

Decoding the Dynamics of Action, Intention, and Error Detection for Conscious and Subliminal Stimuli

Lucie Charles,^{1,2,3} Jean-Rémi King,^{1,2,3} and Stanislas Dehaene^{1,2,3,4}

¹Institut National de la Santé et de la Recherche Médicale, U992, Cognitive Neuroimaging Unit, NeuroSpin Center, Saclay Center, F-91191 Gif-sur-Yvette, France, ²Commissariat à l’Energie Atomique, Division of Life Sciences, Institute of BioImaging, NeuroSpin Center, F-91191 Gif-sur-Yvette, France, ³Université Paris-Sud, Cognitive Neuroimaging Unit, 91405 Orsay Cedex, and ⁴Collège de France, 75231 Paris Cedex 05, France

How do we detect our own errors, even before we receive any external feedback? One model hypothesizes that error detection results from the confrontation of two signals: a fast and unconscious motor code, based on a direct sensory–motor pathway; and a slower conscious intention code that computes the required response given the stimulus and task instructions. To test this theory and assess how the chain of cognitive processes leading to error detection is modulated by consciousness, we applied multivariate decoding methods to single-trial magnetoencephalography and electroencephalography data. Human participants performed a fast bimanual number comparison task on masked digits presented at threshold, such that about half of them remained unseen. By using both erroneous and correct trials, we designed orthogonal decoders for the actual response (left or right), the required response (left or right), and the response accuracy (correct or incorrect). While perceptual stimulus information and the actual response hand could be decoded on both conscious and non-conscious trials, the required response could only be decoded on conscious trials. Moreover, whether the current response was correct or incorrect could be decoded only when the target digits were conscious, at a time and with a certainty that varied with the amount of evidence in favor of the correct response. These results are in accordance with the proposed dual-route model of conscious versus nonconscious evidence accumulation, and suggest that explicit error detection is possible only when the brain computes a conscious representation of the desired response, distinct from the ongoing motor program.

Key words: Action; Consciousness; Error; Intention; MEEG

Introduction

Performance monitoring is a key function of the cognitive control system. When speed is emphasized over accuracy, we often commit a large numbers of errors that we are nonetheless able to correct and detect in a fast automatic manner (Rabbitt, 1966; Gehring et al., 1993). But how can the very same system that commits an error detect it? According to some models of cognitive control (Norman and Shallice, 1986), decision and motor

control are organized as a hierarchy in which higher-level conscious and intentional processes attempt to monitor performance (Posner and Rothbart, 1998), but sometimes arrives too late to modulate ongoing actions (Norman, 1981; Rabbitt, 2002).

In particular, the dual-route model for conscious and non-conscious decision making (Del Cul et al., 2009) hypothesizes that whenever we have to produce a motor response to some stimulus, two parallel routes (Fig. 1*B*) simultaneously accumulate evidence from the sensory input: a fast nonconscious sensory–motor route; and a slower but more accurate conscious route. Crucially, when instructions emphasize speed over accuracy, responses may frequently be emitted by the unconscious route, before the slower conscious route emits its more conservative judgment. Any discrepancy between the outputs of these two routes indicates that an error was committed, signaling a “mismatch” (Coles et al., 2001) or a conflict (Yeung et al., 2004) between actual and intended actions.

Here, we aimed at testing several predictions of this dual-route model. First, this model predicts that at the same time as the subject is making an erroneous action, for instance, clicking on the left-hand button, his or her brain should contain a distinct representation of the required correct action (i.e., clicking right). Second, this signal should be present only on conscious trials, coding for the conscious intention of the subject. Third, we should be able to predict the brain’s capacity for spontaneous

Received June 11, 2013; revised Nov. 27, 2013; accepted Nov. 28, 2013.

Author contributions: L.C. and S.D. designed research; L.C. performed research; L.C. and J.-R.K. contributed unpublished reagents/analytic tools; L.C. and J.-R.K. analyzed data; L.C. and S.D. wrote the paper.

This project was supported by PhD grants from the Direction Générale de l’Armement and the Fondation pour la Recherche Médicale, as well as a senior grant of the European Research Council to S.D. (NeuroConsc Program). The NeuroSpin magnetoencephalography facility was sponsored by grants from Institut National de la Santé et de la Recherche Médicale, Commissariat à l’Energie Atomique, the Fondation pour la Recherche Médicale, the Bettencourt-Schueller Foundation, and the Région Île-de-France. We are grateful to the NeuroSpin infrastructure groups, in particular to the doctors Ghislaine Dehaene-Lambert, Andreas Kleinschmidt, Caroline Huron, and Lucie Hertz-Pannier, and the nurses Véronique Joly-Testault and Laurence Laurier, for their support in subject recruitment and testing; Virginie van Wassenhove, Marco Buiatti, Leila Rogeau, Etienne Labyt, and the NeuroSpin magnetoencephalography team for their technical help; and Bertrand Thirion, Gaël Varoquaux, Alexandre Gramfort, and Fabian Pedregosa for their assistance with decoding methods.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no competing financial interests.

Correspondence should be addressed to Lucie Charles, INSERM-CEA Cognitive Neuroimaging Unit, CEA/SAC/DSV/DRM/NeuroSpin, Bâtiment 145, Point Courrier 156, F-91191 Gif/Yvette, France. E-mail: lucie.charles.ens@gmail.com.

DOI:10.1523/JNEUROSCI.2465-13.2014

Copyright © 2014 the authors 0270-6474/14/341158-13\$15.00/0

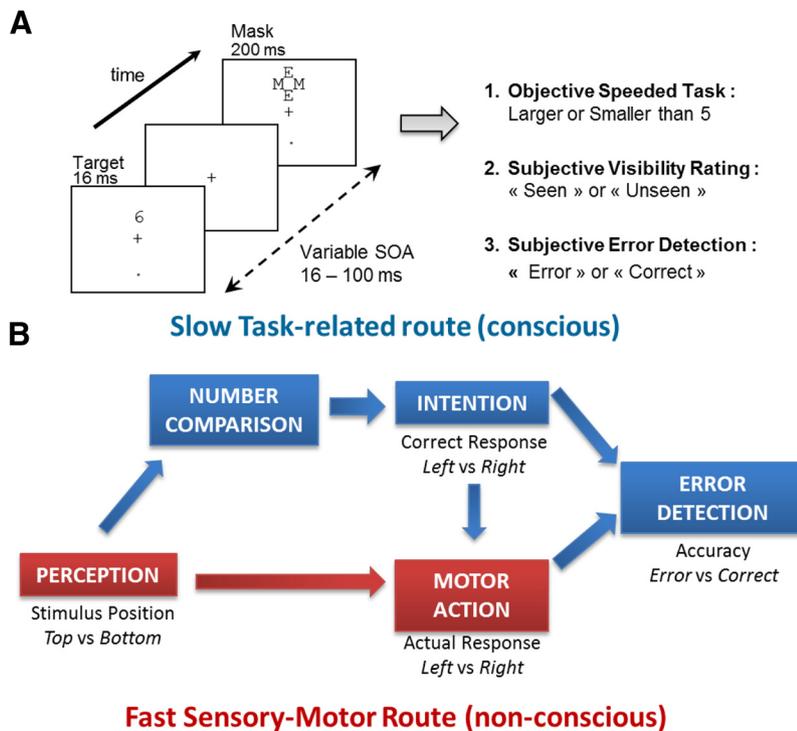


Figure 1. Experimental design and dual-route model. **A**, On each trial, a number was presented for 16 ms at one of two possible locations (top or bottom). It was followed by a mask composed of a fixed array of letters presented at a varying duration after target onset (16, 33, 50, 66, or 100 ms). Participants first performed a fast forced-choice number comparison task where they decided whether the number was smaller or larger than 5. Then, they evaluated the subjective visibility of the target and their own performance in the primary number comparison task. **B**, Dual-route model for error detection. In this model, two routes accumulate sensory evidence in parallel. A response is emitted by whichever route first reaches its decision threshold. The first route corresponds to automatic sensory–motor association and can be triggered nonconsciously to produce fast motor responses. The second route corresponds to the slower, voluntary processing of the stimulus according to task instructions and produces a conscious representation of the required response (i.e., a conscious intention). The comparison of the outputs of these two routes allows participants to detect a discrepancy between their intended and ongoing responses, and therefore to self-evaluate their performance.

error detection by the strength and timing of the discrepancy between the action and intention codes.

In the present study, we investigated this question by attempting to decode, from single-trial brain activity, the chain of cognitive processes linked to action monitoring and to determine how each stage was modulated by consciousness access. Decoding techniques, such as multivariate pattern analysis, have proven powerful in isolating precise cognitive processes in brain activity (Norman et al., 2006) and in distinguishing the processing sequence of a specific task (Bode and Haynes, 2009). Here, we applied these techniques to high-temporal resolution magnetoencephalography (MEG) and electroencephalography (EEG) recordings while participants performed a fast number comparison task on a masked digit and on each trial reported their subjective perception (seen/unseen) of the target (Charles et al., 2013). By training decoders to classify trials according to four different features (stimulus position, actual motor action, required motor action, and accuracy), separately on seen and unseen trials, we assessed how subjective visibility modulated perception, action, intention, and error detection. We then tested the prediction of the dual-route model that error detection results from the comparison of actions and intentions.

Materials and Methods

Participants

Thirteen volunteers (with normal or corrected-to-normal vision) were tested in this MEG/EEG experiment. Event-related potential and event-

related fields of these data have been partially reported previously (Charles et al., 2013). As the present within-subject decoding analysis necessitated a large number of error trials with both left and right motor responses, participants were excluded from the analysis if they did not have at least 20 error responses on both types of motor responses. Six participants (three male, three female) had sufficient numbers of trials in all of the conditions and were kept for analysis. This small number was compensated for by the fact that we systematically examined the within-subject significance of decoding scores, thus obtaining, for each question we raised, six within-subject replications as well as a between-subject nonparametric test.

Design and procedure

The paradigm of this experiment is described in detail by Del Cul et al. (2007) and Charles et al. (2013). Briefly, a target-stimulus (the digit 1, 4, 6, or 9) appeared on a white screen for 16 ms at one of two positions (top or bottom, 2.29° from fixation), with a pseudorandom 50% probability. After a variable delay, a mask appeared at the target location for 250 ms. The mask was composed of four letters (two Es and two Ms; Fig. 1A) 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. The following five SOAs were randomly intermixed: 16, 33, 50, 66, and 100 ms. 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 performed a fast forced-choice task of comparing the target number to the number 5. Responses were collected within 1000 ms after target onset with two buttons using the index finger of each hand (left button press = <5 response; right button-press = >5 response). To induce errors, participants were instructed to respond as fast as they could just after the appearance of the target. Time pressure was increased by presenting an unpleasant sound (mean pitch: 136.2 Hz; 215 ms duration) 1000 ms after target presentation whenever the response time exceeded 550 ms.

At the end of each trial, after another delay of 500 ms, participants were requested to provide two subjective answers with no time pressure. First, they had to indicate whether they saw the target number, or not (visibility task). Second, they had to report whether they thought they had made an error, or not, in the number comparison task (performance evaluation task). For both subjective responses, words corresponding to the two responses (seen/unseen and error/correct) were displayed on the screen, and subjects 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 subjects did not use an automatized button-press strategy.

The experiment was divided in blocks of 96 trials, with 16 trials by SOA condition in which each digit was presented at the two possible target locations (top/bottom). Each participant performed six or seven blocks during MEG/EEG recording. To achieve fast responses, participants were given a training session before the actual recording. They first received 5 min of training during which the target stimulus was not masked. Participants then performed three prerecording blocks of the actual experiment to check that overall performance was suitable for MEG/EEG recording.

Simultaneous EEG and MEG recordings

Simultaneous recording of MEG and EEG data was performed. The MEG system (Neuromag, Elekta) 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 the reference on the nose and the ground on the clavicle bone. Six additional electrodes were used to record electrocardiogram (ECG) and electro-oculogram (EOG; vertical and horizontal) signals.

A three-dimensional Fastrak digitizer (Polhemus) was used to digitize the position of three fiducial head landmarks (nasion and preauricular points) and four coils used as indicators of head position in the MEG helmet, for further alignment. The sampling rate was set at 1000 Hz with a hardware bandpass filter from 0.1 to 330 Hz.

MEG/EEG data preprocessing

MEG data were first processed with MaxFilter software using the signal space separation algorithm. Bad MEG channels were detected both automatically and manually, and were subsequently 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 applied to regress out the stereotypical physiological artifacts. First, artifacted time periods were detected on the EOG and ECG. Second, data were averaged on the onset of each blink and each heart beat separately, and PCA was performed separately for each type of sensor. Then, one to three of the first components characterizing the artifact were manually selected to be further removed.

Data were then entered into Matlab software and processed with Fieldtrip software (<http://fieldtrip.fcdonders.nl/>). An automatic rejection of trials based on signal discontinuities (all signals above 30 and 25 SDs in the 110–140 Hz frequency range) was performed. A low-pass filter at 30 Hz was then applied as well as a baseline correction from 300 to 200 ms before target stimulus onset.

An additional process was applied to the data used to decode stimulus position. Since the mask stimulus was presented at the same position as the target digit, we subtracted out the activity evoked by the mask to minimize the information provided by the mask location and decode only the information about the masked target. To do so, we first aligned each trial on the mask onset. We averaged separately the trials for which no target was presented, corresponding to the mask alone condition. We then subtracted from the rest of the data this mask-related activity and realigned those subtracted data on target onset (Del Cul et al., 2007). While the efficiency of this method is limited by the possible interaction between stimulus and mask presentation as well as trial-by-trial variability in mask-evoked activity, it allows for a large reduction in the response evoked by the mask. As stimulus position was not relevant for the other decoded categories, we did not apply this method to other decoding stages.

Decoding analysis

The support vector machine (SVM) method was used to decode different stages of perceptual decision, from stimulus encoding to performance detection. Briefly, linear classifiers such as SVM allow the discrimination, on a single-trial basis, of two conditions based on their pattern of activity across trials (Chang and Lin, 2011). This is achieved by finding a hyperplane separating the two classes of trials along the dimensions given to the decoder (e.g., sensors or time). We tested eight different decoders with binary SVM classifications for decoding several perceptual stages. Crucially, we split the initial dataset according to visibility reports and tested on each subset of trials how well the classifier could discriminate each of the four following conditions: stimulus position, top versus bottom; actual motor response, left versus right; correct motor response, left versus right; and accuracy, error versus correct.

Importantly, as several of these categories are based on responses of the participants, some of them had an unbalanced number of trials. For instance, more responses were made with the right hand than with the

left hand, probably because subjects who were mostly right handed were more prompt to answer with their dominant hand. Furthermore, the difference between categories could also be partially confounded with different numbers of trials in each subset. For instance, more errors were made when making a right-hand response, for similar reasons, therefore making more erroneous fast guesses with this response hand. These confounds present a problem for the decoding approach: for instance, when trying to decode error versus correct trials, we might obtain better than chance results simply because we are decoding left versus right responses.

To counteract these biases, we applied sample weights to equalize the weight of the trials entered into the classification and belonging to each cell of the potentially confounding category. As noted earlier, this required selecting participants who had >20 trials in each trial subcategory, therefore making sure that each subcategory was sufficiently populated. Sample weights were applied to each trial according to the number of trials in the subcategory to be controlled for. Each sample weight was computed using the following formula:

$$w_{\text{categ}} = n_{\text{tot}} / (4 * n_{\text{categ}}),$$

where w_{categ} designates the weight of all of the samples in the subject category, n_{tot} designates the total number of trials provided to the classifier, and n_{categ} designates the total number of trials in this subcategory. Note that total sum of weights across all trials was equal to n_{tot} , similar to the case where a weighting of 1 is applied to each trial. The subcategories of trials that we controlled for were the required motor response (left or right) for the position decoder and the actual motor response decoder, and the actual motor response (left or right) for the correct response and the accuracy decoders, respectively. Sample weights were entered as parameters in a linear kernel SVM implemented by SciKit-Learn toolbox (Pedregosa et al., 2011). For the decoding of the required response for which the results were the most crucial, we performed an additional step in the analysis to ensure that the decoder was not biased by the unbalanced number of trials among classes. In particular, as decoding the required response is identical to decoding the actual response in correct trials, and these trials were more numerous than error trials, the decoder could simply end up separating the trials according to the motor action (left vs right). To ensure that this was not the case, we separated the data according to the response hand and accuracy, and verified that within each subset the decoder performed above chance in classifying the trials according to the required motor response.

For each participant, MEG/EEG preprocessed data were entered into the classification pipeline (King et al., 2013). Importantly, we used two types of decoders, differing in the features that the decoder was trained on. In the first case, both time and space were used as decoding features, and the decoder was provided with the entire trial time window (0–800 ms after stimulus presentation). In this case, the decoder learned to decode in a high dimensional space with a total of $n_{\text{time point}} \times n_{\text{channel}}$ dimensions. In the second case, we trained a different decoder for each time point, using as a feature only the spatial dimension (n_{channel} dimensions). For the first type of decoder, we obtained only one classification measure for each trial. In the second case, we could reconstruct for each trial the entire time course of classification accuracy, allowing us to study more precisely the dynamics of the related cognitive process.

All decoding stages (including normalization of the MEG/EEG data) were fitted within the cross-validation loop on the training sets only. To obtain valid training/testing datasets, we used a stratified k -folding method, according to the number of trials in each subcategory (as described above). The data were split into seven folds, with each fold composed of a testing set of one of seven trials and a training set of six of seven trials, with the same proportion of trials coming from each subcategory.

To reduce the dimensionality of the data and improve the performance of the classifier, we first applied univariate feature selection (Haynes and Rees, 2006) on each training dataset. To do so, we used a simple ANOVA nested in the cross-validation loop, allowing only the most informative features to be kept. The number of features kept for the analysis was arbitrarily set to 50%. The remaining feature \times trial training data were then rescaled using the mean of z -score transformation. This method allows for all the EEG and MEG channels, which are recorded in

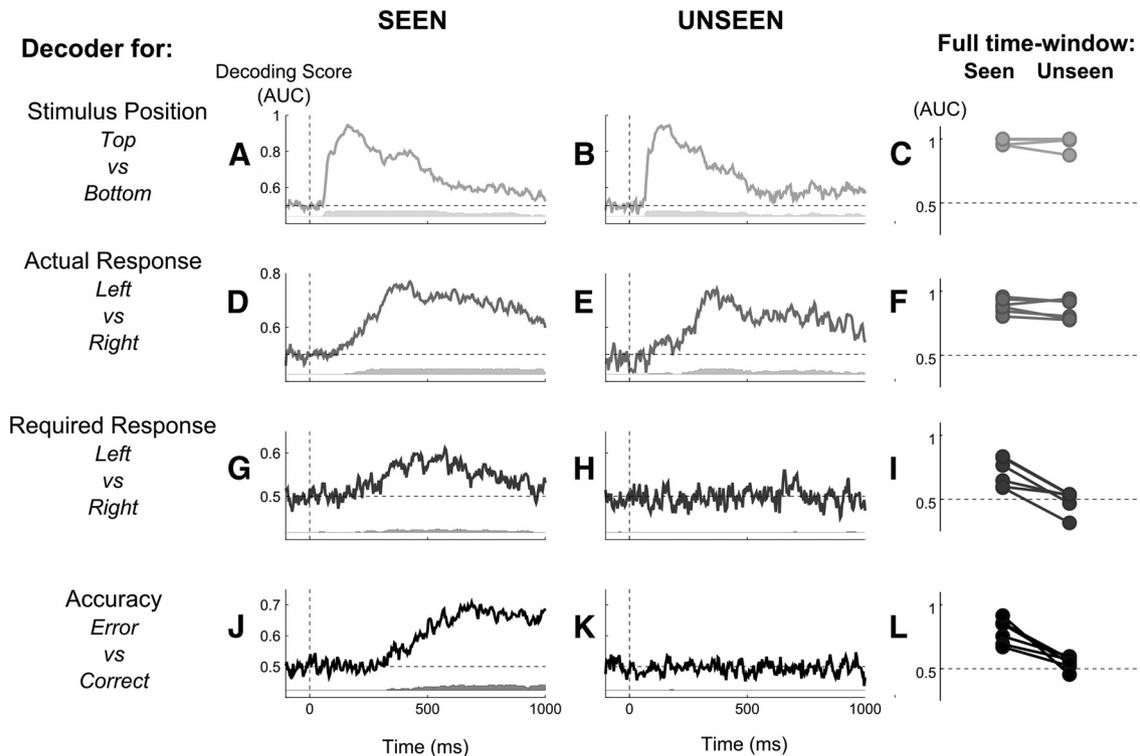


Figure 2. Decoding perception, action, intention, and accuracy, for conscious and nonconscious trials. Multivariate decoding was applied either to each time sample (central columns) or to the full trial time window (right columns). Results demonstrate that while stimulus position and actual response could be decoded in both conscious and nonconscious conditions with high accuracy, the required response and the accuracy could be decoded solely in conscious conditions. *A, B, D, E, G, H, J, K*, Central columns, AUC, a measure of decoding accuracy, is plotted after averaging across subjects, aligned on stimulus onset, separately for the stimulus position decoder (top vs bottom, *A, B*), actual response decoder (left vs right, *D, E*), required response decoder (left vs right, *G, H*), and accuracy decoder (error vs correct, *J, K*), respectively, in seen (*A, C, E, G*) and unseen (*B, D, F, H*) conditions. Gray bars below each graph indicate, for each time point, the number of subjects presenting an above-chance classification score at that instant as computed by cluster analysis. *C, F, I, L*, Right column, For all six subjects, change in classification scores (AUC) between seen (left points) and unseen (right points) conditions are plotted separately for the stimulus position decoder (*C*), actual response decoder (*F*), required response decoder (*I*), and accuracy decoder (*L*). In each case, decoding was applied on all the sensors and time points from the full trial time window (0–800 ms after stimulus presentation).

different physical units, to be put on the same scale and used properly by the classifier. The penalization parameter of the algorithm was then estimated by means of a grid search by nested cross-validation applied within each training dataset (two stratified k -folds), and the best hyperplane was retrieved. Finally, we fitted a cumulative probability distribution function on the decision function of the training dataset using the method of Platt (1999), allowing us to obtain for each trial, not just a discrete output label, but a continuous value bounded by 0 and 1, representing the classifier's estimate of the probability to belong to the first class. Then, exactly the same feature selection and scaling parameters obtained from the training dataset were applied to the testing dataset, and the obtained classifier was applied to the test trials, allowing us to obtain a cross-validated classification measure for each of the test trials. We ensured that we applied all multivariate classification guidelines outlined in the study of Lemm et al. (2011) to minimize classification bias and avoid circular analyses that could result in overfitting the data.

Statistical analysis

Within-subject classification scores. Classification scores across trials were estimated for each subject with a receiver operating characteristic (ROC) curve analysis applied to the obtained classification probabilities and were summarized by the area under the curve (AUC) values. The ROC curve presents the true-positive rate (the proportion of trials belonging to class A and classified as A; i.e., hits) as a function of the false-positive rate (the proportion of trials belonging to class B and classified as A; i.e., false alarms), providing a measure of both the sensitivity and specificity of the decoder. A diagonal ROC curve, which coincides with an AUC of 50%, corresponds to a situation where the number of hits and false alarms are equal, showing a chance level classification score. On the contrary, an AUC of 100%, which corresponds to an ROC curve on the left upper bound of the diagonal,

indicates a perfect positive prediction with no false positives and a perfect decoding score. Importantly, and unlike average accuracy, AUC analysis provides an unbiased measure of decoding accuracy, robust to imbalanced problems and independent of the statistical distribution of the classes.

The classification AUCs were estimated for each subject for the decoders on the entire trial duration (Fig. 2, right column), and above-chance significance within and across subjects was computed by means of a nonparametric Wilcoxon rank sum test. Separately, the AUC was computed for the obtained decoding time series separately for each time point and was averaged across subjects. The middle columns in Figure 2 show the AUC time series averaged across subjects for each decoder.

Within-subject time cluster analysis on decoder time series. To determine the moments at which the decoders performed above chance, we computed within-subject statistics on the obtained trial time series. Using individual data, for each decoder we used a cluster-based nonparametric with Monte Carlo randomization (adapted from Maris and Oostenveld, 2007) on the trial-by-trial time series of decoding probabilities. This method allowed us to identify clusters of time points in which time series of the two learned classes present a significant difference while correcting for multiple comparisons. For each time sample, p values of the difference between the two decoded classes were first computed by means of a nonparametric Mann–Whitney U test. Clusters were then identified by taking all dyads of time samples adjacent in time with $p < 0.05$. The final significance of the cluster was determined by computing the sum of AUC values of the entire cluster and comparing them with the results of Monte Carlo permutations (2000 permutations). Clusters were considered significant at corrected $p < 0.05$ if the probability computed with the Monte Carlo method was inferior to 5% (one-tailed test). The number of sub-

jects presenting a significant cluster at each time point is shown in Figure 2 at the bottom of each graph.

Regression analysis on single-trial amplitude. The dual-route model of error detection predicts that, on each trial, evidence on the required response and the actual response are compared to determine the accuracy of the action. In other words, for a given trial, the amount of evidence that an error was made depends on the discrepancy between action and intention.

To evaluate this prediction, we used the actual and the required response entire-trial decoders as indices of the amount of internal information available on each trial about the action and the intention. This analysis was performed separately on seen and unseen trials. As the chance level was not identical across subjects, we normalized for each subject the trial-by-trial classification probabilities. We then transformed the obtained signal so that it would be centered on 0 and fluctuate between 1 and -1 (instead of 0 and 1; see Fig. 5A). This can be achieved by subtracting on a trial-by-trial basis the decoded probability of belonging to one of the two classes from the probability of belonging to the opposite class (see Fig. 5A): as the sum of the probability is equal to 1, when the probability of belonging to one class is close to 1, the subtraction will be close to either 1 or -1 , while it will be close to 0 when the probability is at chance. We then computed the product of the two obtained indices, because this measure gave us a trial-by-trial index of the discrepancy between action and intention. We then retrieved, for each trial, the output of the accuracy decoder corresponding to the decoded trial-by-trial probability of an erroneous motor response, and correlated it to our index of congruity between action and intention.

We correlated for each subject the trial-by-trial indices obtained by multiplying the motor and intention decoder with the accuracy probability. We used robust linear regression (Holland and Welsch, 1977) to reduce the effect of points with aberrant classification scores (this method did not qualitatively affect our results). A nonparametric test was then performed on the slope of the regression obtained across subjects. As negative values of the computed product should signal erroneous responses, we expected a negative correlation between the two measures, with the smallest negative value being associated with the highest probability of decoding an error.

Regression analysis on single-trial timing. Another prediction of the dual-route model is that one should be able to determine the accuracy of its own response only when information is available on both the required response and the response actually made. This prediction implies that the latest obtained information on either the actual response or the correct response should determine the moment at which an internal estimate of response accuracy can be emitted.

To test this prediction, we searched for a correlation between the time at which the accuracy decoder crossed a threshold and the moment when the latest of the action and intention decoders crossed their threshold. For each subject separately, we normalized the classification probability time series according to the baseline (-100 to 0 before stimulus presentation) to obtain values centered on 0 and ranging from -1 to 1. To increase the signal-to-noise ratio, we converted the SVM probability time series into a cumulative-sum time series (see Fig. 7A), and we extracted for each decoder the moment at which each time series reached 50% of its mean final value across trials (see Fig. 7B). Trials that did not reach the threshold for any of the three decoders were excluded from the analysis, resulting in the selection of about half of the trials for which intention, action, and accuracy could be decoded with high performance. We then took the maximal value of the decision times for the actual response and the required response decoders, and correlated it with the crossing of the threshold of the accuracy decoder. We then performed a nonparametric Wilcoxon signed rank test on the β values across subjects.

Results

Analysis of behavioral data

We first investigated how behavior varied with conscious perception. We split the data according to the trial-by-trial report of visibility. Accuracy in the number comparison task was higher in

seen than in unseen trials ($t_{(5)} = -4.47$, $p < 0.01$). While errors were significantly faster than correct trials on seen trials ($F_{(1,4)} = 18.4$, $p = 0.012$), no such effect was observed on unseen trials. As previously found (Charles et al., 2013), subjective visibility increased in a nonlinear manner with SOA ($F_{(4,20)} = 33.7$, $p < 10^{-4}$), with unseen responses associated mainly with short SOA duration while seen responses dominated for longer SOA values. However, no further effect of SOA was found on reaction times (RTs) or accuracy after splitting the data by visibility. Therefore, in the decoding analysis, trials were split according to subjective visibility alone, regardless of SOA.

Decoding stages of stimulus processing

To determine how consciousness influenced the processing chain leading from stimulus perception to the response and its evaluation, we separated different stages in a decision hierarchy, and we tested whether and when an experimental variable attached to each processing stage could be decoded from the single-trial brain activity, separately for conscious and nonconscious trials. Figure 2 depicts, for each decoder, the individual classification score (AUC; see Materials and Methods) over the entire trial window and the time course of the classification score averaged across subjects.

Decoding early visual processes: stimulus position classifier

Figure 2C shows the result of the classification of stimuli position over the entire trial duration, for both seen and unseen trials. This analysis revealed that stimulus position could be decoded for each individual subject on both seen and unseen trials, with high accuracy. Nonparametric statistics showed that the decoder performed significantly above chance for each subject (Wilcoxon rank sum test on classification probabilities, all $p < 10^{-4}$). The AUC was significantly higher than chance across subjects for both types of trials (Wilcoxon rank sum test, $AUC > 0.5$, $n = 6$, both $p < 0.05$).

Considering the results of the decoding on each time point allowed us to determine precisely the dynamics of perceptual processing of the stimulus, in seen and in unseen conditions (Fig. 2A, B). The peak of performance of the decoder was observed ~ 175 ms after stimulus presentation for seen trials and 130 ms for unseen trials. Within-subject statistical analysis revealed that, for both seen and unseen trials, subjects presented a significant cluster starting ~ 75 ms after onset of the stimulus, which lasted for at least 450 ms. Interestingly, for five of the six subjects the time window of significance lasted longer for conscious than for nonconscious trials.

We then performed a nonparametric test to determine whether the overall difference between the two decoded classes was greater for seen compared with unseen trials. Nonparametric tests across subjects revealed no statistical significance between the two ($p = 0.47$), suggesting that the performance in decoding stimulus position over the entire trial duration were not different in seen compared with unseen trials.

These results suggest that it is possible to classify the stimulus position with high accuracy both for seen compared with unseen trials, showing that early visual processing of the stimulus is largely unimpaired in nonconscious conditions.

Decoding motor decision: actual response decoder

We then turned to the motor response decoder. The aim of this analysis was to determine whether it was possible for a decoder to learn which motor decision was made by the subject. According to our design, a left-hand action implies that the subject's re-

sponse to the stimulus was <5 while a right-hand action corresponded to a response of >5 .

Figure 2D–F shows that the decoder performed significantly above chance to determine whether a left or a right motor response was produced on each trial, both for seen and unseen trials. Again, analysis of the AUCs obtained from the decoding of the motor response over the entire time window revealed that for each subject we were able to decode the motor response better than chance both for seen and for unseen trials (Wilcoxon rank sum test, all $p < 10^{-4}$; Fig. 2F). Similarly, the AUC was significantly higher than chance across subjects both in conscious and nonconscious conditions (Wilcoxon rank sum test, $n = 6$, both $p < 0.05$), confirming that the motor response could be decoded with high accuracy in both cases. Interestingly, comparison between seen and unseen trials revealed no statistical difference between the two, suggesting a comparable decoding accuracy of the motor response in conscious and nonconscious conditions.

The time course of the decoding of the motor response (Fig. 2D,E) revealed that decoding accuracy increased linearly from ~ 120 ms after stimulus presentation both for seen and unseen trials. For five of six subjects, the earliest significant difference between left and right responses was observed at 240 ms after stimulus presentation. Decoding accuracy reached a plateau around the average time of the actual key press (365 and 366 ms, respectively, for seen and unseen trials). The maximal peak was observed at ~ 425 ms for seen trials and at 365 ms for unseen trials, slightly later than the mean RT across subjects. In summary, this analysis revealed that the actual motor response could be decoded with very high accuracy both in conscious and nonconscious conditions.

Required response decoder

One of the main goals of this study was to test whether it is possible to decode, from the time course of brain activity, the presence of a higher-order representation of the required response. We predicted that, on top of the representation of the actual ongoing motor program, there might be a distinct representation of the intended response. On the majority of trials where the response is correct, the intended and actual responses coincide. However, whenever subjects commit an error, the dual-route model predicts that their brain contains a distinct representation of the response that would have been correct. Thus, this neural code should encode the response that should have been made by the subjects, independently of the response that they actually make.

To test this idea, we trained a decoder to classify trials according to the required response, regardless of the actual motor response on the same trial. Importantly, to teach the decoder the proper class, we weighted equally the erroneous and correct trials. Since errors were overall much less frequent than correct trials, we used a weighting technique that ensured that both errors and correct trials were equally used in training the decoder (see Materials and Methods), thus removing the correlation between intended and actual responses.

On seen trials, decoding over the entire time window revealed that we were able to decode the required response for each subject (Wilcoxon rank sum test, all $p < 0.005$; Fig. 2I). Analysis across subjects revealed that the average AUC was significantly above chance (Wilcoxon rank sum test, $n = 6$, $p < 0.05$). However, for unseen trials, we were not able to decode the required response. Analysis of the decoding results showed that the classifier performed at chance for all subjects (Wilcoxon rank sum test, all $p > 0.35$), except for one subject for which the classifier performed

significantly below chance (Wilcoxon rank sum test, $p < 10^{-4}$; Fig. 2I). Similarly, the average decoding score across subjects did not differ from chance (Wilcoxon rank sum test, $n = 6$, $p = 0.35$). This resulted in a significant effect of visibility on the decoding scores across subject (Wilcoxon rank sum test, $n = 6$, $p < 10^{-3}$).

When training the decoder on each time sample (Fig. 2G,H), within-subject statistical analysis revealed a significant cluster for all subjects in the seen condition (Fig. 2G). Three subjects presented an identical significant temporal cluster between 350 and 750 ms after stimulus presentation, while the remaining subjects presented a shorter period of significance in this time window. Interestingly, decoding performance varied in time across subjects, with some subjects presenting above-chance decoding accuracy only starting on average at 500 ms after stimulus presentation. No such decoding was possible for unseen trials (Fig. 2H). Cluster-level significance was not achieved for most of the subjects. For one subject, a 10 ms time window of significance was found, unlikely to reflect a solid effect (subject 6, 700–710 ms after stimulus onset). For another subject, a more sustained cluster was found, but at a time that is unlikely to be meaningful (subject 5, 925–960 ms after stimulus onset). Therefore, these results suggest that a representation of the required response can be decoded from brain activity in seen trials, but that not enough information is available on unseen trials for the classifier to extract this representation.

As the decoding of the required response was performed on highly unbalanced datasets, where correct trial were more numerous than error trials, we verified that the decoder on seen trials was not simply picking up the motor activity on correct trials. Thus, we separated the trials according to the actual motor response (left vs right) and the actual performance (error vs correct). Crucially, we verified that, within each such subset, the decoder performed above chance in classifying the trials according to the required motor response (Fig. 3). If the intention decoder simply used information related to the ongoing motor action, its results should be at chance when testing its ability to decode the required response for a fixed actual motor action, and it should be significantly below chance when testing error trials only. Therefore, these two analyses ensured that the present decoder was indeed decoding information related to the required response, regardless of the motor response and the accuracy (Fig. 3).

We first tested separately the trials with a fixed response hand. When considering the decoding of the required response over the entire time window, average classification scores across subjects were above chance for both right and left motor response (Wilcoxon rank sum test, $AUC > 0.5$, $n = 6$, $p = 0.016$ and $p = 0.03$, respectively). When considering decoding time series, the decoder for the required response performed above chance for five of six subjects within right-hand motor responses, while within left-hand motor responses for which fewer trials were available, significance was achieved in two subjects. For both motor responses, time windows of significance overlapped on a 410–650 ms time period after stimulus onset. We then tested whether above-chance decoding could be observed for error trials specifically. Average decoding scores across subjects did not significantly exceed chance when considering the entire time window. However, when considering decoding time series, decoding performance reached significance in three subjects, within a time window of 430–530 ms. For two of the remaining three subjects, a trend was seen in the appropriate direction (i.e., responses tended to be classified according to the required response and therefore opposite to the actual response).

In conclusion, while this analysis was limited by the small number of trials available within each subcategory of trials, it confirmed that the intention decoder learned to classify trials according to the required response and not the actual motor response made by the subject.

Accuracy decoder

We then determined whether our recordings contained decodable single-trial information about the accuracy of the motor decision, separately for seen and unseen trials. The dual-route model postulates that to determine the accuracy of their decisions, participants compare their actual motor response to the response that they should have made, and evaluate the discrepancy between these two internal representations. As we were not able to decode the representation of the required response on unseen trials, the model predicted that we should also not be able to decode accuracy on these trials. That is indeed what we found. Considering the entire time window, we were able to decode with high performance the accuracy of the response at a trial-by-trial level for all six subjects on seen trials (Wilcoxon rank sum test, all $p < 10^{-4}$), resulting in an above-chance classification score across subjects (Wilcoxon rank sum test, $n = 6$, $p < 0.05$). Importantly, we were not able to decode the accuracy of the motor response on unseen trials except for one subject (Wilcoxon rank sum test, all $p = 0.03$), resulting in a chance-level decoding score across subjects (Wilcoxon rank sum test, $n = 6$, all $p = 0.08$). This resulted in a significant effect of visibility on the decoding scores across subjects (Wilcoxon rank sum test, $n = 6$, $p < 10^{-3}$).

Considering the decoding analysis for each time sample, the peak of decoding performance on seen trials was reached at the latest time window rather than for previous action and intention decoders, ~ 600 ms after stimulus presentation. On unseen trials, decoding scores remained at chance over the entire time window.

Following the dual-route prediction model, our results therefore suggest that the brain encodes a representation of response accuracy that can be decoded with high accuracy on conscious trials, but that on nonconscious trials, when no information is available on the required response, the accuracy of the motor response cannot be predicted.

Effect of visibility on early versus late processing stages and effect of SOA

Our results suggest that only the early stages of processing of the stimulus, containing either visual or motor activity, can be decoded equally well on conscious and nonconscious trials, while higher-order representations of the goal of the action and its accuracy are available only in conscious conditions. To support this, we performed an ANOVA separately, with visibility and decoder type (perceptual and motor vs intention and accuracy) as main factors. A significant interaction ($F_{(1,29)} = 21.07$; $p = 10^{-4}$) revealed that, indeed, while early stages could be decoded with equal performance in conscious and nonconscious conditions, late stages could be decoded only in conscious trials.

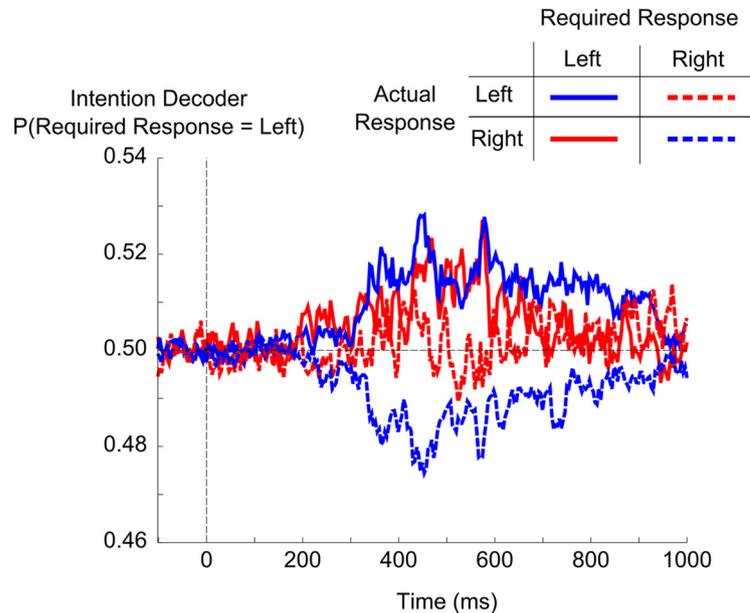


Figure 3. Decoding conscious intention independently of motor action. This figure demonstrates that, while subjects are preparing for a given response (correct or erroneous) their brain activity contains decodable information about the response that they should make (the required response). The graph shows the output of the intention decoder (i.e., the estimated probability of the required response being left, averaged across subjects). The decoder was trained on all seen trials. Trials were then sorted according to the actual response and the required response. Errors are plotted in red, and correct trials in blue. Time 0 corresponds to the onset of the stimulus.

The above seen/unseen comparison is partially confounded with differences in objective conditions of stimulation, as the majority of seen trials comes from trials with long SOAs, while unseen trials correspond in majority to short SOAs (Charles et al., 2013). We performed an additional analysis to ensure that visibility and SOA were disentangled. To do so, we split the decoding results according to SOA and computed decoding scores separately for each subset of data. Due to the unbalanced number of trials for each SOA, this analysis could not be performed for each SOA conditions for all subjects. However, for the intermediate SOA of 33 ms, a sufficient number of trials were available both for seen and unseen trials (Fig. 4) for four subjects. On these trials, the pattern of results was unchanged: decoding of the stimulus position and the actual response could be performed on both conscious and nonconscious trials for each SOA conditions (Fig. 4), but decoding of the required response and the accuracy could be performed only on conscious trials. Indeed, an ANOVA with visibility and decoder type (perceptual and motor vs intention and accuracy) as main factors for this intermediate SOA value revealed a significant interaction ($F_{(1,19)} = 6.653$; $p < 0.05$; $n = 4$), suggesting that subjective visibility, above objective variations in stimulation, influenced the decoding of late decision stages.

Trial-by-trial test of predictions of the dual-route model

Congruity between action and intention correlates with the strength of error detection

The dual-route model states that if no representation of the required response is available, as seems to be the case in the unseen condition, then the accuracy of one's performance cannot be determined. A related prediction is that trial-by-trial variation in the amount of evidence, concerning either the required response or the actual response, should be predictive of the amount of evidence concerning decision accuracy. In particular, the more evidence one has on what the required response is, the better one can determine whether one's performance is correct or not.

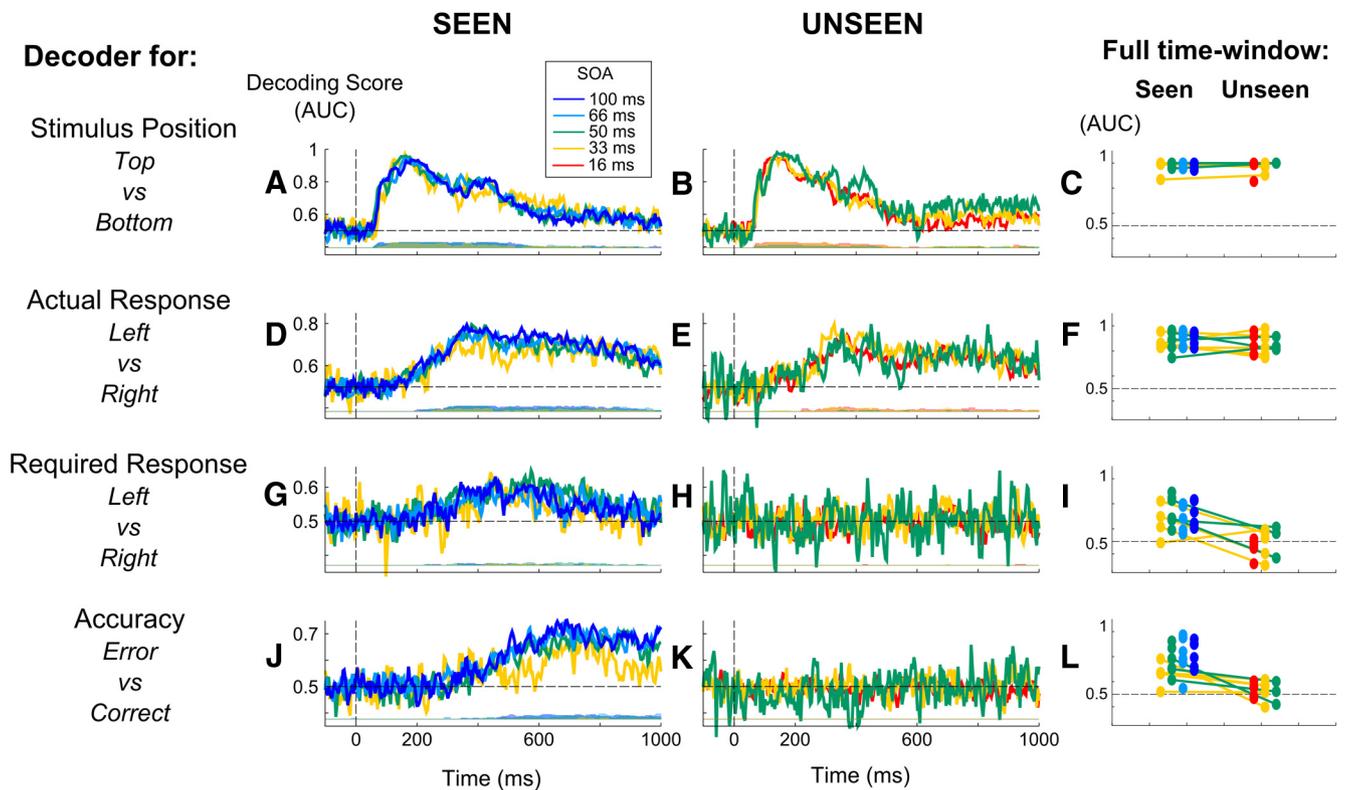


Figure 4. Decoding perception, action, intention, and accuracy for conscious and nonconscious trials according to SOA. Multivariate decoding was applied either to each time sample (central columns) or to the full trial time window (right columns), and results were spited by SOA condition. Results demonstrate that the required response and the accuracy could be decoded in conscious conditions for each SOA condition. **A, B, D, E, G, H, J, K**, Central columns, AUC, a measure of decoding accuracy, is plotted for each SOA condition after averaging across subjects, aligned on stimulus onset, separately for the stimulus position decoder (top vs bottom, **A, B**), actual response decoder (left vs right, **D, E**), required response decoder (left vs right, **G, H**), and accuracy decoder (error vs correct, **J, K**). Due to reduced trial numbers, only the shortest SOAs (16, 33, and 50 ms) are presented for unseen trials (**B, E, H, K**), while only longer SOAs (33, 50, 66, and 100 ms) are included for seen trials (**A, D, G, J**). Bars below each graph indicate, for each time point, the number of subjects presenting an above-chance classification score at that instant as computed by cluster analysis. **C, F, I, L**, Right column, For each subject and each SOA condition, individual measures of AUC are plotted for seen (left) and unseen (right) trials, separately for the stimulus position decoder (**C**), actual response decoder (**F**), required response decoder (**I**), and accuracy decoder (**L**). In each case, decoding was applied on all the sensors and time points from the full trial time window (0–800 ms after stimulus presentation), and results were then split according to SOA condition.

To test this prediction, we collected the trial-by-trial classification probabilities computed by the three main decoders separately for seen and unseen trials and used them as indices of the amount of evidence available for the required response, the actual motor response, and the accuracy (see Materials and Methods). Our main goal was to determine whether for each trial, the discrepancy between action and intention predicted the decoding of accuracy.

We computed for each trial an intention index and a motor index varying from -1 to 1 across trials and coding for the amount of information that this trial contained, respectively, about the intended response and the motor response (Fig. 5B; -1 corresponds to a sure left response and 1 to a sure right response). According to the dual-route model, the product of the intention and action indices, which evaluates their congruency, should predict the accuracy of the response. If the product is positive, it means that the action and the intention are congruent, and the actual motor response is therefore likely to be correct. On the contrary, if the product is negative, it means that intention and action vote in favor of different responses, and the actual motor response is likely to be incorrect. Note that if one of the indices is close to 0 (i.e., no information is available either on the action or on the intention), the product is also close to 0 , so the model predicts that the accuracy of the response cannot be predicted. Across trials, accuracy evidence should therefore be correlated with the product of action and intention indices.

After transforming the classification probability of the actual motor response and the required response into signed indices of action and intention strength (Fig. 5A), we computed for each trial the product of these two indices, obtaining a measure, for each trial, of the congruity between intention and action (see Materials and Methods). We then retrieved from the accuracy decoder the estimated probability that the response was erroneous on the same trial. Finally, we tested whether these two measures were correlated, as predicted by the dual-route model. As the obtained indices were signed values, we expected a negative correlation between the two measures: a negative product indicated a discrepancy between action and intention, and therefore a greater probability of error (Fig. 5B).

Figure 5C depicts this correlation for each subject for seen trials. Linear regression was performed for each subject, and we tested whether the slope differed from 0 . All regression slopes were negative (Fig. 5C), and the Wilcoxon rank sum test on the slopes across subjects confirmed that the average slope was significantly different from 0 ($n = 6$, $p = 0.016$). These findings indicated that, indeed, trial-by-trial fluctuations in the congruity between intention and action signals correlated with fluctuations in the strength of error representation in the participants' brain, as predicted by the dual-route model. Unsurprisingly, such correlation did not reach significance in unseen trials for which no intention signal could be decoded.

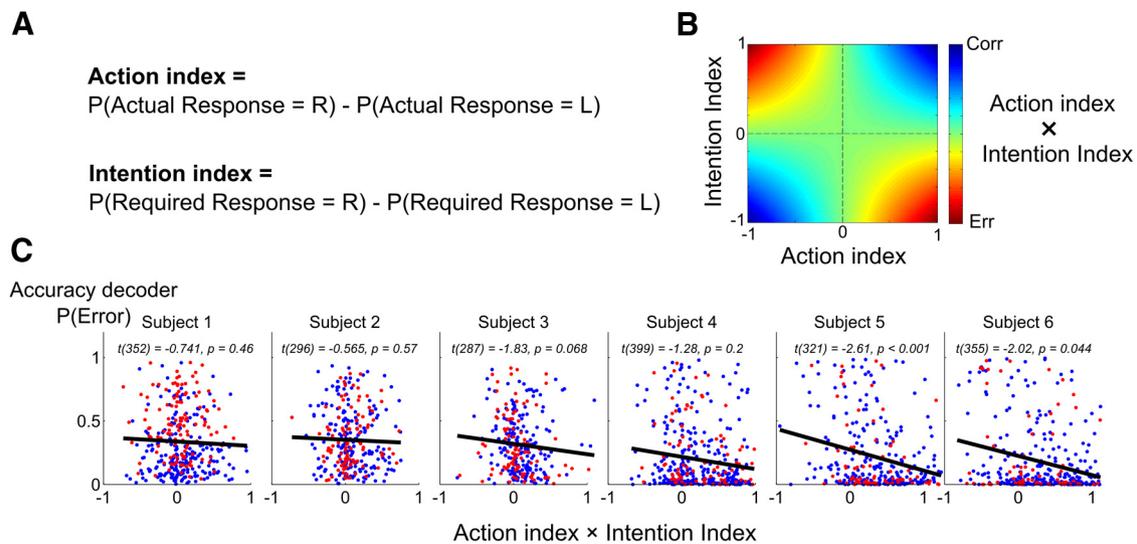


Figure 5. Congruity between action and intention correlates with the strength of error decoding. **A**, To obtain a trial-by-trial measure of the strength of internal representations of action and intention, we first transformed the output of the classifiers by subtracting the classification probability of the left response from the classification probability of the right response, thus yielding for each trial a measure ranging from -1 (i.e., certainty of a left response) to 1 (certainty of a right response). This computation was done separately for the actual response and for the required response, thus yielding two single-trial indices of the strength of internal representations, the action index, and the intention index. **B**, The product of the intention and action indices reflects the congruity between intended and executed actions. Positive values (blue) are obtained when both action and intention are congruent (the values of the two indices are of the same sign), indicating a high probability of being correct. On the opposite side, negative values (red) indicate a discrepancy between action and intention, and therefore a high probability of committing an error. Note that when no information is available on either the action or the intention, the product is close to 0 and does not allow distinguishing error from correct trials. **C**, Correlation results of the product of action and intention indices with the decoded error probability for each subject. Each dot corresponds to a single seen trial (red = errors, blue = correct). A negative correlation confirms that the internal representation of an upcoming error is stronger when the discrepancy between internal representations of action and intention is larger.

The timing of error detection correlates with the slowest of the internal codes for action and intention

Another prediction of the dual-route model is that the timing of error detection should be predictable from the timing of the computation of the action and intention codes. To investigate this question more precisely, we realigned the obtained decoding time series on the onset of the response, to gain a clearer view of how the dynamics of error detection varied with the timing of the response. Figure 6 depicts the time courses of the classification scores realigned on the onset of the motor response in seen trials. Above-chance decoding of the motor response (Fig. 6A) and the required response (Fig. 6B) occurred before the onset of the actual key press. Classification performance was significantly better than chance in the time window of -150 to -100 ms before the key press for the actual response decoder (Wilcoxon rank sum test, $n = 6, p < 0.05$) and -100 to -50 ms for the required response decoder (Wilcoxon rank sum test, $n = 6, p < 0.05$). Crucially, decoding of the accuracy was possible immediately after this point, in the time window just preceding the motor response (-50 to 0 ms before key-press; Wilcoxon rank sum test, $n = 6, p < 0.05$; Fig. 6C), suggesting that error detection followed the computation of the actual response. Interestingly, decodability of the accuracy continued to increase after this point, in the time window of the error-related negativity (ERN) and the following positive deflection, the Pe. The highest classification scores were obtained at the time of the Pe, 250 ms after the motor response. Indeed, as the Pe corresponds to a long-lasting component while the ERN corresponds to a sharper peak after the motor response, it is not surprising that decoding aligned on the stimulus revealed higher decoding scores at the time of the Pe when data were realigned a posteriori on the motor response.

Nonetheless, these results suggest that error detection immediately follows the erroneous motor action. However, when

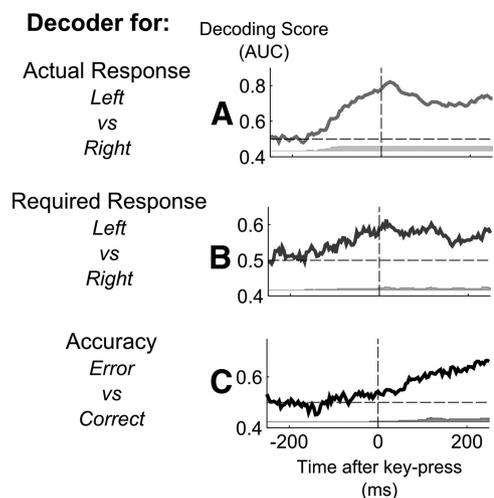


Figure 6. Decoding action, intention, and accuracy before and after the actual key press. **A–C**, For seen trials only, the figure shows the time course of decoding the actual response (**A**), the required response (**B**), and the accuracy (**C**) relative to actual key-press. The curves were realigned on motor onset and an average measure of decoding success (AUC) was computed across subjects. Gray bars below graphs indicate for each time point the number of subjects presenting an above-chance classification score at that instant, as computed by cluster analysis.

speed pressure is imposed, the dual-route model predicts that a motor response may be emitted early on, before a clear intention has been computed from the stimulus. In this case, error detection should be possible only once the intention is determined. Overall, the timing of error detection should vary on a trial-by-trial basis according to the availability of both intention and action signals, whichever comes last.

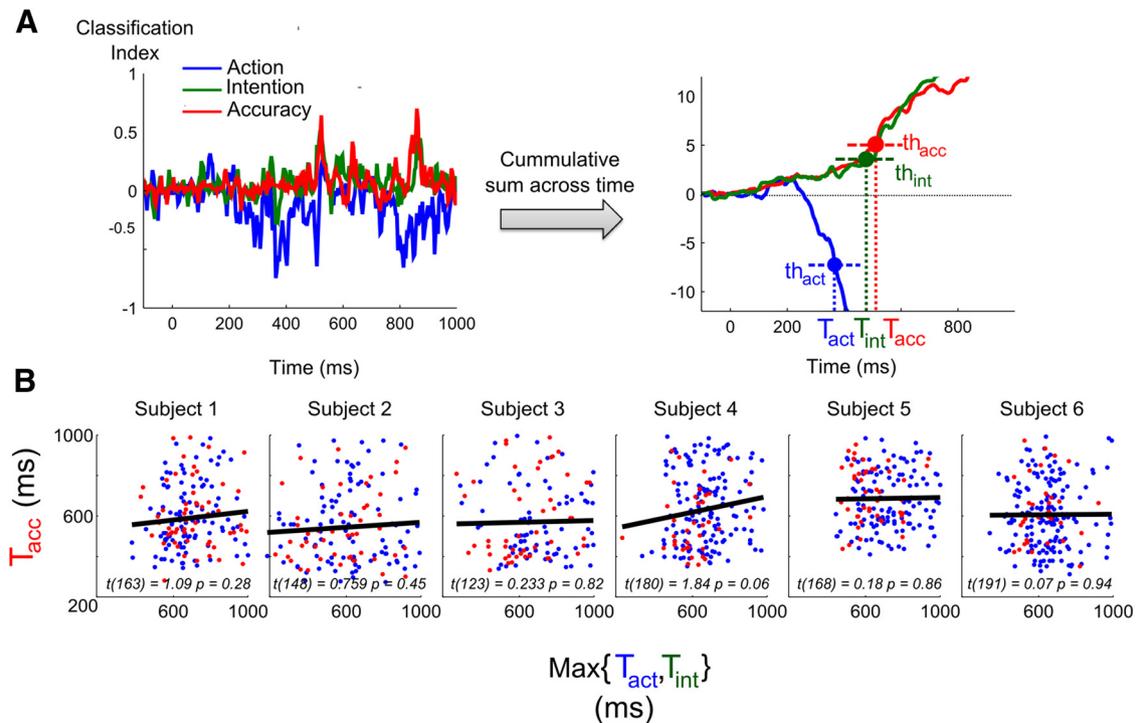


Figure 7. The timing of error detection correlates with the slowest of two signals for action and intention. **A**, Example of a single-trial computation of decoding time. To improve the signal-to-noise ratio, we computed the cumulative sum, across time, of the probability values obtained from each of the three decoders for actual response, required response, and accuracy. Threshold values for each decoder were defined, and the timing of threshold crossing for each time series was taken as an index of the time when this code first became available on this trial. Thus, three values were obtained for each trial (T_{act} , T_{int} , and T_{acc}), corresponding respectively to the time for threshold crossing of the actual response, the required response, and the accuracy decoder. **B**, Correlation results of the slowest (maximum) time index between T_{int} and T_{act} with the time index of error detection T_{acc} . Each dot corresponds to a single seen trial (red = errors, blue = correct). A positive correlation indicates that, as predicted, error information becomes available only once both action and intention codes have been computed.

To test this prediction, we computed for each trial the moment at which each of the three decoders (action, intention, and accuracy) crossed a given threshold (see Materials and Methods). We therefore obtained for each trial three time measure indices (T_{int} , T_{act} , and T_{acc}), corresponding respectively to the timing of intention, action, and accuracy detection (Fig. 7A). We then tested how these times correlated with one another.

We first verified whether T_{act} correlated with the actual trial-by-trial RT. This was indeed the case; the slope of a linear regression was significantly >0 across subjects (Wilcoxon rank sum test, $p = 0.016$).

As only the latest event between action and intention should determine when one can detect making an error, we then computed, for each trial, the maximum value between T_{int} and T_{act} and correlated it with T_{acc} , as shown in Figure 7A. None of the regressions reached significance at the single-subject level (all $p > 0.05$). However, a nonparametric test on the slope of the regression across subjects revealed a significant positive correlation (Wilcoxon rank sum test, $p = 0.016$; Fig. 7B), suggesting a correlation between the timing of performance detection and the timing of action and intention, as predicted by the dual-route model.

Discussion

We showed that MEG/EEG signals contain decodable information on the correct motor response, independently of the ongoing motor plan. Such information was present only on seen trials and not on unseen trials, while lower-level perceptual information and motor action were decodable on both types of trials. These findings suggest that, when the stimulus

is masked below the threshold for conscious access, the brain is unable to compute a clear representation of the required action for that stimulus given the task instructions. Furthermore, the accuracy of the motor decision was also decodable from conscious trials only, with a magnitude and at a point in time correlating with the information that is decodable about the actual and the required action. These results fit with the prediction of the dual-route model of error detection, according to which accuracy can be determined, on conscious trials only, by comparing the output of two distinct cortical routes for conscious and nonconscious processes, which compute intention and action, respectively.

The crucial finding of this study is that, for conscious trials, a representation of the required response can be decoded in brain activity, independently of the ongoing motor action. This finding builds upon our previous work (Charles et al., 2013) where we showed that when performing a task on masked stimuli, the ERN, a known brain marker of error detection is present only on a conscious trial. In the present study, we replicated this finding using multivariate analysis, showing that the accuracy of motor decisions can be decoded only in conscious conditions. Crucially, we now show the presence, in brain activity, of an intention signal, which is modulated in exactly the same fashion by subjective visibility, and might serve as an input to error detection and the triggering of the ERN.

The presence of an accurate intention signal independent from the action itself, but contemporaneous with it, readily explains how errors can be detected and corrected, sometimes

nearly instantaneously after the wrong key-press (Rabbitt, 2002), or why the ERN starts nearly simultaneously with the erroneous response itself (Rodríguez-Fornells et al., 2002). Indeed, the existence of such a signal had been postulated in several previous models that proposed that error detection results from a comparison (Bernstein et al., 1995; Coles et al., 2001; Maier et al., 2008) or a conflict (Van Veen and Carter, 2002b; Yeung et al., 2004) between the executed and the required response.

A dissociation between intention and action was previously reported by Desmurget et al. (2009) and Desmurget and Sirigu (2012), who found that, during intracranial stimulation of the right inferior parietal region, subjects reported a strong intention to move, without any actual electromyographic activity. Similarly, decisions can be decoded from the activity of prefrontal cortex before any motor preparation (Haynes et al., 2007). Our finding provides further evidence of a brain representation distinct from the ongoing motor plan that nonetheless carries information about the intended action. Our decoding method, operating on sensor-level data, did not allow us to investigate directly which brain regions carried this intention signal. However, previous findings suggest that premotor cortex (Gallivan et al., 2011), precuneus (Soon et al., 2013), medial prefrontal cortex (Haynes et al., 2007), and parietal cortex (Desmurget et al., 2009) could be plausible candidates for decoding intentions. In the present experiment, since decoding the required response coincided with decoding the result of the number comparison task, regions involved in number processing (Dehaene et al., 2003) might also be involved. Further research, dissociating these factors, will be needed to specify the precise source of the intention signal that serves as a basis for error detection.

Which computational models may explain how the same system produces an initial error and its subsequent correction? According to some models of decision making, a single decision system accounts for both the initial incorrect response and the subsequent corrective action (Kiani and Shadlen, 2009; Resulaj et al., 2009; Pleskac and Busemeyer, 2010), which corresponds to a late “change of mind” (Resulaj et al., 2009) signaling the commission of an error. This single-representation model is challenged by our findings, which demonstrate the simultaneous presence of two orthogonal patterns of brain activity coding respectively for the ongoing and the required response, and suggest that neural codes for the desired response and the executed motor response are not activated in parallel. In connectionist models of conflict, two decision unit responses compete to produce the motor action errors resulting from the initial activation of the incorrect decision unit, while activity in the correct decision unit builds up more slowly. The overlap of these two activations triggers a conflict signal reflected by ERN (Botvinick et al., 2001; Yeung et al., 2004). Similarly, in models of error detection as a comparison, an “efference copy” of the motor response is kept in memory while further processing of the stimulus leads to the computation of the correct response (Falkenstein et al., 2000; Coles et al., 2001), the mismatch between these two representations leading to an error detection signal. While these two models predict that a representation of the required response should be decodable from both correct and error trials, such information should be available only at later stages, concomitant with the moment when accuracy first becomes decodable. Our findings, which indicate a parallel activation of the correct and incorrect responses, do not fit with these predictions.

An alternative model that fits with the observed data is a dual-route model for conscious versus nonconscious processes (Del Cul et al., 2009) in which intentions emerge from the computation of a slower but more accurate route for conscious evidence accumulation. We previously suggested that the dual-route model could account for our observation of an all-or-none error detection reflected by the ERN and triggered only in conscious trials (Charles et al., 2013). Indeed, the ERN and its following positive component, the Pe (Falkenstein et al., 2000), as well as patterns of brain activity originating from cingulate cortex (Debener et al., 2005; Charles et al., 2013) could have been at the origin of the signals used by the present performance accuracy decoder. Another related prediction of the model is that the size of the discrepancy (Scheffers and Coles, 2000) or conflict (Steinhauser and Yeung, 2010) between intended and executed action should predict the size of the internal error signal. Indeed, we found that on conscious trials the trial-by-trial product of action and intention decoding scores correlated with the decoded probability of accuracy. Likewise, several studies found that the ERN and the Pe vary with the objective amount of evidence in favor of the correct response (Hughes and Yeung, 2011; Steinhauser and Yeung, 2012; Charles et al., 2013) as well as the subjective identification of the required response (Scheffers and Coles, 2000; O’Connell et al., 2007; Dhar et al., 2011; Hughes and Yeung, 2011; Wessel et al., 2011; Shalgi and Deouell, 2012). Furthermore, we found that the time at which this accuracy code emerged correlated with the slowest of the action and intention codes, in accordance with the prediction that the latency of error detection should reflect the latest of the two available signals for action and intention (Van Veen and Carter, 2002b; Yeung et al., 2004). More detailed investigations will be needed to determine precisely how our results relate to findings regarding the determinants of the amplitude and timing of the ERN and the Pe. In particular, our data need to be reconciled with findings obtained when aligning evoked responses with corrective responses (Burle et al., 2008). However, overall, our conclusions are in accordance with models that view these components as essential steps in the error detection process (Steinhauser and Yeung, 2010; Wessel et al., 2011; Wessel, 2012).

The present study also sheds light on the distinction between subliminal and conscious processing. We found a dissociation between early and late stages of stimulus processing, consistent with the findings that automatized perceptual, cognitive, and motor operations are preserved even under subliminal conditions (Del Cul et al., 2006; Melloni et al., 2007), while later stages show an all-or-none dissociation between conscious and nonconscious trials (Sergent and Dehaene, 2004; Del Cul et al., 2007). Nonetheless, cognitive processes related to performance monitoring may also be partially triggered nonconsciously (Nieuwenhuis et al., 2001; Cohen et al., 2009; Logan and Crump, 2010). Indeed, in our previous analysis of the present dataset, we found that subjects could detect the accuracy of motor decisions with above-chance accuracy even on unseen trials (Charles et al., 2013), suggesting that some performance evaluation processes distinct from the ERN operate nonconsciously. According to the dual-route model, the level of evidence reached by the nonconscious route is a noisy indicator of the confidence in the response (Galvin et al., 2003; Pleskac and Busemeyer, 2010), and thus may be used as a subliminal index of accuracy. Crucially, however, this mechanism is only statistical in nature, and thus unable to confi-

dently and categorically label a given trial as correct as erroneous. Our results suggest that such categorical meta-cognitive knowledge cannot be attained unconsciously, but requires an explicit representation of the required action.

References

- Bernstein PS, Scheffers MK, Coles MG (1995) “Where did I go wrong?” A psychophysiological analysis of error detection. *J Exp Psychol Hum Percept Perform* 21:1312–1322. [CrossRef Medline](#)
- Bode S, Haynes JD (2009) Decoding sequential stages of task preparation in the human brain. *Neuroimage* 45:606–613. [CrossRef Medline](#)
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108:624–652. [CrossRef Medline](#)
- Burle B, Roger C, Allain S, Vidal F, Hasbroucq T (2008) Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *J Cogn Neurosci* 20:1637–1655. [CrossRef Medline](#)
- Chang C, Lin C (2011) LIBSVM: a library for support vector machines. *ACM TIST* 2:article 27. [CrossRef](#)
- Charles L, Van Opstal F, Marti S, Dehaene S (2013) Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 73:80–94. [CrossRef Medline](#)
- Cohen MX, van Gaal S, Ridderinkhof KR, Lamme VA (2009) Unconscious errors enhance prefrontal-occipital oscillatory synchrony. *Front Hum Neurosci* 3:54. [CrossRef Medline](#)
- Coles MG, Scheffers MK, Holroyd CB (2001) Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biol Psychol* 56:173–189. [CrossRef Medline](#)
- Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *J Neurosci* 25:11730–11737. [CrossRef Medline](#)
- Dehaene S, Piazza M, Pinel P, Cohen L (2003) Three parietal circuits for number processing. *Cogn Neuropsychol* 20:487–506. [CrossRef Medline](#)
- Del Cul A, Dehaene S, Leboyer M (2006) Preserved subliminal processing and impaired conscious access in schizophrenia. *Arch Gen Psychiatry* 63:1313–1323. [CrossRef Medline](#)
- Del Cul A, Baillet S, Dehaene S (2007) Brain dynamics underlying the non-linear threshold for access to consciousness. *PLoS Biol* 5:e260. [CrossRef Medline](#)
- Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132:2531–2540. [CrossRef Medline](#)
- Desmurget M, Sirigu A (2012) Conscious motor intention emerges in the inferior parietal lobule. *Curr Opin Neurobiol* 22:1004–1011. [CrossRef Medline](#)
- Desmurget M, Reilly KT, Richard N, Szathmari A, Mottolese C, Sirigu A (2009) Movement intention after parietal cortex stimulation in humans. *Science* 324:811–813. [CrossRef Medline](#)
- Dhar M, Wiersema JR, Pourtois G (2011) Cascade of neural events leading from error commission to subsequent awareness revealed using EEG source imaging. *PLoS One* 6:e19578. [CrossRef Medline](#)
- Falkenstein M, Hoormann J, Christ S, Hohnsbein J (2000) ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol* 51:87–107. [CrossRef Medline](#)
- Gallivan JP, McLean DA, Valyear KF, Pettypiece CE, Culham JC (2011) Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *J Neurosci* 31:9599–9610. [CrossRef Medline](#)
- Galvin SJ, Podd JV, Drga V, Whitmore J (2003) Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon Bull Rev* 10:843–876. [CrossRef Medline](#)
- Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. *Psychol Sci* 4:385–390. [CrossRef](#)
- Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534. [CrossRef Medline](#)
- Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. *Curr Biol* 17:323–328. [CrossRef Medline](#)
- Holland PW, Welsch RE (1977) Robust regression using iteratively re-weighted least-squares. *Commun Stat Theory Methods* 6:813–827. [CrossRef](#)
- Hughes G, Yeung N (2011) Dissociable correlates of response conflict and error awareness in error-related brain activity. *Neuropsychologia* 49:405–415. [CrossRef Medline](#)
- Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759–764. [CrossRef Medline](#)
- King JR, Fugeras F, Gramfort A, Schurger A, El Karoui I, Sitt JD, Rohaut B, Wacongne C, Labyt E, Bekinschtein T, Cohen L, Naccache L, Dehaene S (2013) Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* 83:726–738. [CrossRef Medline](#)
- Lemm S, Blankertz B, Dickhaus T, Müller KR (2011) Introduction to machine learning for brain imaging. *Neuroimage* 56:387–399. [CrossRef Medline](#)
- Logan GD, Crump MJ (2010) Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science* 330:683–686. [CrossRef Medline](#)
- Maier ME, Steinhauser M, Hübner R (2008) Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. *J Cogn Neurosci* 20:2263–2273. [CrossRef Medline](#)
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190. [CrossRef Medline](#)
- Melloni L, Molina C, Pena M, Torres D, Singer W, Rodriguez E (2007) Synchronization of neural activity across cortical areas correlates with conscious perception. *J Neurosci* 27:2858–2865. [CrossRef Medline](#)
- Steinhauser M, Yeung N (2012) Error awareness as evidence accumulation: effects of speed-accuracy trade-off on error signaling. *Front Hum Neurosci* 6:1–12. [CrossRef Medline](#)
- Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GP, Kok A (2001) Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology* 38:752–760. [CrossRef Medline](#)
- Norman DA (1981) Categorization of action slips. *Psychol Rev* 88:1–15. [CrossRef](#)
- Norman D, Shallice T (1986) Attention to action: willed and automatic control of behavior. In: *Consciousness of self-regulation. Advances in research and theory*, Vol 4 (Davidson RJ, Schwartz GE, Shapiro D, eds), pp 1–18. New York: Plenum.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430. [CrossRef Medline](#)
- O’Connell RG, Dockree PM, Bellgrove MA, Kelly SP, Hester R, Garavan H, Robertson IH, Foxe JJ (2007) The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. *Eur J Neurosci* 25:2571–2579. [CrossRef Medline](#)
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. *J Machine Learn Res* 12:2825–2830.
- Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: *Advances in Large-Margin Classifiers (Neural Information Processing)* (Smola AJ, Bartlett P, Schölkopf B, Schuurmans D, eds.), pp 61–74.
- Pleskac TJ, Busemeyer JR (2010) Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev* 117:864–901. [CrossRef Medline](#)
- Posner MI, Rothbart MK (1998) Attention, self-regulation and consciousness. *Philos Trans R Soc Lond B Biol Sci* 353:1915–1927. [CrossRef Medline](#)
- Rabbitt PM (1966) Error correction time without external error signals. *Nature* 212:438. [CrossRef Medline](#)
- Rabbitt P (2002) Consciousness is slower than you think. *Q J Exp Psychol A* 55:1081–1092. [CrossRef Medline](#)
- Resulaj A, Kiani R, Wolpert DM, Shadlen MN (2009) Changes of mind in decision-making. *Nature* 461:263–266. [CrossRef Medline](#)
- Rodriguez-Fornells A, Kurzbuch AR, Münte TF (2002) Time course of error detection and correction in humans: neurophysiological evidence. *J Neurosci* 22:9990–9996. [Medline](#)

- Scheffers MK, Coles MG (2000) Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *J Exp Psychol Hum Percept Perform* 26:141–151. [CrossRef Medline](#)
- Sergent C, Dehaene S (2004) Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. *J Physiol Paris* 98:374–384. [CrossRef Medline](#)
- Shalgi S, Deouell LY (2012) Is any awareness necessary for an Ne? *Front Hum Neurosci* 6:124. [CrossRef Medline](#)
- Soon CS, He AH, Bode S, Haynes JD (2013) Predicting free choices for abstract intentions. *Proc Natl Acad Sci U S A* 110:6217–6222. [CrossRef Medline](#)
- Steinhauser M, Yeung N (2010) Decision processes in human performance monitoring. *J Neurosci* 30:15643–15653. [CrossRef Medline](#)
- Van Veen V, Carter CS (2002a) The timing of action-monitoring processes in the anterior cingulate cortex. *J Cogn Neurosci* 14:593–602. [CrossRef Medline](#)
- Van Veen, Carter CS (2002b) The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiol Behav* 77:477–482. [CrossRef Medline](#)
- Wessel JR (2012) Error awareness and the error-related negativity: evaluating the first decade of evidence. *Front Hum Neurosci* 6:88. [CrossRef Medline](#)
- Wessel JR, Danielmeier C, Ullsperger M (2011) Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. *J Cogn Neurosci* 23:3021–3036. [CrossRef Medline](#)
- Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev* 111:931–959. [CrossRef Medline](#)