



## Preserved metacognition for undetected visuomotor deviations

Michael Pereira, Rafal Skiba, Yann Cojan, Patrik Vuilleumier, Indrit Bègue

### ► To cite this version:

Michael Pereira, Rafal Skiba, Yann Cojan, Patrik Vuilleumier, Indrit Bègue. Preserved metacognition for undetected visuomotor deviations. *Journal of Neuroscience*, 2023, 43 (35), pp.6176-6184. 10.1523/JNEUROSCI.0133-23.2023 . hal-04199348

**HAL Id: hal-04199348**

**<https://cnrs.hal.science/hal-04199348>**

Submitted on 7 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# ***Preserved metacognition for undetected visuomotor deviations***

**Abbreviated title:** Preserved metacognition for undetected deviations

Michael Pereira <sup>1</sup>, Rafal Skiba <sup>2,3,4</sup>, Yann Cojan <sup>2</sup>, Patrik Vuilleumier <sup>2</sup>, Indrit Bègue <sup>2, 5, 6\*</sup>

1. Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

2. Laboratory for neurology and imaging of cognition, Department of Basic Neuroscience, University of Geneva, 1211 Geneva, Switzerland

3. BC Mental Health and Addictions Research Institute, Vancouver, BC, Canada

4. Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

5. Adult psychiatry division, Department of psychiatry, University hospitals of Geneva, 1211 Geneva, Switzerland

6. Laboratory for clinical and experimental psychopathology, Department of psychiatry, University of Geneva, 1211 Geneva, Switzerland

**\* Corresponding author:**

Email: [indrit.begue@unige.ch](mailto:indrit.begue@unige.ch)

Number of pages: 52

Number of Figures: 3

Number of Tables: 2

Number of words for abstract: 150

Number of words for introduction: 567

Number of words for discussion: 1652

The authors declare no conflict of interest

**Acknowledgements:** This study was conducted on the imaging platform at the Brain and Behavior Lab (BBL), at the University of Geneva, Switzerland and benefited from support of the BBL technical staff. Special thanks to Christophe Mermoud for programming of the experimental protocol. We thank Nathan Faivre for constructive comments on the manuscript.

## 34 Abstract

36 Humans can successfully correct deviations of movements without conscious detection of such  
38 deviations, suggesting limited awareness of movement details. We ask whether such limited  
awareness impairs confidence (metacognition). We recorded functional magnetic resonance  
40 imaging data while 31 human female and male participants detected cursor deviations during a  
visuomotor reaching task and rated their confidence retrospectively. We show that participants  
42 monitor a summary statistic of the unfolding visual feedback (the peak cursor error) to detect  
visuomotor deviations and adjust their confidence ratings, even when they report being unaware of  
44 a deviation. Crucially, confidence ratings were as metacognitively efficient for aware and unaware  
deviations. At the neural level, activity in the ventral striatum tracks high confidence, whereas a  
46 broad network encodes cursor error but not confidence. These findings challenge the notion of  
limited conscious action monitoring and uncover how humans monitor their movements as they  
unfold, even when unaware of ongoing deviations.

## 48 Statement of relevance

We are unaware of the small corrections we apply to our movements as long as our goals are  
50 achieved. Here, although we replicate the finding that participants deny perceiving small deviations  
that they correct, we show that their confidence reliably reflects the presence or absence of a  
52 deviation. This observation shows that they can metacognitively monitor the presence of a  
deviation, even when they deny perceiving it. We also describe the hemodynamic correlates of  
54 confidence ratings. Our study questions the extent to which humans are unaware of the details of  
their movements and describes a plausible mechanism for metacognition in a visuomotor task,  
56 along with its neural correlates and has important implications for the construction of the sense of  
self.

## 58 Introduction

Whether reaching for popcorn while viewing a movie or biking while enjoying the scenery, humans  
60 rely on reciprocal intricate connections between vision and motor processing to perform efficient  
behavior. Such visuomotor loops seem to occur mostly in the absence of awareness. Indeed,  
62 seminal work from (Fournieret & Jeannerod, 1998) showed that participants are unaware of their  
true hand position under imposed visuomotor deviations, although they appropriately correct their  
64 movement trajectories. Humans also neglect small spatial incongruences in feedback about their  
own movements (Farrer et al., 2008) and can reach targets that they cannot consciously report  
66 (Binsted et al., 2007) or that are displaced without being noticed (Goodale et al., 1986). These  
findings support the notion that participants show limited *awareness* of the details of an ongoing  
68 movement, as long as their goal is achieved (Blakemore et al., 2002; Custers & Aarts, 2010;  
Gaveau et al., 2014; Tsay et al., 2020).

70  
Rather than awareness of movement details, other studies tested whether humans can monitor the  
72 accuracy of decisions based on movement details (Sinanaj et al., 2015; Arbuzova et al., 2020;  
Charles et al., 2020; Locke et al., 2020). This ability to monitor (and control) one's internal  
74 processes is referred to as *metacognition* (Flavell, 1979; Koriat, 2006). A standard measure of  
metacognition in humans is to ask them to make a decision (about a stimulus or an action) and  
76 subsequently rate their confidence in this decision (Rahnev et al., 2020). Together, these two lines  
of research suggest that, although humans have limited *awareness* of their movement details, they  
78 can *metacognitively monitor* decisions about these movements' details. As the discrimination tasks  
used previously (Arbuzova et al., 2020; Charles et al., 2020; Locke et al., 2020) were not  
80 specifically designed to contrast trials with and without awareness, it remains unknown whether  
awareness is necessary for metacognitive monitoring during motor control.

82  
Indeed, during perceptual decision-making, participants are better at using perceptual information  
84 to scale their confidence about the presence of a stimulus than about the absence of a stimulus,  
showing that metacognitive *efficiency* is impaired when perceptual information is insufficient to  
86 reach awareness (Kanai et al., 2010; Meuwese et al., 2014; Mazor et al., 2020). It remains  
unresolved whether awareness has a similar effect on metacognitive efficiency in visuomotor  
88 decisions. In a previous study on metacognition of movement with a visuomotor *detection task*,  
participants reported being aware of only some of the experimental deviations in the visual

90 feedback of their movements. Individual differences in metacognitive efficiency were associated  
with variability in brain gray matter volume in prefrontal cortex, insular, and visual areas (Sinanaj et  
92 al., 2015). However, the degree to which metacognitive efficiency is related to awareness of  
visuomotor deviation and to specific movement parameters, as well as their functional neural  
94 underpinning, were not examined in our first study.

Here, we sought to explain how confidence is computed in the presence or absence of reported  
96 awareness – measured using introspective subjective reports (Mudrik & Deouell, 2022) – of a  
visuomotor deviation and unveil brain regions mediating this process using functional magnetic  
98 resonance imaging (fMRI) data. We could thus determine how visual information about  
participants' movements was integrated into their confidence ratings, while dissociating trials for  
100 which they reported being aware of the deviation from those for which they reported not being  
aware of it. Finally, we identified brain regions whose activity correlated with confidence, detection  
102 response, and monitoring of performance during a visuomotor task.

## 104 Method

### Participants

106 We recruited 32 healthy right-handed participants based on convenience sampling. One participant did not  
complete the experimental task, therefore, the final sample included 31 participants of either sex (age: 26  
108 years  $\pm$  4.7). Participants gave written informed consent prior to the experiment and received 20 Swiss  
francs per hour as compensation. They had normal or corrected-to-normal vision and reported no  
110 neurological or psychiatric disorder. The study was approved by the Ethics Committee of the University of  
Geneva and University Hospitals of Geneva (CER:11-214/NAC 11-077). All participants read and signed an  
112 informed consent form, and were screened for contraindications to MRI with a standard safety questionnaire.

### Experimental procedure

114 We asked participants to perform a visuomotor reaching task (Fournier & Jeannerod, 1998; Farrer et al.,  
2008). After a short preparation period (white triangle became red, duration 1-2 seconds, jittered) participants  
116 had to reach a centrally-located target on the screen with a cursor moving at constant speed in a direction  
controlled by the joystick (Figure 1A). The experimental manipulation (79% of the trials) consisted in  
118 introducing deviations in the mapping between the joystick and the cursor direction: the cursor direction was  
deviated by a certain deviation angle, positive for right (clockwise) deviations and negative for left  
120 (anticlockwise) deviations. These deviations were gradually applied (0.3 s ramp), starting after participants  
reached a fixed distance from the starting point corresponding to 13% of the vertical distance between the  
122 initial cursor error and the final target. Participants were informed that these externally originating deviations  
would not occur all the time, however, when they occurred, participants had to correct for these deviations  
124 using the joystick in order to reach the target. After reaching the target, participants reported whether they  
noticed any externally originating deviations of their trajectory ("yes" detection responses) or whether they  
126 did not ("no" responses), and subsequently rated their confidence in their own judgment on a scale ranging  
from 1 = not certain to 5 = completely certain (Figure 1A). Participants did not receive feedback about the  
128 accuracy of either detection, or confidence judgments. We encouraged participants to use the whole  
confidence scale. Participants selected "yes" or "no" responses as well as confidence ratings through joystick  
130 handle movements on the right and left, respectively, then pressing a button on the joystick.

132 We asked participants to perform a training run outside the scanner, consisting of 30 non-deviated  
trajectories to familiarize them with the joystick and experimental environment. For each experimental  
134 session, we ran an adaptive staircase procedure that made the task more difficult after two consecutive  
correct responses by decreasing the next deviation by 2.64°, but made it easier after an incorrect response  
136 by increasing the next deviation by 1°. After the training session, participants entered the scanner and  
performed a 'threshold' session of 80 trials in order to stabilize the staircase procedure. After the threshold

session, participants completed two experimental runs. A structural T1 image was acquired between these two runs. Overall, there were 208 trials (21% without experimental deviations). Each trial lasted 11.5 s and was followed by a blank screen with a jittered duration (3 to 6 s).

## Behavioral analyses

We defined a trial with an experimental deviation that was reported as such by participants as a *hit* and as a *miss* in case it was not reported. A trial was a *correct rejection* when there was no deviation and participants correctly reported no deviation and a *false alarm* if participants reported a deviation. We grouped hits and false alarms into “yes” responses and misses and correct rejections into “no” responses. We computed the sensitivity  $d'$  and criterion  $c$  using signal detection theory. Cursor and joystick positions were defined every 10 ms. *Cursor error* was defined as the horizontal distance between the cursor position and the midline between the starting point (triangle at a lower central position on the screen) and the target (top central position). For each trial, we defined *peak cursor error* as the peak value of this cursor error over time. *Onset error* (onset err.) was defined as the cursor error at the onset of the deviation. We also defined the *average position* (avg. pos.) as the absolute value of the average position of the cursor with respect to the sagittal line (the latter can be negative) as well as the *average cursor error* (avg. err.) as the average of the distance between the cursor and midline (always positive). A trial with a large deviation to the right followed by a large deviation to the left would thus have an average position close to zero but a high average cursor error.

To build two-dimensional histograms of the cursor trajectories (Figure 1B), we mirrored trajectories when the deviation was towards the right (only for hits and misses). We then computed two-dimensional histograms of the position of the cursor in all trials of a signal detection theory category (hit, miss, correct rejection and false alarms) and normalized the resulting histograms by the number of trials in that category. Finally, we averaged across participants. Since the cursor at time  $k$  ( $c_k$  in the complex plane) can be defined knowing cursor position at time  $k-1$  and the angle of the joystick handle at time  $k$  ( $\alpha_k$ )

$$c_k = c_{k-1} + p \exp(j(\alpha_k + \delta))$$

with  $c_0 = 0 + j*0$ ,  $p$  a small real constant defining the speed and  $\delta$  the angle of the experimental deviation, we could reconstruct the trajectories of the cursor, had there not been any deviation and construct histograms as described above for Figure 1B:

$$\hat{c}_k = \widehat{c_{k-1}} + p \exp(j(\alpha_k))$$

For statistics, we defined Binomial link mixed-effects models to analyze detection responses and cumulative link mixed-effects models for confidence ratings using the ordinal package (Christensen, 2019) in R. Inclusion of random effects was guided by regression model selection based on Bayesian information criterion and led to the inclusion of all factors and interactions as random effects. Since most models are



172 multivariate, we ensured that variance inflation factors that measure how much the variance of one  
coefficient is increased because of collinearity were under 3, indicated weak collinearity that does not  
174 warrant any correction (Belsey et al., 1980). All statistical tests were two-tailed. For Figure 1E and 1F, we  
performed similar analyses but including a factor either for cursor or for joystick position. As these factors are  
176 defined every 10 ms time point, we fitted one mixed-effect model per time point and corrected p-values for  
multiple comparisons using the false-discovery rate. To assess the amount of information in the visual  
178 feedback (e.g. the visual trajectory displayed as subjects performed the visuomotor reaching task), we  
performed receiving operator characteristics analyses computing the true- and false-positive rate for a range  
180 of criterion values, from 0 to its maximum by steps of 0.01. We then searched for the criterion leading to the  
same false-positive (false-alarm) rate as found in the data and compared the corresponding true-positive rate  
182 (hit rate) to the one observed in the data.

184 For Figure 2D, we binned the peak cursor error into five quantiles computed independently for each  
participant but for all signal detection theory categories together. For Figure 2E, we plotted model predictions  
186 of the cumulative link model's fixed effects for each signal detection theory category and increasing values of  
peak cursor error. To estimate metacognitive efficiency, or to what extent the information available to the  
188 detection decision was used to scale their confidence, we used the M-ratio between meta-d' (Maniscalco &  
Lau, 2014) and d' estimated in a response-specific version of the HMeta-d' toolbox (Fleming, 2017). These  
190 meta-d' and d' allow to compute the M-Ratio: a ratio between meta-d' and d', which quantifies metacognitive  
efficiency, or how well information from first-order performance informs the metacognitive process. We used  
192 the default parameters of three chains of 10'000 samples with 1000 *burn in* samples for the MCMC  
procedures with no thinning. Visual inspection of MCMC and Rhat values well under 1.1 indicated good  
194 convergence. Of note, as the d' does not vary according to the response, differences in the M-Ratio are  
solely driven by differences in the meta-d'.

196

### **fMRI data collection, preprocessing and analyses**

198 We acquired functional MRI images with a 3T whole-body scanner (Trio TIM, Siemens, Germany) with a 12-  
channel head-coil. Functional images were acquired with a susceptibility weighted EPI sequence with the  
200 following parameters: TR/TE = 2100/30 ms, flip angle = 80 degrees, PAT factor = 2, 64 × 64 voxel, 3.2 × 3.2  
mm, 36 slices, 3.2 mm slice thickness, 20% slice gap. We acquired structural images using a T1-weighted  
202 3D sequence using the following parameters: MPRAGE, TR/TI/TE = 1900/900/2.32 ms, flip angle = 9°, voxel  
dimensions: 0.9 mm isotropic, 256 × 256 × 192 voxels. We presented task stimuli on a back-projection screen  
204 inside the scanner bore using an LCD projector (CP-SX1350, Hitachi, Japan). We recorded responses via  
buttons placed on the joystick used for the visuomotor reaching task (HH-JOY-4, Current Designs Inc., USA

206

We used the SPM8 software (<http://www.fil.ion.ucl.ac.uk/spm/>) for statistical analyses of functional data with a standard pipeline. We first corrected for head movements between scans by an affine registration and realignment to the mean of all images. The anatomical image was spatially normalized on the T1 template. The functional images were also normalized to the EPI template, which were thereby transformed into standard stereotaxic space and resampled with a 3×3×3mm voxel size. The normalized images were spatially smoothed using an 8mm full-width at half-maximum (FWHM) Gaussian kernel. We used the general linear model (GLM) framework implemented in SPM to analyze our data. We modeled the convolved standard hemodynamic response function with a delta (or “stick”) function at the onset of the preparatory phase (appearance of white triangle – PREP regressor), at the onset of the joystick movement (start of movement – MOV regressor), and at the onset of the response screen (“yes” vs. “no”, RESP regressor). To examine brain regions whose activity fluctuated with trial-by-trial confidence, we took a parametric modulation approach (e.g. Fleming et al., 2018; Pereira et al., 2020): the “MOV” regressor event regressors were modulated by four additional parametric factors (Wood et al., 2008) representing in order, the trial-by-trial values of the angle of the deviation, of the detection response “yes” vs. “no”, of the peak cursor error, and of the confidence rating, leading to the following regression equation:

$$BOLD \sim deviation\_angle + detection\_response + peak\_cursor\_error + confidence$$

However, we also replicated our results using four single regressor models ( $BOLD \sim deviation\_angle$ ,  $BOLD \sim detection\_response$ ,  $BOLD \sim peak\_cursor\_error$ ,  $BOLD \sim confidence$ ). To account for head motion-related variance, we included the six differential parameters derived from the realignment process [x, y, and z translations (in millimeters) plus pitch, roll, and yaw rotations] as regressors of no interest. Low-frequency signal drifts were filtered using a cut-off period of 128s. Global scaling was applied, with each fMRI value rescaled to a percentage value of the average whole-brain signal for that scan.

Contrast images from one-sample t-tests corresponding to each event (PREP, MOV, RESP) and their parametric modulators, were fed into a second-level random-effect analysis. All second-level results are reported at a significance-level of  $p < 0.05$  using cluster-extent family-wise error (FWE) correction with a voxel-height threshold of  $p < 0.001$ . In Figure 3, activations are displayed at a cluster-size threshold of 30 voxels, using MRICroGL (<http://www.cabiatl.com/mricrogl/>). Data and analysis scripts from this study will be made freely available upon acceptance.

## 238 Results

### Participants correct for deviations regardless of detection

240 For both detected (hits) and undetected (misses) deviations, all trajectories correctly ended on the  
target (Figure 1B) showing that participants always adjusted to the experimentally induced  
242 deviations. Reconstructions of what trajectories would be without deviation for hits and misses  
confirmed that participants would have clearly missed the target, even for miss trials (Figure 1C),  
244 showing that they corrected for deviations even when not reporting them. We also assessed the  
relation between the lateral angle of the joystick (i.e. corrective motor command) and the cursor  
246 position (i.e. visual feedback). For this, we cross-correlated cursor position and joystick angle  
vectors across every trial, averaged over trials and participants, and observed that the strongest  
248 correlation occurred with a lag of  $-0.45 \text{ s} \pm 0.01 \text{ s}$  (mean  $\pm$  95% confidence interval; Figure 1D).  
Furthermore, this correlation was significant for detected ( $t(30) = -41.30$ ,  $p < 0.001$ ) and undetected  
250 ( $t(30) = -36.06$ ,  $p < 0.001$ ) deviations, showing that participants corrected deviations regardless of  
whether they reported them or not.

252

### Detection responses rely on cursor error

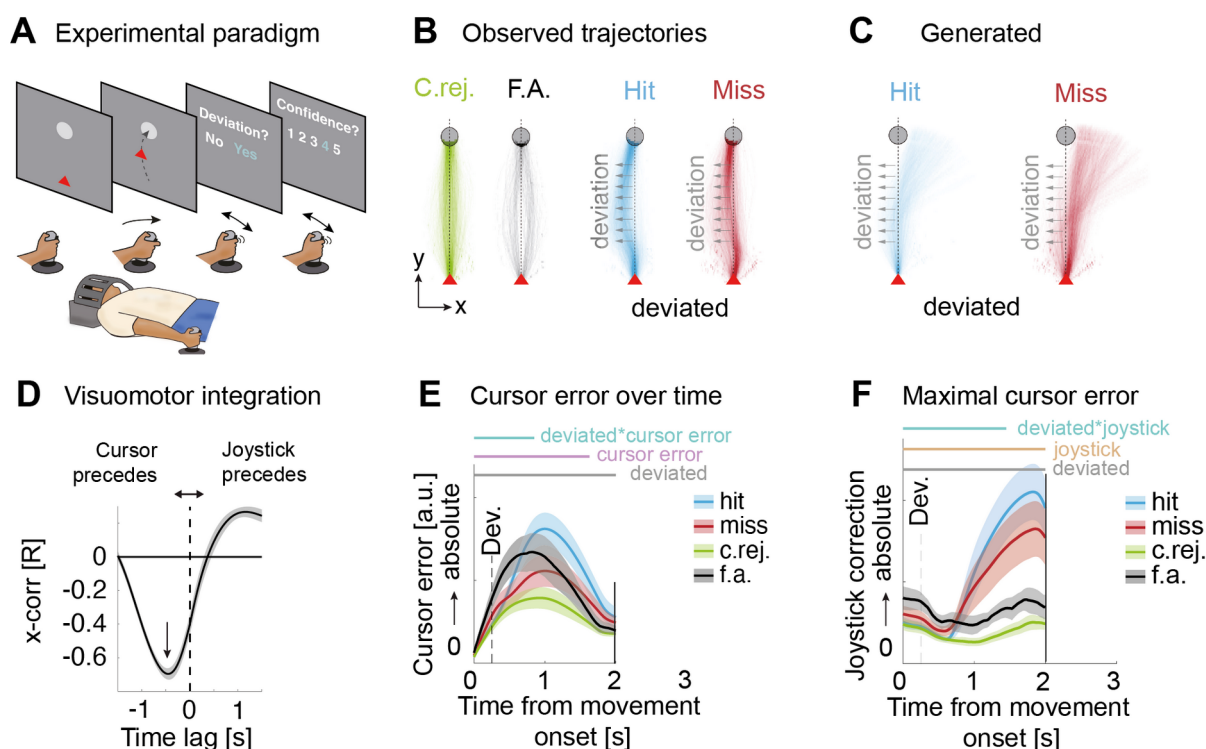
254 A one-up/two-down staircase procedure was used throughout the experiment, resulting in an  
average deviation angle of  $20.59^\circ \pm 2.03$  overall. The detection performance was  $67.71 \% \pm 1.74$   
256 correctly reported deviations ( $d' = 1.69 \pm 0.14$ ), indicating that participants' performance was well  
above chance level. Participants were conservative in their response ( $c = 0.37 \pm 0.08$ ,  $t(30) = 4.63$ ;  
258  $p < 0.001$ ). For deviated trials, neither small variations in the imposed deviation angle due to the  
staircase procedure ( $t(5160) = 0.91$ ,  $p = 0.36$ ), nor the index of the trial ( $t(5160) = -0.11$ ;  $p = 0.91$ )  
260 had an effect on detection responses. We therefore considered perceptual difficulty to be constant  
over time, which is standard in perceptual metacognition (e.g. Fleming et al., 2010).

262

When analyzing cursor position over time, we computed the horizontal distance between the  
264 cursor and the midline as an (unsigned) measure of *cursor error* (Figure 1E). We then regressed  
detection responses using cursor error and trial type (deviated vs. non-deviated) as fixed effects  
266 (generalized mixed effect models for every time point). We found a significant interaction effect  
during the first 0.85 s of the movement ( $p < 0.05$ ; corrected), indicating that during this time, the  
268 relation between detection responses and cursor error depended on whether the trial was actually  
deviated or not. Furthermore, there was a main effect of cursor error without an interaction in a

270 later time window up to 1.63 s after movement onset ( $p < 0.05$ ) and a main effect of trial type  
 272 during the whole movement ( $p < 0.001$ ), suggesting that during this part of the movement, cursor  
 error had similar influence on detection responses regardless of whether the trial was deviated or  
 not, with an additive bias due to trial type. We verified that cursor error also had an effect on  
 274 detection responses in deviated trials ( $p < 0.05$  between 0.49 and 1.94 s) and for undeviated trials  
 ( $p < 0.05$  between 0 and 1.61 s), confirming that the effect of cursor error on detection responses  
 276 was genuine.

278



**Figure 1. Experimental paradigm and detection responses.** A) Participants started every trial by using the joystick to bring a visible cursor (red triangle) to a target (circle) on top of the screen. A deviation was applied to the cursor in 79% of the trials, titrated so as to reach 71% of detection accuracy. Titrated detection rates can be found in Figure 1-1. After reaching the target, participants reported whether the cursor was deviated or not (detection) and how confident they were about their answer. B) Trajectories of the cursor for correct rejections (green), false alarms (black), hits (blue) and misses (red). Note that deviations could be towards the left or the right, but right deviation trials are mirrored and pooled with left trials for display purposes. C) Generated trajectories obtained by recomputing the cursor position had there been no deviation for hits (blue) and misses (red). D) Cross-correlation between cursor error and joystick lateral position. The vertical arrow indicates the time lag at which cross-correlation is strongest. E) Cursor error over time for different signal detection theory categories: hits (blue), misses (red), correct rejections (green) and false alarms (black). The significant ( $p < 0.05$ ; FDR corrected) main effects of cursor error (respectively, of deviation) over time is depicted by the purple (respectively, grey) line; and the significant interaction effects between deviated trials and cursor error is depicted by the cyan line. Model comparison, peak cursor error distributions and discriminability analyses can be found in Figures 1-2, 1-3, 1-4 and 1-5. F) Idem, for joystick correction over time. In all panels, shaded areas indicate 95% confidence intervals.

296 Our results are consistent with a threshold operation on the visual feedback, with higher cursor  
error leading to deviation detection (hit and false-alarms; Figure 1E). Therefore, detection  
298 responses should be best explained by the peak cursor error over the time course of the  
movement (i.e. deviations are detected if the cursor error reaches a threshold value at some point  
300 during the trial). Indeed, this model had a better Bayesian Information Criterion (BIC) than models  
with cursor error sampled at single time points (Figure 1-1A), as well as other measures of visual  
302 feedback (Figure 1-1B). Similarly, models based on putative motor or proprioceptive feedback  
yielded worse BIC (Figure 1-1C). The two main effects in this model were significant (peak cursor  
304 error:  $t(6403) = 5.31$ ;  $p < 0.001$ ; trial type:  $t(6403) = 7.47$ ;  $p < 0.001$ ) and there was no significant  
interaction ( $t(6403) = -1.32$ ;  $p = 0.19$ ), confirming that the occurrence of deviations did not change  
306 how visual information (quantified by peak cursor error) is integrated into detection responses, but  
nevertheless influenced conscious detection (deviation was more frequently reported on deviated  
308 than non-deviated trials overall). Nonetheless, the information carried by the peak cursor error  
would not have been sufficient for participants to have such performance (Figures 1-2 and 1-3),  
310 showing that they must have relied on additional signals, whether those were based on visual,  
motor or proprioceptive information.

312

### **Confidence scales with peak cursor error**

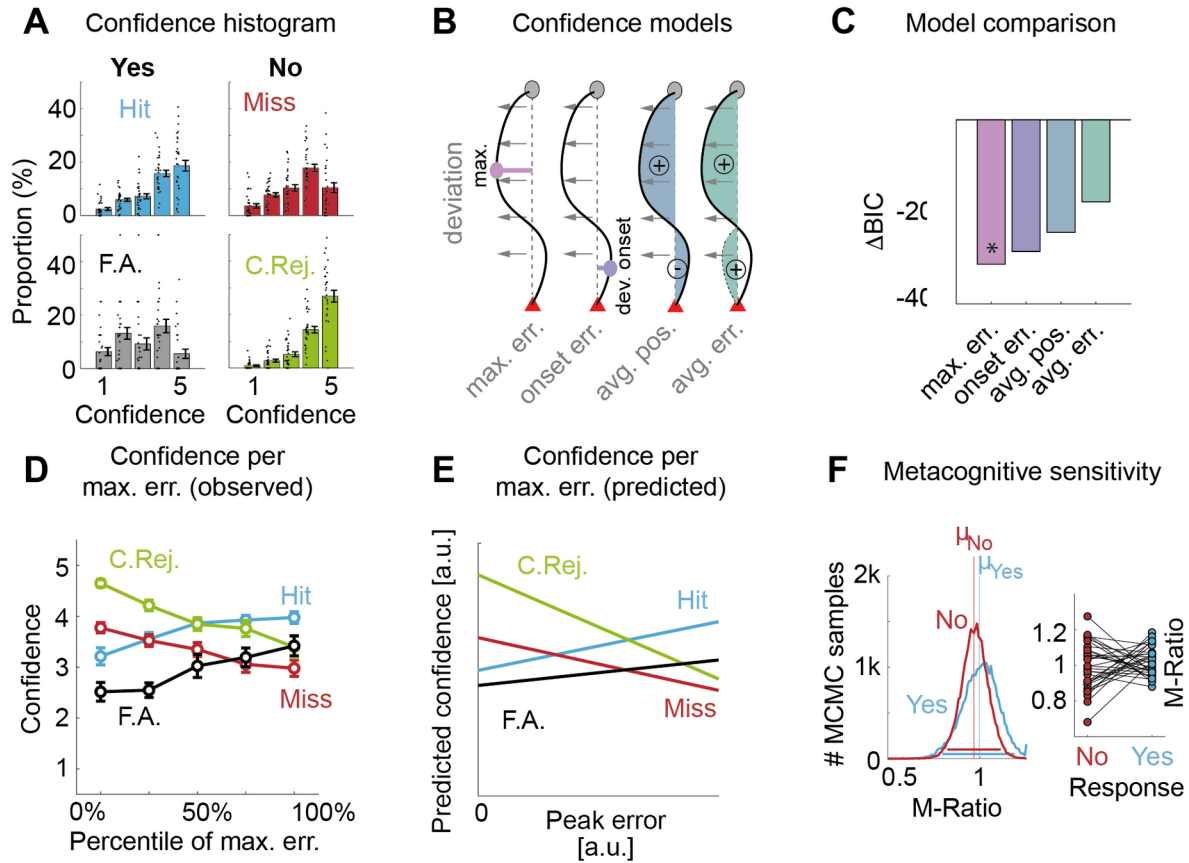
314 We then examined confidence ratings across signal detection theory categories (Figure 2A). To  
understand how visual information is integrated into confidence ratings, we regressed confidence  
316 using various summary statistics of cursor error over time (Figure 2B). Our baseline model  
investigated the relation between detection responses and confidence, and thus included only a  
318 predictor for trial type (deviated vs. non-deviated), as well as for detection response (to model the  
fact that confidence is conditioned to the response; interaction effect:  $z = 8.98$ ,  $p < 0.001$ ). We then  
320 determined the best model in terms of BIC, which was obtained by adding a predictor for peak  
cursor error over time (peak err.). This model showed a lower BIC compared to the baseline model  
322 with a single predictor for detection response (Figure 2C;  $\Delta\text{BIC} = -313.80$ ;  $p < 0.001$ ). Alternative  
models, such as including the cursor error at the onset of the deviation (onset err.), the average  
324 (signed) cursor position (avg. pos.), or the average of the cursor error (avg. err.), yielded lower  
improvements in BIC.

326

Using the winning model, we found an interaction between peak cursor error and detection  
328 responses (Figure 2D-E;  $z = 4.69$ ,  $p < 0.001$ ), showing that confidence ratings were influenced by

the peak cursor error conditioned to the detection response (i.e. when the peak cursor error is high, confidence in detected deviations is high but confidence in undetected deviations is low). There was also an interaction between trial type and peak cursor error ( $z = 4.20$ ,  $p < 0.001$ ) and between detection response and trial type ( $z = 7.087$ ,  $p < 0.001$ ) but no triple interaction between detection response, trial type, and peak cursor error ( $z = -1.38$ ,  $p = 0.17$ ). These results show that participants rated their confidence by conditioning the peak cursor error to their conscious detection, with a confidence bias for non-deviated trials that depended on their response (e.g. more confident for correct rejections versus misses; less confident for false alarms vs. hits).

To better understand the relation between confidence and peak cursor error, we fitted the data independently for deviated and non-deviated trials. We confirmed the interaction of response and peak cursor error for both trial types (deviated:  $z = 5.77$ ,  $p < 0.001$ ; non-deviated:  $z = 5.76$ ,  $p < 0.001$ ). Importantly, and contrary to the hypothesis of limited monitoring, we also found that confidence was significantly related to cursor error size for all trial categories ( $p < 0.01$  for hits, misses, correct rejections, and false alarms). Together, these results show that confidence increases with peak cursor error when participants report a deviation but decreases with peak cursor error when participants do not report a deviation, independently of whether the deviation was actually due to an experimental manipulation (e.g. during misses) or to their own intrinsic visuomotor variability (e.g. during correct rejections) (Figure 2D-E).



**Figure 2. Confidence ratings and metacognition.** A) Distribution of confidence ratings for hits (blue), misses (red) and correct rejections (green). Each black dot represents the data of one participant. B) Schematic depiction of the four regressors used for the four confidence regression models tested: peak err.: peak cursor error; onset. err.: cursor error at the onset of the deviation; avg. pos.: cursor position (signed) averaged along the trajectory; avg. err.: cursor error (unsigned) averaged along the trajectory. C) Improvement in BIC (compared to a model with no cursor information). Note that the peak err. model shows the largest improvement (\*). Additional model comparisons can be found in Figure 2-1. D) Confidence for different percentiles of peak cursor error (peak err.) for hits (blue), misses (red) and correct rejections (green). E) Fixed effects predictions of confidence for comparable levels of peak error (normalized) for hits (blue), misses (red), correct rejections (green) and false-alarms (black). F) Hierarchical Bayesian estimation of response-specific metacognitive efficiency using the M-Ratio. Left: posterior probability for “yes” (blue) and “no” (red) responses. Vertical lines show the mean M-Ratio and horizontal bars show the 95% confidence interval. Right: Single participant values of the M-Ratio. In all panels, shaded areas and whiskers indicate 95% confidence intervals.

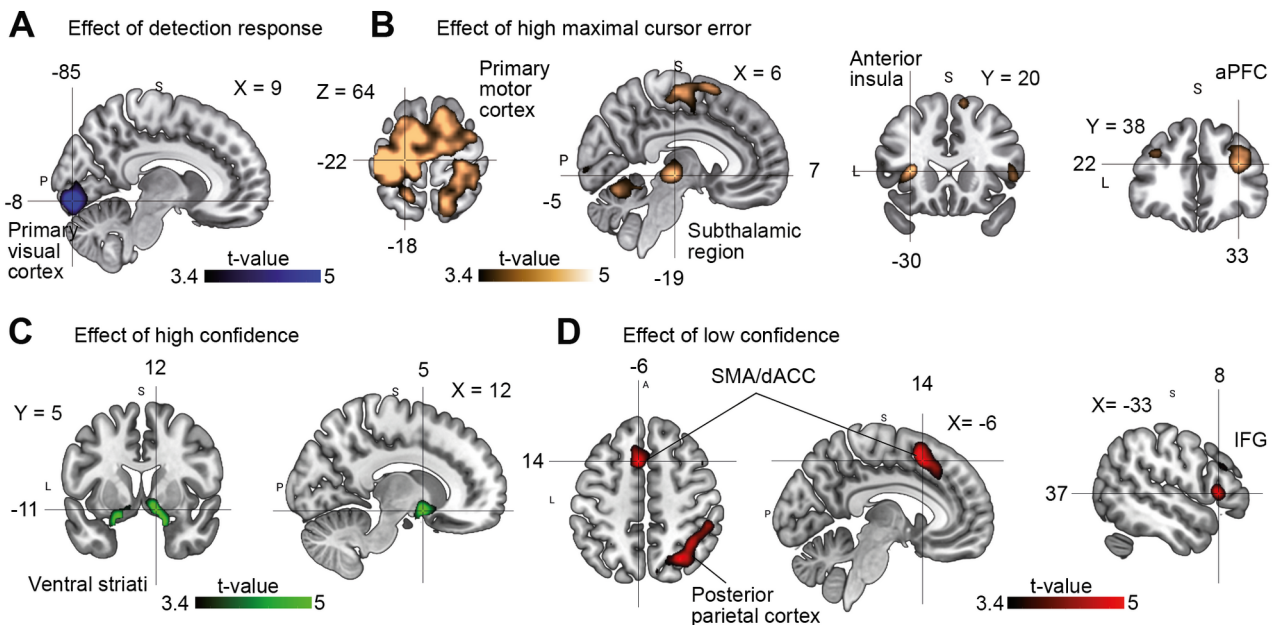
## Preserved metacognitive efficiency for unreported deviations

Finally, to confirm that participants reliably monitored their visuomotor actions, we estimated their (metacognitive) efficiency by measuring to what extent the information available to the detection decision was used to scale their confidence. For this, we fitted a response-specific hierarchical Bayesian model based on signal detection theory (Fleming, 2017) which estimates metacognitive efficiency for “yes” and “no” responses while controlling for task performance (this model used all “yes” and “no” trials in the fitting procedure). Results from this analysis revealed a metacognitive efficiency of 0.96 for the “no” responses (correct rejections and misses), indicating that participants efficiently used all information available for the decision to adjust their confidence ratings.

Metacognitive efficiency for “Yes” responses was similar (0.99; Figure 2F). These results show that metacognitive efficiency was optimal in the sense that confidence ratings were as informative as what would be expected based on detection performance, and thus confirm that reporting deviations does not lead to better metacognitive efficiency.

### Brain correlates of visuomotor performance

To investigate the neural substrate of the cognitive processes described above, we used trial-by-trial measures (confidence, response, peak cursor error, and deviation angle) as parametric regressors to model the fMRI BOLD signal during each movement. These regressors showed only limited correlation among themselves (maximal absolute mean  $R = 0.41$ ; Figure 3-1). First, we found positive relationships between the BOLD signal and detection responses in the right primary visual cortex, meaning that activity increases when participants detected the deviations (Figure 3A; Table 1). We found no significant negative relationships. We also found that larger peak cursor error yielded widespread activity in visual, motor, and subthalamic regions, as well as in the left insula, cingulate, and lateral prefrontal cortex (right inferior frontal gyrus and anterior prefrontal cortex (aPFC); Figure 3B; Table 1). Smaller cursor error yielded no differential activity beyond the statistical threshold and we did not find any neural correlate of deviation angle.



**Figure 3. fMRI results.** Statistical maps of parametric modulation contrast for A) explicit detection (“yes” responses), B) high peak cursor error, C) high confidence and D) low confidence. Note that colors represent different parametric regressors and are independent from Figure 1. Results are displayed at  $p < 0.001$  uncorrected, but all clusters displayed were significant after FWE correction ( $p < 0.05$ ). aPFC: anterior prefrontal cortex; SMA: supplementary motor area; dACC: dorsal anterior cingulate cortex. IFG: inferior frontal gyrus. See text Tables 1 and 2 for other brain activations. Correlation between regressors can be



396 *found in Figure 3-1. Figure 3-2 shows beta values averaged across participants for each level of confidence  
are shown for each ventral striatum using MarsBar (Brett et al., 2002)..*

Brain region	MNI Coordinates mm (xyz)			Cluster extent (voxels)	Significance
“Yes” responses					
Right primary visual cortex	9	-85	-8	602	T= 6.64, $p_{FWEc} < 0.001$
High peak cursor error					
Left primary motor cortex	-18	-22	64	3563	T= 8.18, $p_{FWEc} < 0.001$
Right inferior occipital gyrus	45	-67	1	927	T= 6.17, $p_{FWEc} < 0.001$
Left anterior insula	-30	20	7	632	T= 5.33, $p_{FWEc} < 0.001$
Right subthalamic region	6	-19	-5	285	T= 5.26, $p_{FWEc} < 0.001$
Right lateral anterior prefrontal cortex	33	38	22	229	T= 5.13, $p_{FWEc} < 0.001$
Right mid-cingulate gyrus	15	-19	37	106	T= 4.98, $p_{FWEc} = 0.028$
Right inferior frontal gyrus	48	29	1	279	T= 4.47, $p_{FWEc} < 0.001$
Left calcarine cortex	-12	-85	-8	147	T= 4.41, $p_{FWEc} = 0.008$
Left superior occipital gyrus	-21	-79	28	108	T= 4.16, $p_{FWEc} = 0.026$
No surviving voxels for the contrasts: Low peak error, No responses; $p_{FWEc} = p$ corrected for multiple comparisons at the cluster level					

400 **Brain correlates of confidence**

402 We then turned to confidence and found positive relationships with the BOLD signal, meaning that higher confidence was related to increased bilateral ventral striatum activity, extending to the left amygdala (Figure 3C; Table 2). Conversely, we found negative relationships, meaning that lower confidence was associated with increased activity in the left supplementary motor area (SMA), extending to the dorsal anterior cingulate cortex (dACC), together with the left inferior and middle frontal gyri, as well as the right posterior parietal cortex (Figure 3D; Table 2).

**Table 2. Summary of whole brain regions' activation for confidence**

Brain region	MNI Coordinates mm (xyz)			Cluster extent (voxels)	Significance
High confidence					
Right ventral striatum	12	5	-11	103	T=5.85, p <sub>FWEc</sub> = 0.040
Left amygdala	-21	-1	-17	96	T= 5.80, p <sub>FWEc</sub> =0.049
Extending in the ventral striatum	-15	8	-14		T=4.69
Low confidence					
Left dACC/SMA	-6	14	49	221	T=5.33, p <sub>FWEc</sub> = 0.002
Left inferior frontal gyrus	-51	20	4	119	T=4.97, p <sub>FWEc</sub> = 0.024
Left middle frontal gyrus	-33	8	37	134	T=4.51, p <sub>FWEc</sub> = 0.016
Right posterior parietal cortex/angular gyrus	36	-49	40	403	T=5.43, p <sub>FWEc</sub> < 0.001
Abbreviations: dACC = dorsal anterior cingulate cortex; SMA = supplementary motor area					

## 412 Discussion

414 We studied the behavioral and neural correlates of detection and confidence in a visuomotor  
416 reaching task with perturbations. We found that while participants often reported not being aware  
418 of deviations applied experimentally to the trajectory of their movement, their metacognitive  
420 efficiency was not impaired by this reported lack of awareness. Thus, they adjusted confidence  
422 judgments to the accuracy of their detection responses even when reporting no deviations.  
424 Furthermore, regression model selection revealed that regardless of whether participants were  
aware of the deviation or not, their judgment relied on a summary of visual feedback (i.e. peak  
cursor error), whose magnitude correlated with activity in a specific visuomotor brain network.  
Finally, we uncovered neural activity patterns associated with confidence, with high confidence  
engaging subcortical areas in the ventral striatum, and low confidence engaging a cortical network  
in medial and lateral frontal cortex as well as posterior parietal cortex.

In terms of detection responses, we found that participants tended to report an experimentally  
induced deviation when the visible cursor error was high, and conversely did not report one when  
the cursor error was low. We interpret these detection reports as awareness of the deviation  
although as with any measure based on introspection, it is conceivable that participants had partial  
awareness while reporting being unaware (Mudrik & Deouell, 2022). These findings are consistent  
with awareness of a deviation occurring when cursor error exceeds a threshold: when the peak  
cursor error is higher than this threshold, a deviation is reported. Other regression models of  
detection responses or confidence could not fit the data as well, whether they were based on other  
measures of visual feedback, or on putative proprioceptive signals (indexed by joystick position).  
Still, a part of the variance of participants' performance was not explained by the peak cursor error,  
suggesting that participants use other dimensions of the visual feedback not captured by the peak  
cursor error, or other motor or proprioceptive signals. This idea is in line with previous works  
showing that participants at least partially rely on additional information, possibly stemming from  
the proprioceptive inputs (Farrer et al., 2008) and/or the comparison of the perceived cursor  
position relative to a self-generated prediction, based on an efferent copy of their motor command  
(Wolpert et al., 1995). The later assumption agrees well with past research on motor control and  
agency (Haggard, 2017). As the contribution of these different signals is still unclear in the context  
of motor control (e.g. Kasuga et al., 2022), future studies will be required to disentangle their  
contribution to deviation detection and its associated confidence.

To further assess how well participants monitored their movements, we examined their metacognitive efficiency when they reported being unaware of deviations that they nonetheless successfully corrected. Remarkably, metacognitive efficiency was similar when participants reported deviations and when they did not, suggesting that even if the visuomotor information was not enough to report a deviation, it could still be harnessed to rate confidence. Moreover, we found that confidence increased with peak cursor error when participants reported the deviation, and instead decreased with peak cursor error when participants did not report the deviation. This behavior reveals a judicious use of visuomotor signals to guide confidence, irrespective of the reported awareness of the deviation. We therefore surmise that participants were able to consciously monitor their unfolding movements by having access to at least a summary of visuomotor cues that guided their explicit detection and confidence responses, similar to what could occur for internally accumulated sensory signals (Pereira et al., 2021). The factors that constitute the evidence used to compute confidence are still unknown. Since metacognitive efficiency was as good as expected considering detection performance, it is possible that they directly derive from the decisional signal used for detection responses. Therefore, our findings question the notion that small deviations are corrected mostly in the absence of awareness (Blakemore & Frith, 2003; Custers & Aarts, 2010): people might be unaware of some deviations that they correct, but they still have a calibrated ‘feel’ of their performance.

Our finding of preserved metacognition for “no” responses thus sharply contrasts with visual metacognition studies describing lower metacognition for putatively unaware stimuli (Kanai et al., 2010; Mazor et al., 2020; Pereira et al., 2021). Considering that the factors responsible for metacognitive inefficiency are still unclear (Shekhar & Rahnev, 2021), we can only speculate on why metacognitive performance does not decrease for unaware deviations in our task. Confidence for unaware visual stimuli was proposed to depend on monitoring attention instead of perceptual evidence (Kanai et al., 2010; Mazor et al., 2020). According to this view, confidence for aware and unaware stimuli might involve different mechanisms. In other studies, reduced metacognitive efficiency for unaware stimuli was accounted for by a single mechanism (Kellij et al., 2021; Pereira et al., 2021), based on the fact that the variances for noise and stimuli differ. It is therefore possible that, in our study, equal variance between the noise and the signal allowed metacognition to be preserved for “no” responses. Indeed, the absence of response-specific differences in metacognitive sensitivity suggests that neither unequal variance nor response-specific

metacognition are good candidates to explain our data. It is therefore possible that unequal  
478 variance occurs only in perceptual metacognition studies with perceptual uncertainty on the  
stimulus (which are either weak or embedded in noise). In our study, there is no perceptual  
480 uncertainty on peak cursor error, it is simply smaller for unaware deviations, therefore not leading  
to unequal variance nor lower metacognitive sensitivity for unaware deviations.

482

Our fMRI results showed extended activations associated with increasing peak cursor error,  
484 involving sensorimotor regions, as well as occipital areas, anterior prefrontal and mid-cingulate  
cortex, insula, inferior frontal gyrus, and subthalamic regions. These results suggest the existence  
486 of widespread action-monitoring processes (Limanowski et al., 2017), possibly responsible for the  
integration of visual and proprioceptive feedback signals with motor action planning. Interestingly,  
488 the aPFC – including the frontal pole and posterior regions of Brodmann area 46 (Neubert et al.,  
2014) – has been extensively linked to perceptual metacognition in studies using voxel-based  
490 morphometry (Fleming et al., 2010), TMS (Rahnev et al., 2016) or patients with focal lesions  
(Fleming et al., 2014). In our study, activity in the aPFC was modulated by the peak cursor error,  
492 and not by the confidence regressor, suggesting that the aPFC might not be involved in confidence  
*per se* but rather in monitoring the performance of actions. Further in line with our interpretation,  
494 patients with prefrontal lesions reported fewer deviations than healthy controls despite similar  
corrective behavior in a task similar to ours (Slachevsky et al., 2001). In sum, our finding of  
496 increased aPFC activity reflecting cursor errors and not confidence pleads in favor of a role of the  
aPFC for monitoring action performance rather than the accuracy of decisions.

498

Finally, we found that low confidence was associated with higher activity in a network of medial  
500 frontal cortex, left inferior and middle frontal gyri, and right posterior parietal cortex, providing novel  
support for a role of fronto-parietal regions in metacognitive processes via graded confidence  
502 computation (Hebart et al., 2016; Vaccaro & Fleming, 2018), extending here to the motor rather  
than purely perceptual domain. Conversely, high confidence engaged the ventral striatum, again  
504 corroborating fMRI findings in the perceptual domain (Hebart et al., 2016; Guggenmos et al., 2016;  
Mazor et al., 2020). Apart from its well-known involvement in reward-based learning (Daniel &  
506 Pollmann, 2014), the ventral striatum also computes pseudo-reward prediction errors – defined as  
reward predictions errors related to the subjectively perceived progress in a given task, rather than  
508 merely driven by external (e.g. monetary) reward (Westbrook et al., 2016). Our results therefore  
dovetail with a broader putative role of the ventral striatum for valuation information predicting

510 reward (Schultz et al., 1992; Daniel & Pollmann, 2014), whereby in the absence of feedback, the  
valuation information corresponds to confidence (Daniel & Pollmann, 2014). Taken together, these  
512 findings support a key role of the ventral striatum in monitoring decisional signals for confidence  
(Daniel & Pollmann, 2014; Hebart et al., 2016; Vaccaro & Fleming, 2018) that can be used to adapt  
514 subsequent behavior even in absence of external feedback (Guggenmos et al., 2016).

516 To conclude, our study sheds new light on both cognitive and neural mechanisms that underpin the  
monitoring of visuomotor behavior: participants base their detection and confidence reports on at  
518 least some summary statistic of visual feedback. We also map these monitoring and correcting  
processes to activity in an extended brain network including the aPFC, calling for a revision of its  
520 role in tracking confidence only. Importantly, although participants did report not being aware of  
some deviations, their confidence ratings were still calibrated to their actual performance on a trial-  
522 by-trial basis and guided by the same summary statistics as detection responses. These results  
may offer a plausible explanation for a paradox: that humans perform corrective actions in the  
524 absence of awareness but are good at attributing actions to themselves or to an external agent (de  
Vignemont & Fournieret, 2004). We argue that even if participants are unaware of their corrections,  
526 they can still accurately report their confidence through an efficient summary statistic of action  
parameters. They only become aware that something is wrong when that summary statistic  
528 exceeds what could be expected from their own intrinsic motor variability. Elucidating these  
mechanisms bears important clinical implications as deficits in the awareness of action have been  
530 linked to psychiatric diseases (Blakemore & Frith, 2003) such as schizophrenia (Frith et al., 2000;  
Voss et al., 2010). Our methodology should catalyze future research in the visuomotor domain to  
532 uncover metacognitive deficit in schizophrenia, or more generally, psychosis spectrum (Rouy et al.,  
2021), and what is the underlying defective process. It will be important to examine whether these  
534 clinical populations employ the same mechanisms to compute confidence as we describe here,  
and if so, how deficits in such mechanisms relate to specific pathophysiologic dimensions (positive  
536 i.e. psychotic and negative i.e. amotivational symptoms) and to particular brain substrates.

## 538 References

Arbuzova, P., Peters, C., Röd, L., Koß, C., Maurer, H., Maurer, L. K., Müller, H., Verrel, J., & Filevich, E.  
540 (2020). *Measuring Metacognition of Direct and Indirect Parameters of Voluntary Movement*  
[Preprint]. Neuroscience.

- 542 Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- 544 Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan). Region of interest analysis using an SPM
- 546 toolbox.
- 548 Binsted, G., Brownell, K., Vorontsova, Z., Heath, M., & Saucier, D. (2007). Visuomotor system uses target features unavailable to conscious awareness. *Proceedings of the National Academy of Sciences*, 104(31), 12669–12672.
- 550 Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, 6(6), 237–242.
- 552 Blakemore, S.-J., & Frith, C. (2003). Self-awareness and action. *Current Opinion in Neurobiology*, 13(2), 219-224.
- 554 Charles, L., Chardin, C., & Haggard, P. (2020). Evidence for metacognitive bias in perception of voluntary action. *Cognition*, 194, 104041.
- 556 Christensen, R. H. B. (2019). ordinal—Regression Models for Ordinal Data. R package version 2019.12-10. <https://CRAN.R-project.org/package=ordinal>.
- 558 Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. *Science*, 329(5987), 47–50.
- 560 Daniel, R., & Pollmann, S. (2014). A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiology of Learning and Memory*, 114, 90–
- 562 de Vignemont, F., & Fournieret, P. (2004). The sense of agency: A philosophical and empirical review of the “Who” system. *Consciousness and Cognition*, 13(1), 1–19.
- 564 Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of Distorted Visual Feedback on the Sense of Agency. *Behavioural Neurology*, 19(1–2), 53–57.
- 566 Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive- developmental Inquiry. *American Psychologist*, 34(10), 906–911.
- 568 Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating Introspective Accuracy to Individual Differences in Brain Structure. *Science*, 329(5998), 1541–1543.
- 570 Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal Contributions to Metacognition in Perceptual Decision Making. *Journal of Neuroscience*, 32(18), 6117–6125.
- 572 Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
- 574 Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137(10), 2811-2822.
- 576 <https://doi.org/10.1093/brain/awu221>



578 Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in  
metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137(10), 2811–2822.

580 Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from  
confidence ratings. *Neuroscience of Consciousness*, 2017(1).

582 Fournieret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal  
subjects. *Neuropsychologia*, 36(11), 1133–1140.

584 Francken, J. C., Beerendonk, L., Molenaar, D., Fahrenfort, J. J., Kiverstein, J. D., Seth, A. K., & van Gaal, S.  
(2022). An academic survey on theoretical foundations, common assumptions and the current state  
of consciousness science. *Neuroscience of Consciousness*, 2022(1)

586 Frith, C. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain  
Research Reviews*, 31(2–3), 357–363.

588 Gaveau, V., Pisella, L., Priot, A.-E., Fukui, T., Rossetti, Y., Pélisson, D., & Prablanc, C. (2014). Automatic  
online control of motor adjustments in reaching and grasping. *Neuropsychologia*, 55, 25–40.

590 Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not  
depend on vision of the hand or perception of target displacement. *Nature*, 320(6064), 748–750.

592 Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide  
perceptual learning in the absence of external feedback. *ELife*, 5, e13388.

594 Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4), 196–207.

596 Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2016). The Relationship between Perceptual  
Decision Variables and Confidence in the Human Brain. *Cerebral Cortex*, 26(1), 118–130.

598 Kanai, R., Walsh, V., & Tseng, C. (2010). Subjective discriminability of invisibility: A framework for  
distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*,  
19(4), 1045–1057.

600 Kasuga, S., Crevecoeur, F., Cross, K. P., Balalaie, P., & Scott, S. H. (2022). Integration of proprioceptive and  
visual feedback during online control of reaching. *Journal of Neurophysiology*, 127(2), 354–372.

602 Kellij, S., Fahrenfort, J., Lau, H., Peters, M. A. K., & Odegaard, B. (2021). An investigation of how relative  
precision of target encoding influences metacognitive performance. *Attention, Perception, &  
Psychophysics*, 83(1), 512–524.

604 Limanowski, J., Kirilina, E., & Blankenburg, F. (2017). Neuronal correlates of continuous manual tracking  
under varying visual movement feedback in a virtual reality environment. *NeuroImage*, 146, 81–89.

606 Locke, S. M., Mamassian, P., & Landy, M. S. (2020). Performance monitoring for sensorimotor confidence: A  
visuomotor tracking study. *Cognition*, 104396.

610 Maniscalco, B., & Lau, H. (2014). Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-d',  
Response-Specific Meta-d', and the Unequal Variance SDT Model. In S. M. Fleming & C. D. Frith  
(Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 25–66). Springer Berlin Heidelberg.

612 Mazar, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for  
detecting, but not discriminating visual stimuli. *ELife*, 9, e53900.

614 Miele, D. B., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating Neural Correlates of Action  
Monitoring and Metacognition of Agency. *Journal of Cognitive Neuroscience*, 23(11), 3620-3636.  
616 [https://doi.org/10.1162/jocn\\_a.00052](https://doi.org/10.1162/jocn_a.00052)

Mudrik, L., & Deouell, L. Y. (2022). Neuroscientific Evidence for Processing Without Awareness. *Annual  
618 Review of Neuroscience*, 45(1), 403-423.

Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. S. (2014). Comparison of Human  
620 Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal  
Cortex. *Neuron*, 81(3), 700–713. <https://doi.org/10.1016/j.neuron.2013.11.012>

622 Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., Van De  
Ville, D., & Millán, J. del R. (2020). Disentangling the origins of confidence in speeded perceptual  
624 judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, 117(15),  
8382–8390.

626 Pereira, M., Megevand, P., Tan, M. X., Chang, W., Wang, S., Rezai, A., Seeck, M., Corniola, M., Momjian,  
S., Bernasconi, F., Blanke, O., & Faivre, N. (2021). Evidence accumulation relates to perceptual  
628 consciousness and monitoring. *Nature Communications*, 12(1), 3261.

Rouy, M., Saliou, P., Nalborczyk, L., Pereira, M., Roux, P., & Faivre, N. (2021). Systematic review and  
630 meta-analysis of metacognitive abilities in individuals with schizophrenia spectrum disorders.  
*Neuroscience & Biobehavioral Reviews*, S0149763421001329.

632 Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S., & D'Esposito, M. (2016). Causal evidence for frontal  
cortex organization for perceptual decision making. *Proceedings of the National Academy of  
634 Sciences*, 113(21), 6059–6064.

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L.  
636 Y., Balci, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A.,  
Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., ... Zylberberg, A. (2020). The Confidence  
638 Database. *Nature Human Behaviour*, 4(3), 317–325

Shekhar, M., & Rahnev, D. (2021). Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences*,  
640 25(1), 12–23.

Schultz, W., Apicella, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum  
642 related to the expectation of reward. *The Journal of Neuroscience*, 12(12), 4595–4610.

Sinanaj, I., Cojan, Y., & Vuilleumier, P. (2015). Inter-individual variability in metacognitive ability for  
644 visuomotor performance and underlying brain structures. *Consciousness and Cognition*, 36, 327–  
337.

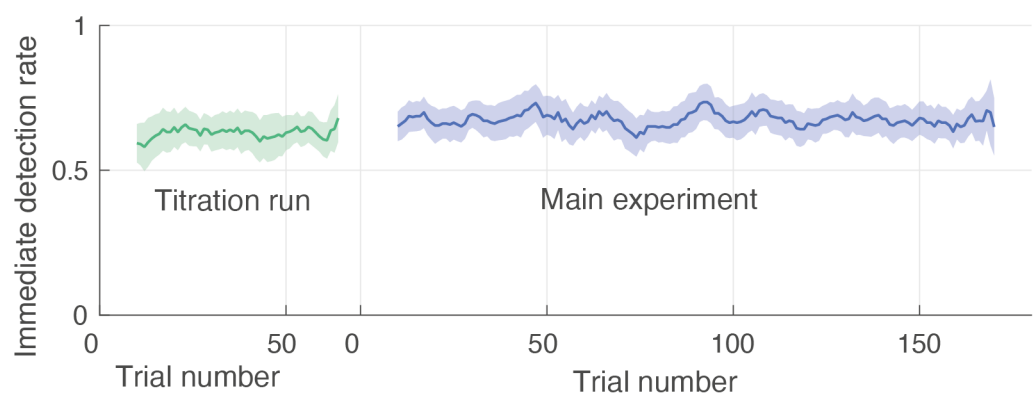
- 646 Slachevsky, A., Pillon, B., Fournieret, P., Pradat-Diehl, P., Jeannerod, M., & Dubois, B. (2001). Preserved  
Adjustment but Impaired Awareness in a Sensory-Motor Conflict following Prefrontal Lesions.  
648 *Journal of Cognitive Neuroscience*, 13(3), 332–340.
- Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of  
650 neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, 2,  
239821281881059.
- 652 Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in  
schizophrenia: A specific deficit in predicting action consequences. *Brain*, 133(10), 3104–3112.
- 654 Westbrook, A. and Braver, T.S. (2016) Dopamine Does Double Duty in Motivating Cognitive Effort. *Neuron*  
89, 695–710
- 656 Wolpert, D., Ghahramani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*,  
269(5232), 1880–1882.
- 658 Wood, G., Nuerk, H.-C., Sturm, D., & Willmes, K. (2008). Using parametric regressors to disentangle  
properties of multi-feature processes. *Behavioral and Brain Functions*, 4(1), 38.

660

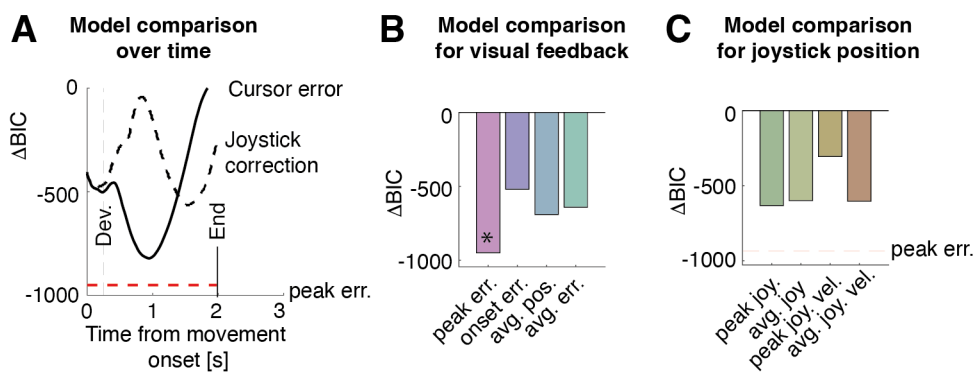
662

# Extended data

664



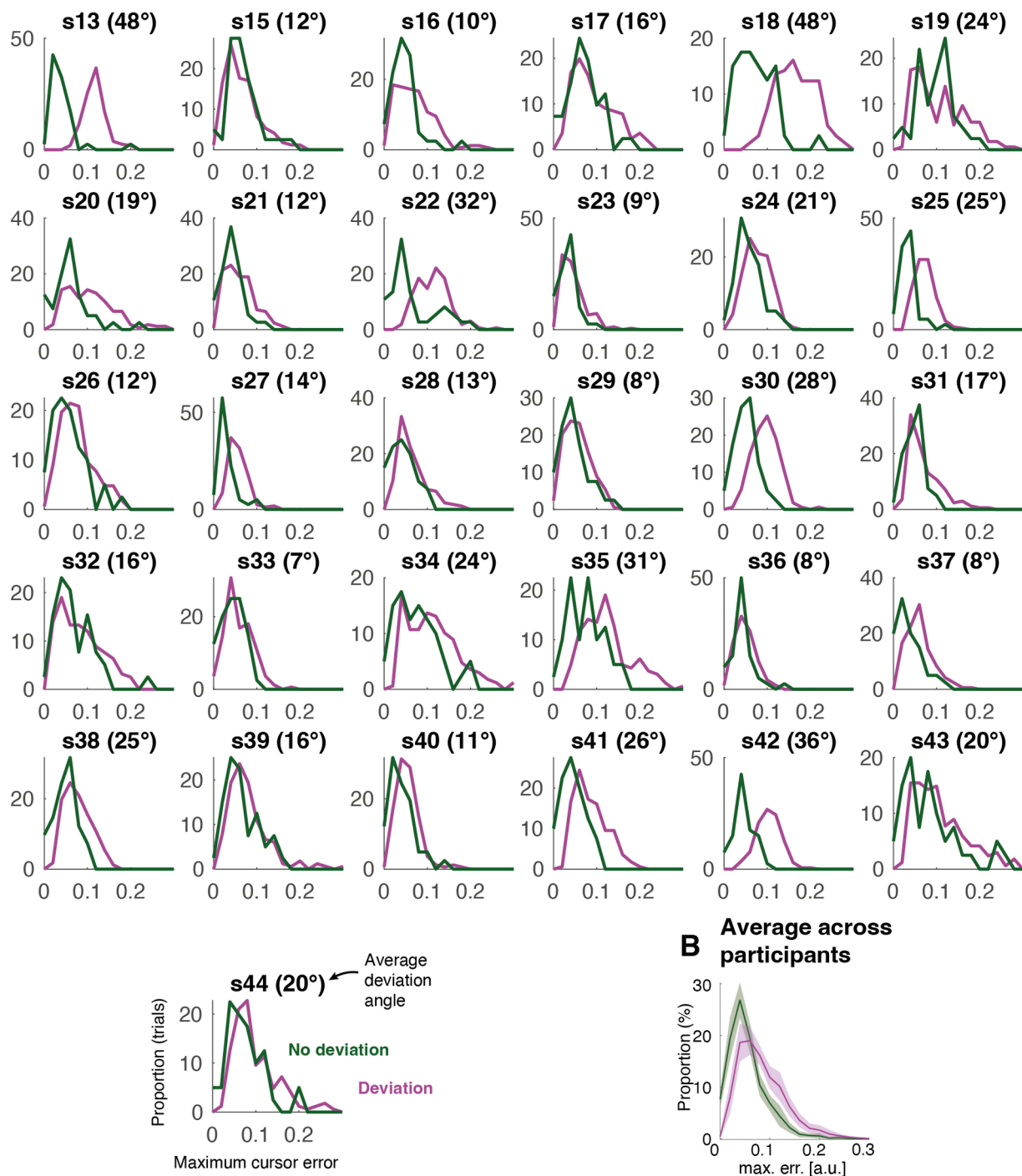
666 **Figure 1-1**  
668 **Staircase convergence.** Immediate detection rate estimated using a 10-trial running average excluding  
670 catch trials for the 62 trials titration run (80 trials in total) in green and the main experiment in blue. All  
672 participants had a minimum of 8 reversals missing a deviation in trial  $n-1$  and detecting it in trial  $n$  or  
detecting a deviation in trial  $n-1$  and missing it in trial  $n$  during the titration run showing good convergence.  
There are not differences between the number of reversals in the first ( $36.13 \pm 1.45$ ) and second run ( $35.45 \pm 1.46$ ) of the main experiment ( $z = 0.41$ ,  $p = 0.68$ ).



674 **Figure 1 – 2**  
676 **Regression model comparison**  
678 Bayesian information criteria (BIC) difference of different models compared to a model with only trial type  
(deviated or non deviated) as fixed effect. The lower the BIC, the better the model. A) BIC differences  
680 compared to the baseline model with no cursor error for models using the cursor error (solid trace) or the  
joystick correction (absolute value of the lateral position of the joystick; dashed line) at different times from  
682 movement onset. The horizontal purple dashed shows the BIC difference when using the peak cursor error,  
which is consistently lower. B) BIC differences for different measures of cursor position: peak err.: maximal  
684 cursor error; onset. err.: cursor error at the onset of the deviation; avg. pos.: cursor position (signed)  
averaged along the trajectory; avg. err.: cursor error (unsigned) averaged along the trajectory. The asterisk  
686 shows the best model (peak cursor error; peak err.) C) BIC differences for different measures of joystick  
position: peak joy.: peak joystick horizontal position, avg. joy.: joystick horizontal position averaged the  
688 trajectory, peak joy. vel.: peak joystick horizontal velocity, avg. joy. vel.: joystick horizontal velocity averaged  
the trajectory. The horizontal purple dashed line shows the BIC difference when using the peak cursor error  
(peak err.), which is consistently lower.

690

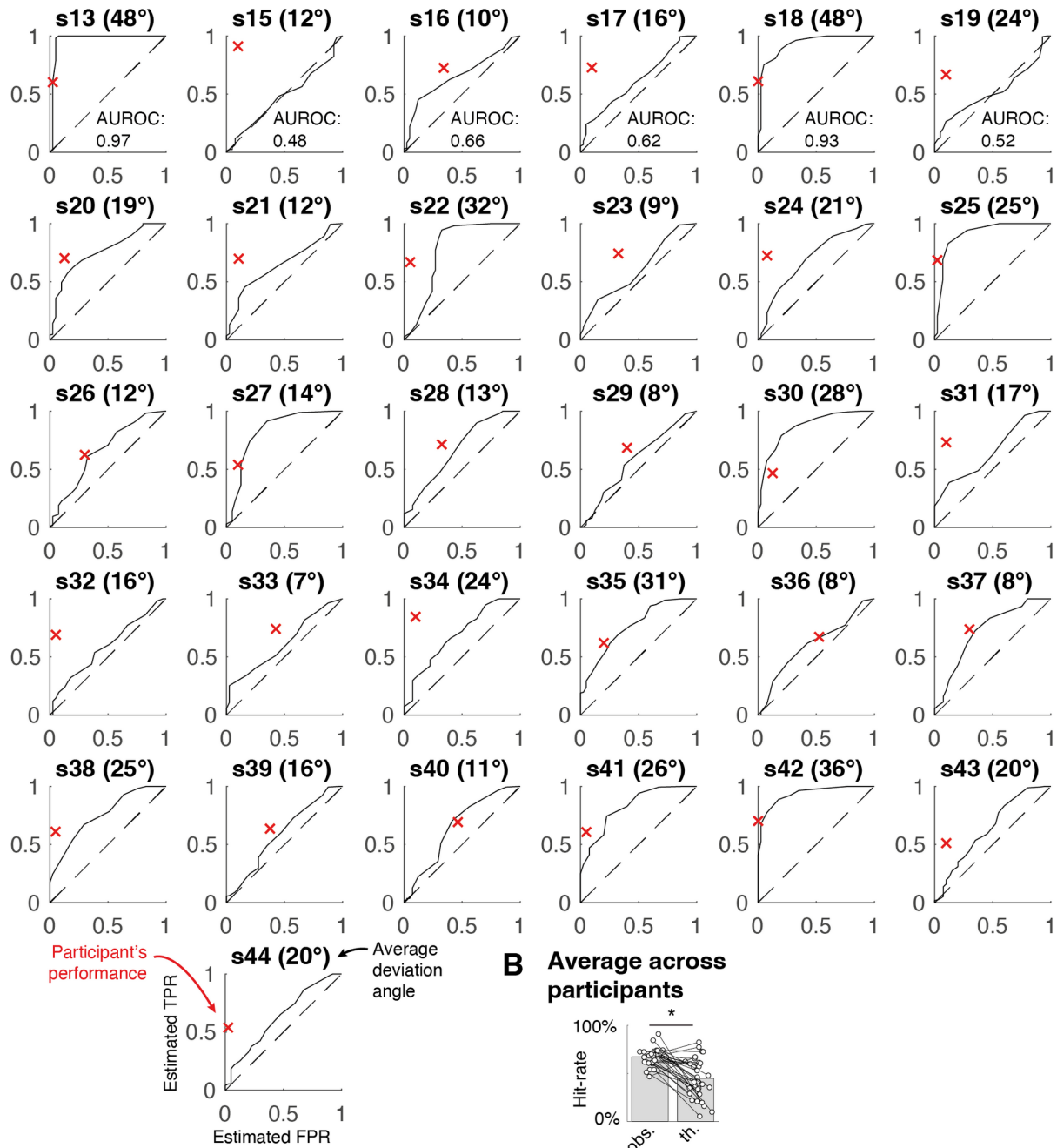
## A Peak cursor error distribution for each participant



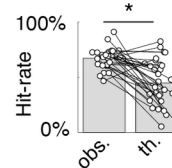
**Figure 1 - 3**  
**Distribution of maximal cursor error**

A) For deviated (light blue) and non-deviated (dashed black) trials for every participant. The average deviation angle (i.e. easiness of the task) obtained through the staircase procedure is displayed in the title of each panel. Only participants who failed to reach a low average deviation angle showed highly separable distributions (e.g. s13, s18, s42). B) Idem, but averaged across participants. Shaded area represents the 95% confidence interval.

## A AROC for each participant



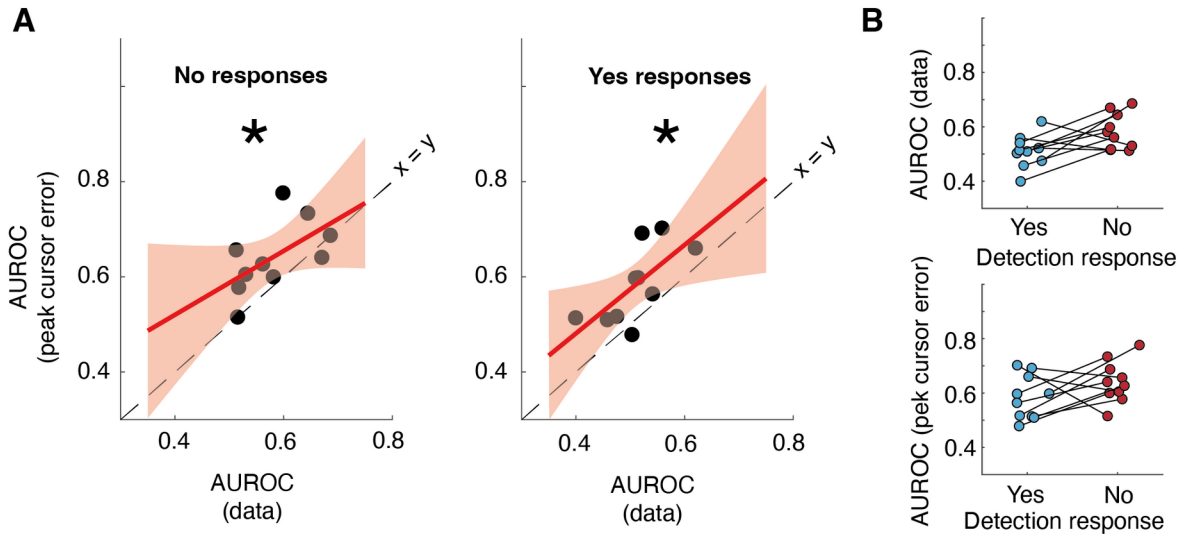
## B Average across participants



**Figure 1 - 4**  
**Receiving operating curve (ROC) analysis for detection.**

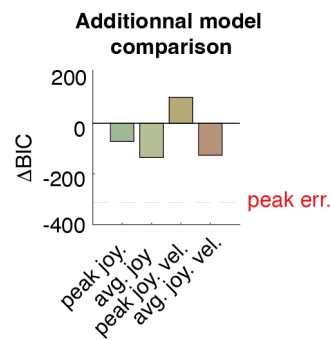
A) To compare individual participants' performances against the information contained in the maximal cursor error, we estimated the ROC curve (black trace) by sliding the criterion along the horizontal axis in Figure 1 - 2 while computing true positive rate (estimated TPR) and false positive rate (estimated FPR). The dashed black trace represents the theoretical ROC curve for random performance (i.e. for non-separable distributions of maximal cursor error between deviated and non-deviated trials). We then displayed the actual performance from each participant in terms of TPR and FPR (red cross). If participants would rely only on the maximal cursor error to detect deviations, we would expect their performance to be under (or close to) the estimated ROC curve. This was the case only for a few participants (e.g. s30, s36, s40). These results indicate that most participants used additional information to detect deviations. We obtained very similar results using other indices (see Figure 2B) such as cursor error at deviation onset, averaged cursor position, and averaged cursor error, suggesting that participants did not use another more integrative strategy either, even based on other visual cues from the cursor. B) Across participants, this theoretical hit-rate, corresponding to the theoretical performance that could be achieved using peak cursor error alone, was much lower (th.; 42 %  $\pm$  4) than the observed hit-rate (obs.; 67%  $\pm$  2;  $t(30) = 5.16$ ;  $p < 0.001$ ). Similar results

were obtained using other summary statistics of cursor error over time (Figure 2B), such as cursor error at the onset of the deviation (onset err.), the average (signed) cursor position (avg. pos.), or the average of the cursor error (avg. err.). These result suggest that participants could not have reached the observed detection performance relying on the peak cursor error information alone. C) Between-participant correlation between the estimated area under the ROC curve (AUROC) quantifying the information contained in the maximal cursor error and the average deviation angle achieved by the staircase procedure. This relation shows that the less good participants were at the task (high deviation angle), the more information they could use from the maximal cursor error.



**Figure 1 – 5.**

**Receiving operating curve (ROC) analysis for confidence.** A. Simulated AUROC (vertical axis) compared to AUROCs in the data (horizontal axis) for yes (detected ; left) and no (undetected ; right) responses. Each point represents one of the 10 participants for which we had enough false alarms. Red lines represent the regression line with shaded 95 % confidence intervals. B. Differences of AUROC in the data (top) and simulated from the peak cursor error (bottom) between yes and no responses for each participant confirming the absence of evidence for an increase in metacognitive sensitivity for detected deviations (Yes responses).



**Figure 2 – 1**

**Additional model comparison:** BIC differences for different (unsigned) measures of joystick position: peak joy.: peak joystick horizontal position, avg. joy.: joystick horizontal position averaged the trajectory, peak joy. vel.: peak joystick horizontal velocity, avg. joy. vel.: joystick horizontal velocity averaged the trajectory. The horizontal purple dashed shows the BIC difference when using the peak cursor error, which is consistently lower.

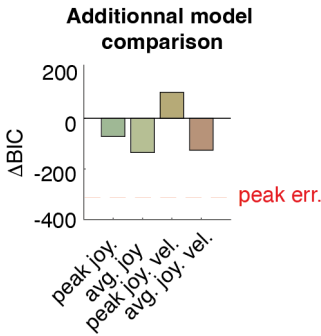
744

CORRELATION	Angle	Detection	Peak error	Confidence
Angle	1.00			
Detection	0.41 ± 0.02	1.00		
Peak error	0.30 ± 0.03	0.37 ± 0.04	1.00	
Confidence	-0.11 ± 0.03	0.02 ± 0.03	-0.04 ± 0.04	1.00

**Figure 3 – 1**  
**Correlation table of the four parametric regressors used in our fMRI.**

746

748



**Figure 3 – 2**  
**Beta values averaged across participants for each level of confidence.** For each subject, we modeled each level of confidence as an individual condition. Thus each subject had 5 regressors, one regressor for condition. Using MarsBar (Brett et al., 2002), we then extracted BOLD estimates for each confidence level with a sphere (radius of 3 mm) around the peak voxel for the left (MNI: 12, 5, -11) and right ventral striatum (MNI: -15, 8, -14) of each participant.

752

754

756