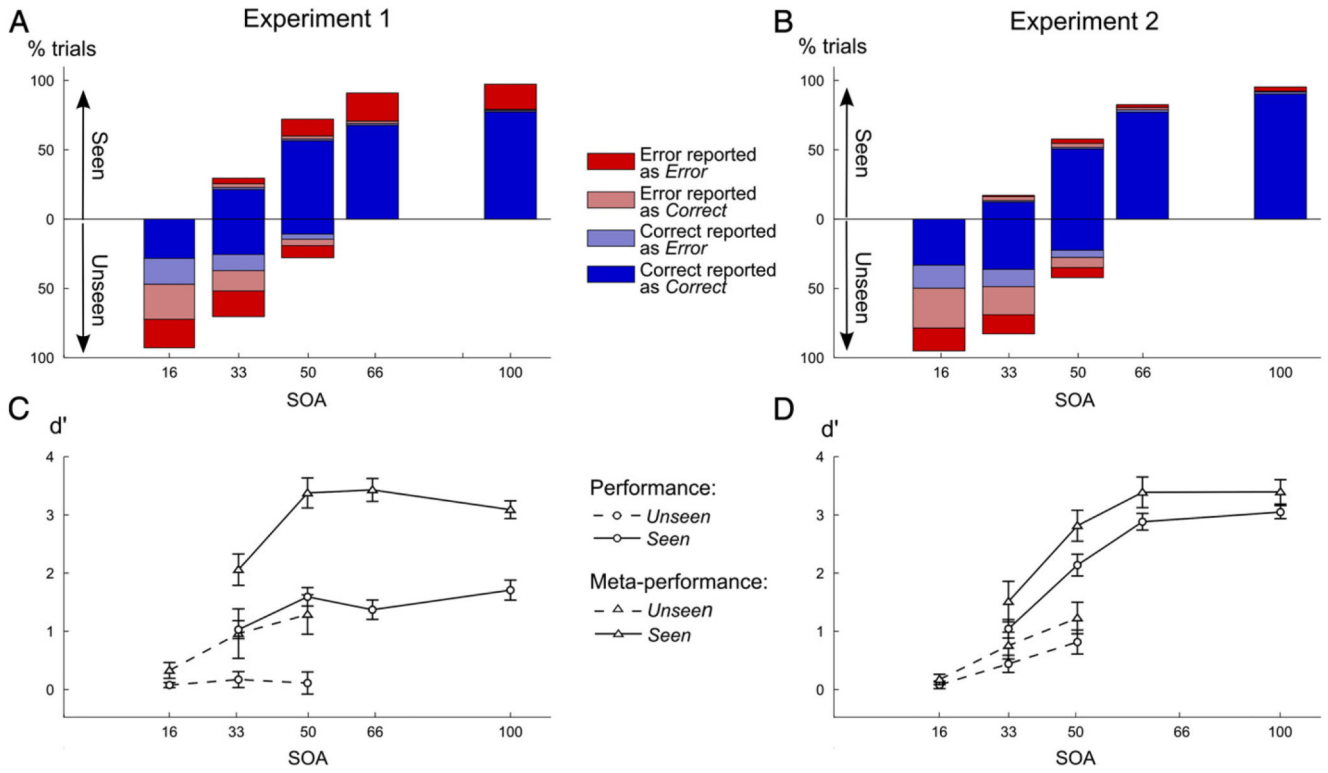

Fig. 3.

Performance and meta-performance according to visibility and SOA in both experiments (left column, experiment 1; right column, experiment 2). (A–B) Proportions of *unseen* (below midline) and *seen* trials (above midline) were computed for each SOA. For each type of trials and each SOA, the relative percentage of each category of trials was derived according to objective performance and subjective report of performance (same color code as in Fig. 2). (C-D) Unbiased measures of performance (d', circles) and meta-performance (*meta-*d', triangles) were computed separately for *seen* (solid line) and *unseen* (dashed-line) trials and each SOA value. All error-bars represent standard error.

Europe PMC Funders Author Manuscripts

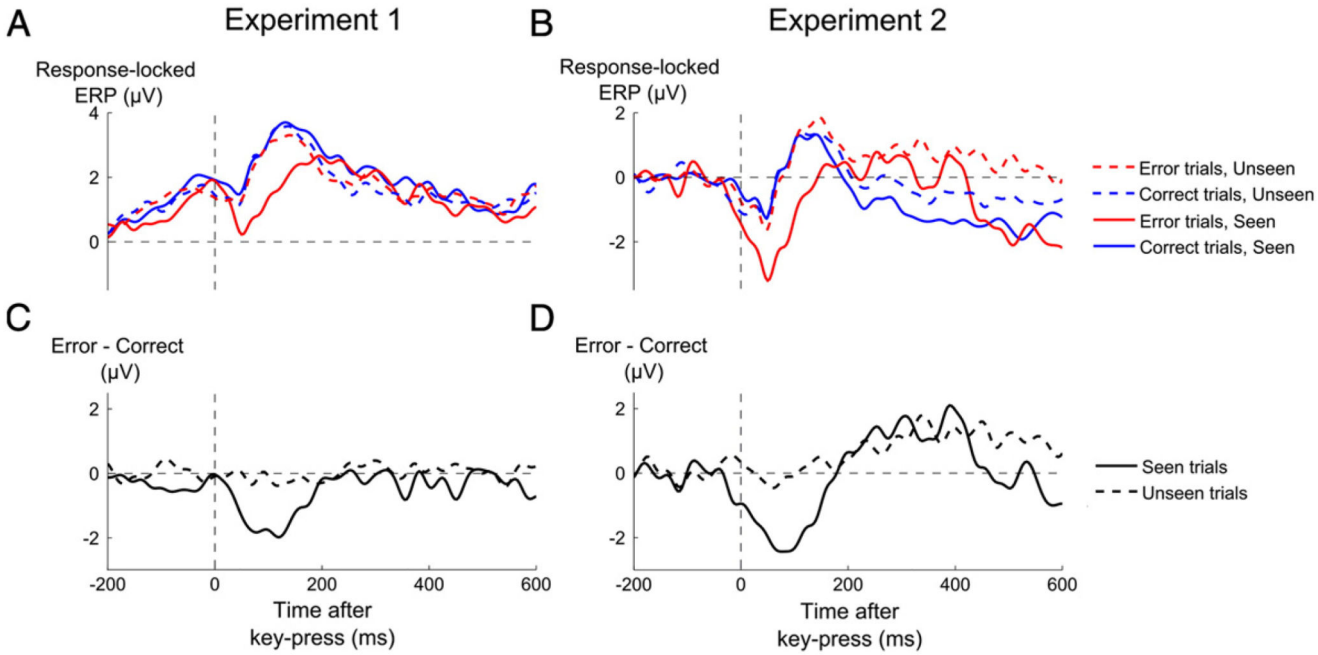

Fig. 4.

Time courses of event-related potentials as a function of objective performance and visibility. (A,B) Grand-average event-related potentials (ERPs) recorded from a cluster of central electrodes (FC1, FC2, C1, Cz, C2), sorted as a function of whether performance was erroneous (red lines) or correct (blue lines), and whether the target was *seen* (solid lines) or *unseen* trials (dashed lines), for experiment 1 (A) and experiment 2 (B). (C,D) Difference waveforms of error minus correct trials, separately for *seen* (solid line) and *unseen* (dashed line) trials.

Europe PMC Funders Author Manuscripts

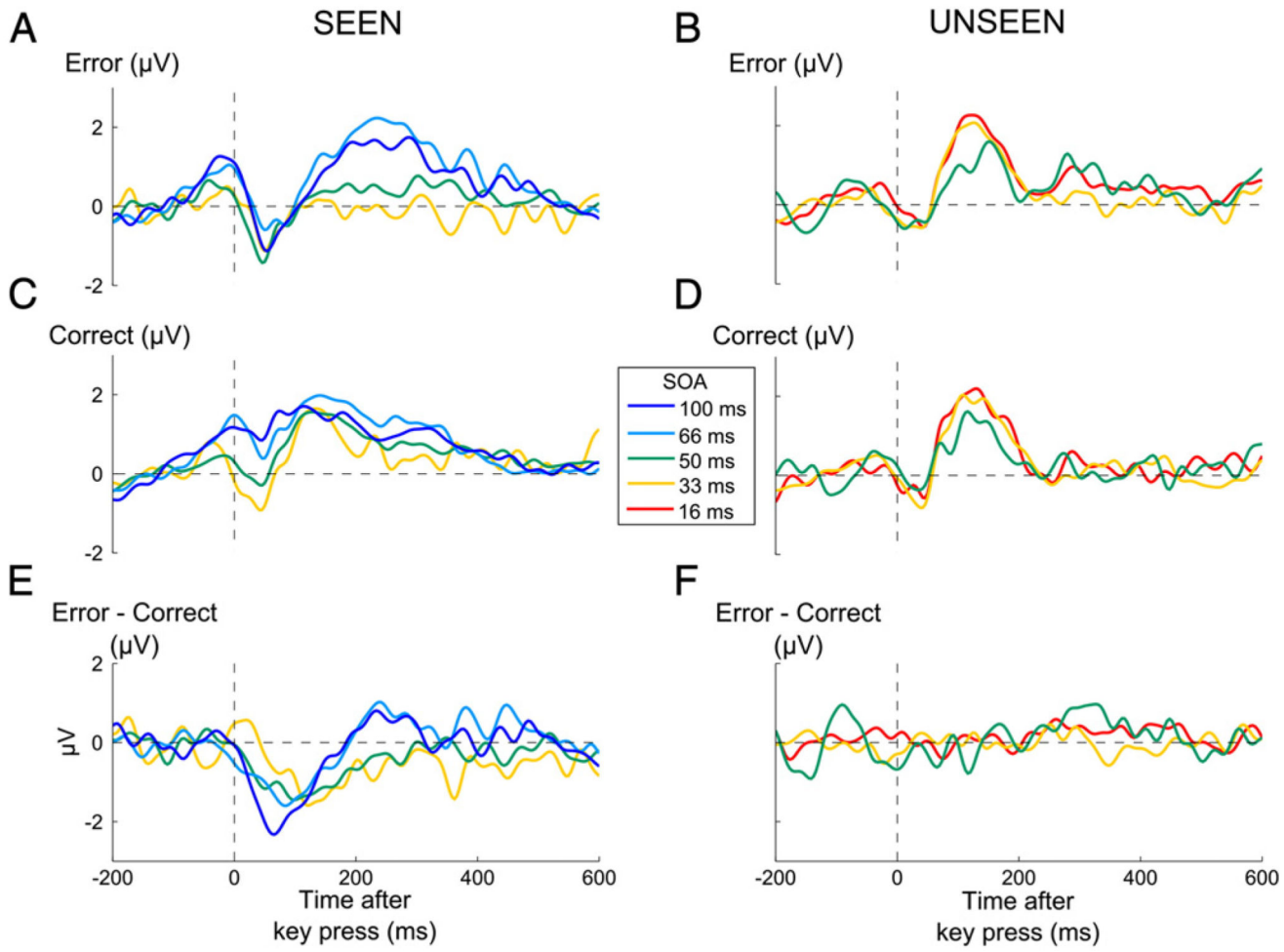

Fig. 5.

Time courses of event-related potentials as a function of SOA and objective performance for *seen* and *unseen* trials. (A–D) Grand-average event-related potentials (ERPs) by SOA condition for error (top raw, A and B) and correct (middle raw, C and D) trials in s*een* (left column, A and C) and *unseen* (right column, B and D) conditions for experiment 1 on a cluster of central electrodes (FC1, FC2, C1, Cz, C2). (E,F) Difference waveforms of error minus correct for *seen* (solid line) and *unseen* (dashed line) trials, by SOA. Due to reduced trial numbers, only the shortest SOA (16, 33 and 50) ms are presented for *unseen* trials while only longer SOAs (33 ms, 50 ms, 66 ms and 100 ms) are included for s*een* trials.

Europe PMC Funders Author Manuscripts

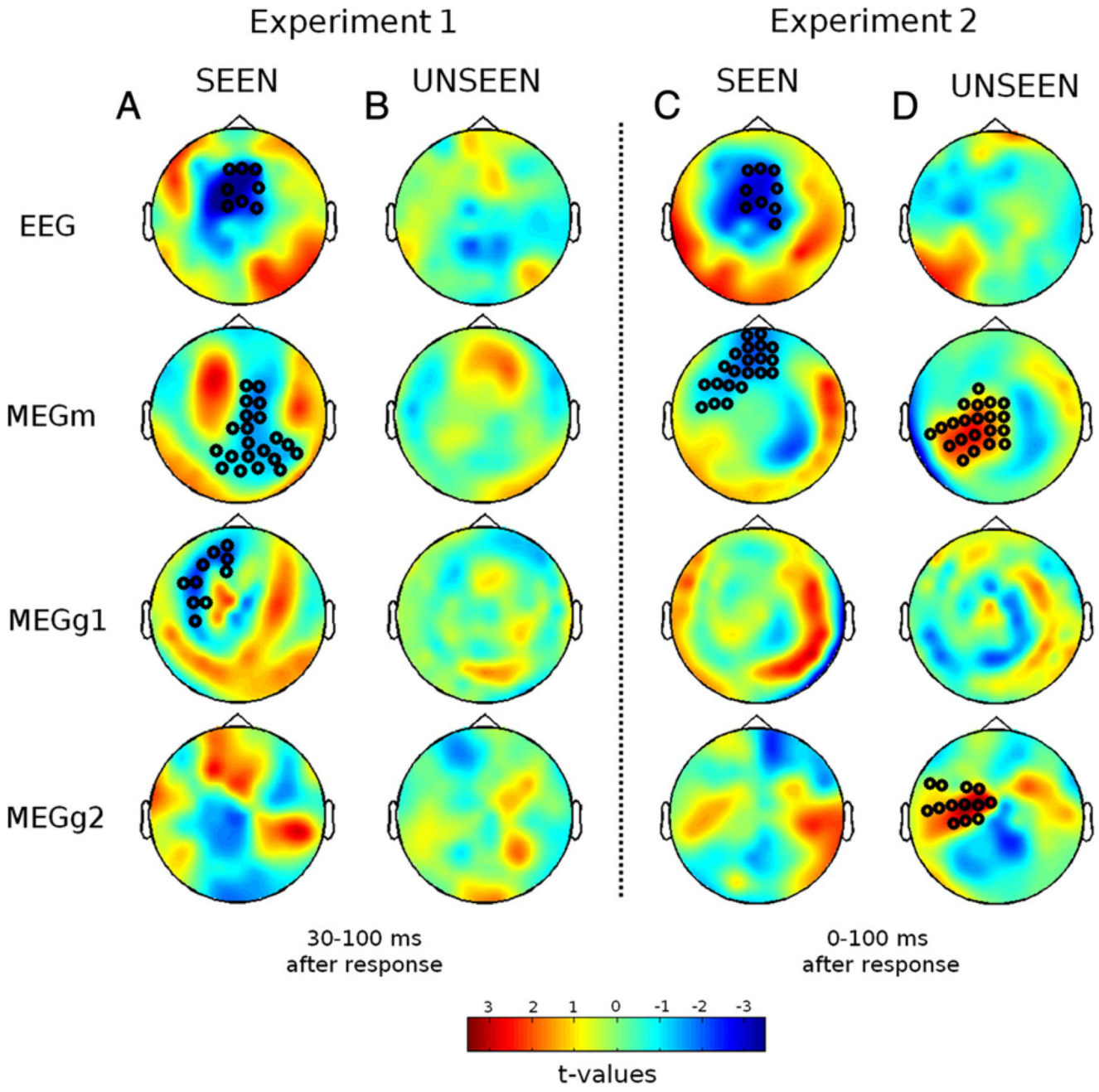

Fig. 6.

Error-related MEEG topographies as a function of target visibility. Each plot depicts the scalp topography of the t-value for a difference between correct and error trials, averaged across a 30–100 ms time window for experiment 1 and 0–100 ms for experiment 2 following the motor response, separately for each type of sensors (EEG, magnetometers [MEGm], longitudinal gradiometers [MEGg1], latitudinal gradiometers [MEGg2]) and for the *seen* and *unseen* trials, in experiments 1 (A) and 2 (B). Black circles indicate sensors belonging to a spatiotemporal cluster showing a significant difference ($p < 0.025$) between error and correct conditions using a Monte-Carlo permutation test.

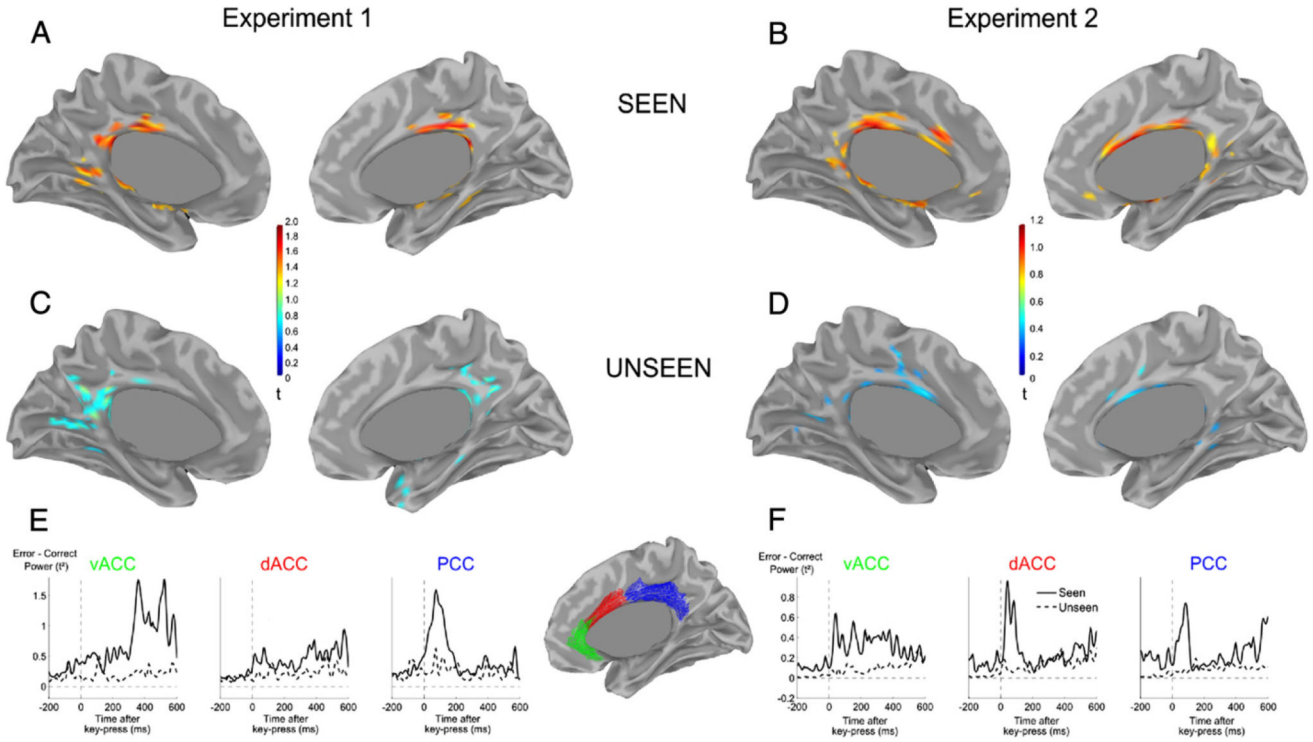

Fig. 7.

Difference of source estimates between error and correct MEEG signals. (A–D) View of the medial surface of the left and right hemispheres, for experiment 1 (A,C) and experiment 2 (B,D), for *seen* (A–B) and *unseen* (C–D) trials. Data are thresholded at 66% of maximum activity within each condition. Brain activity was averaged in a 30–100 ms time-window for experiment 1 (A,C) and 0–100 ms for experiment 2 (B,D). (E–F) Time-courses of brain activity in three bilateral regions of interest located in ventral Anterior Cingulate Cortex (vACC), dorsal Anterior Cingulate Cortex (dACC) and Posterior Cingulate Cortex (PCC), for experiment 1 (E) and experiment (2), for *seen* (solid-line) and *unseen* (dashed-line) trials. Values correspond to instantaneous power in the region of interest (average, across vertices, of the square current density t-maps).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Statistical analyses of performance and meta-performance scores, relative to chance level, as a function of visibility, for experiment 1 and 2.

		Pooling all SOAs		**SOA 33 ms**		**SOA 50 ms**	
Performance	exp 1	$t_{12} = 10.5$	$p < 10^{-4}$	$t_{12} = 5.20$	$p < 10^{-4}$	$t_{12} = 6.9921$	$p < 10^{-4}$
	exp 2	$t_{12} = 12.5$	$p < 10^{-4}$	$t_{12} = 3.70$	$p = 0.0015$	$t_{12} = 5.08$	$p = 0.0001$
Meta-performance	exp 1	$t_{12} = 9.42$	$p < 10^{-4}$	$t_{12} = 2.719$	$p = 0.0093$	$t_{12} = 4.507$	$p = 0.0003$
	exp 2	$t_{12} = 8.73$	$p < 10^{-4}$	$t_{12} = 1.677$	$p = 0.0597$	$t_{12} = 5.15$	$p = 0.0001$

Europe PMC Funders Author Manuscripts

Table 2

Statistical increase in performance and meta-performance with SOA for experiment 1 and 2.

	Experiment 1	Experiment 2
d'	$F_{3,36} = 8.776$, $p = 0.0002$	$F_{3,36} = 49.677$, $p < 10^{-4}$
meta-d′	$F_{3,36} = 8.12$, $p = 0.0003$	$F_{3,36} = 10.3$, $p < 10^{-4}$