Exogenous and endogenous sources of uncertainty inform global performance monitoring

Marie Chancel[1], Elisa Filevich[2], Nathan Faivre[1]

[1]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France
[2]Institute of Systems Motor Science, Humboldt Universität zu Berlin, 10099 Berlin, Germany

Keywords
performance monitoring, metacognition, confidence, sensory uncertainty, attentional cueing

Abstract

The present study investigates the contribution of first-order accuracy and uncertainty to global performance monitoring. After a set of four trials of an orientation matching task, participants first reported their perceived mean response, and then a region around that mean estimate corresponding to their estimation of their responses' dispersion. We could assess how first-order performance, endogenous uncertainty (estimated from the variability of first-order performance), and exogenous uncertainty impacted global performance monitoring. In two experiments, we found that participants were able to track and use the average and dispersion of their first-order performance to monitor global performance. The calibration of metacognitive judgments to first-order performance was better when endogenous uncertainty was lower. Similarly, exogenous sources of uncertainty (i.e., stimulus- and attention-related) also modulated the calibration between global metacognitive reports and first-order performance. These results suggest that people can reliably estimate the mean and variability of their performance and use it together with exogenous uncertainty to inform their global performance monitoring. However, this capacity decreases in the presence of both internal and external uncertainty. We discuss these results in light of the role of uncertainty in perceptual metacognition and the relationship between local and global performance monitoring.

Introduction

Imagine hanging a frame on a wall, trying to align it with the wall's edges or with other frames. Before drilling a hole and permanently altering the wall, a certain level of confidence in the frame orientation has to be reached. To know when this acceptable orientation has been achieved, one needs to monitor one's own performance, to build a metacognitive judgment over a series of actions. Metacognition, the ability to critically assess our perception and actions (Dunlosky 2008; Fleming, Dolan, and Frith 2012; Koriat 2007) is thus fundamental to shaping and adjusting behavior in such a task (Desender, Boldt, and Yeung 2018). In studying this crucial aspect of cognition, an increasing number of studies investigate the mechanisms and factors explaining how metacognitive judgements are formed, most commonly by assessing how we compute confidence estimates about our performance on a task (Kepecs and Mainen 2012; Mamassian 2016). Several sources of information for confidence have been identified. Sensory information and prior knowledge contributing to perceptual decisions are also used in confidence judgments (Kiani and Shadlen 2009), although they might not contribute equally (Constant et al. 2022). Confidence also appears to depend on post-decisional processes (Balsdon, Mamassian, and Wyart 2021; van den Berg et al. 2016; Murphy et al. 2015; Pereira, Perrin, and Faivre 2022; Pleskac and Busemeyer 2010), and on the monitoring of action-related signals (Faivre et al. 2018, 2020; Filevich, Koß, and Faivre 2020; Gajdos et al. 2019; Pereira et al. 2020).

Most of the research so far has focused on local confidence judgements following one unique event or task, but recently, the need to develop new paradigms to approach metacognition has been highlighted as a crucial goal for the future of the field (Rahnev et al. 2022) and a few studies saw their interest shift toward the metacognitive evaluation of a series of events (Lee, de Gardelle, and Mamassian 2021) and more global estimations of self-performance (Rouault, Dayan, and Fleming 2019). In the frame example, evaluating one's own performance requires assessing how far from the target orientation the frame is, but may also depend on the magnitude of each consecutive adjustment, i.e., the variability of performance in time. Such broadening to consider performance monitoring following several events is further motivated by the importance of global beliefs in shaping our decisions and actions (Bandura 1977; Elliott et al. 1996; Zacharopoulos et al. 2014). With this line of research in mind, the current study sought to investigate if global performance monitoring pertaining to several repetitions of the same task was formed using similar performance and stimulus-related cues as the metacognitive judgments referring to a single trial, and if different sources of uncertainty (endogenous or exogenous) had comparable impacts.

Humans are able to assess both the mean and variance information about a group of visual stimuli at the perceptual level and use it to guide behavior (Desender et al. 2018; de Gardelle and Summerfield 2011; Ji and Hayward 2021; Michael, de Gardelle, and Summerfield 2014). It has also been shown that individuals are able to monitor their sensorimotor performance when asked to continuously track a visual target on a screen and use it for a subsequent confidence judgment (Locke, Mamassian, and Landy 2020). Moreover, a recent study suggests that both exogenous uncertainty (linked to the stimulus) and endogenous uncertainty (linked to the variance in participants' responses when answering about stimuli) influence confidence (Geurts et al. 2022). Based on these studies, we expected participants to evaluate their own performance based on their mean performance and corresponding variability, in addition to the uncertainty related to the stimuli they observed. Although most studies examine confidence judgments on binary tasks (correct versus incorrect trials), variability in binary responses is calculated in a less straightforward manner than for continuous responses. Here, by measuring performance monitoring following a continuous orientation reproduction task, we could assess how internal response variability affects global metacognitive estimates.

In each trial, participants were asked to reproduce the orientation of a stimulus they had just seen for a brief period of time. After four trials involving the same target orientation (+/- a jitter), we asked participants for two types of global performance monitoring: (1) a report of the mean orientation response across a set of four orientation-matching trials. This report required participants to track their individual trial responses and monitor the accuracy of their global performance over these individual trials. In the case of a participant with perfect performance monitoring, this report should be equal to the average of their actual responses on a set. (2) a report of a confidence zone, i.e., how far around this mean orientation response the participants thought their responses landed in the corresponding set of four trials. This report required the participants to monitor the dispersion of their own responses over these individual trials. In the case of a participant with performance monitoring, this report should overlap perfectly with the dispersion of responses across a set of four trials.

We predicted that participants would use their mean performance, endogenous uncertainty (performance-related), and exogenous uncertainty (stimulus-related) to compute these metacognitive reports. Exogenous uncertainty was manipulated by displaying ordinal or cardinal orientations which are known to differ in terms of sensory noise (i.e., *oblique effect;* Appelle 1972; Girshick, Landy, and Simoncelli 2011). We examined another source of exogenous uncertainty in a second experiment, where we additionally manipulated the allocation of attention by varying the proportion of valid versus invalid exogenous cueing across the set of four orientation matching trials. Such exogenous manipulation is known to increase uncertainty (Carrasco 2011), impacting both perceptual decisions and confidence for isolated events (Denison et al. 2018). We predicted that this change in participants' focus would also influence the participants' global metacognitive reports.

Methods

The study design and analysis plan were registered prior to data acquisition on a public repository (https://osf.io/knufx). All procedures were performed in accordance with ethical standards and were approved by our institutional research committee (CERGA-Avis-2022-16).

Participants

Participants were recruited via the Prolific marketplace (https://www.prolific.co/). Data was collected on a server from the Pavlovia platform (https://pavlovia.org/), and the experimental scripts were written using HTML/JavaScript/CSS, and the JsPsych library (https://www.jspsych.org/7.0/). We adopted an open-ended sequential Bayes factor design, and stopped data collection when our statistical model reached strong evidence for either H0 (the variable of interest does not contribute to reported confidence) or H1 (the variable of interest contributed to reported confidence) for our main variable of interest (SD_{set} = standard deviation of the participant's response within a set of trials), or any interaction including this variable. Thus, data were acquired until a Bayes factor equal or inferior to 0.2 or equal or superior to 5 was obtained. In experiment 1, 40 participants were recruited and 36 participants were included (19 females, mean age +/- SD: 26.2 +/- 7 yo, see exclusion criteria); in experiment 2, 40 participants were recruited and 38 participants were included (18 females, mean age +/- SD: 26.2 +/- 8 yo, see exclusion criteria).

Procedure

Experiment 1

Experiment 1 assessed the relative contributions of external and internal uncertainty to global metacognitive judgments. We manipulated exogenous uncertainty by setting the stimulus orientation to be either cardinal or oblique, and we approximated the endogenous uncertainty using participants' behavioral variability. Thus, our exogenous/endogenous distinction for uncertainty relied on the origin of the change in uncertainty: experimental (i.e., exogenous) or from the participants' response (i.e., endogenous). The first part of the experiment ensured a similar difficulty level across participants using a staircase procedure (Levitt 1971). After a fixation cross (shown for 1 s), a visual stimulus (Gabor patch) made to occupy 80% of the screen's height, was displayed on a dark gray background (HEX color code #222222) for 500 ms, generated by overlaying a 2D Gaussian window on a sine-wave grating (spatial frequency = 2 cycles/°), with an initial Michelson contrast of 0.6, oriented at cardinal (0º, 90º) or oblique (45º, 135º) orientations. The visual stimulus was immediately followed by a mask made with concentric circles and Gaussian noise presented for 300 ms to avoid after-effects (Barbosa and Kouider 2018). Once the mask disappeared, a white response bar appeared at a random initial angle. Participants were instructed to click on the bar and drag it until its orientation matched the orientation of the previously seen Gabor patch. When participants were satisfied with the bar orientation, they validated and ended the trial by pressing the spacebar (Fig. 1. A). If the absolute error between the actual target angle and the participants' response was above 10º, the contrast of the next Gabor patch increased by 0.005. If the absolute error between the actual target angle and the participants' response was under 10º twice in a row, the contrast of the next Gabor patch decreased by 0.005. This staircase-like procedure stopped after 80 trials, and the final contrast level was kept for the rest of the experiment. This procedure allowed us to match the task difficulty between participants. Participants also had the possibility to click on a button to skip a trial when the response bar appeared if they missed the corresponding Gabor patch (e.g. if they looked away when the visual stimulus was presented). They were instructed to use this button only if they did not see the visual stimulus, not if they did not remember or were unsure about its orientation (see exclusion criteria section for the processing of these "missed" trials). This initial part of the experiment lasted around 5 minutes.

The main part of the experiment started after this initial calibration phase. Once again, participants were asked to move the response bar to match the orientation after seeing the same sequence as in the calibration phase (Gabor patch – mask – response bar). However, this time, this sequence was presented four times in a row, with the same angular orientation and a predetermined jitter of either -5, -2.5, 2.5, or 5° with a randomized order. After performing these four trials, participants were asked to assess their overall performance (Fig. 1. A.). To do so, a circle of the same size as the Gabor patch was presented to the participants, with a bar to mark the diameter (at a random orientation) that could be moved and an interactive red "confidence" area around this bar. They were requested (1) to rotate the central bar to match the mean of the orientation they reported in the previous set of trials and then (2) to adjust the size of the confidence area following the instruction: "now that you reported your target orientation, how precise were you in matching this orientation across the past four trials?". Each step was validated by pressing the spacebar. Each orientation level ([0º, 45º, 90º, or 135º]) was repeated 24 times in a pseudorandomized order, and the experiment was divided into 4 blocks. Between each block, the participants were encouraged to take a short break. This part of the experiment lasted approximately 40 min for a total of 384 trials (96 sets).

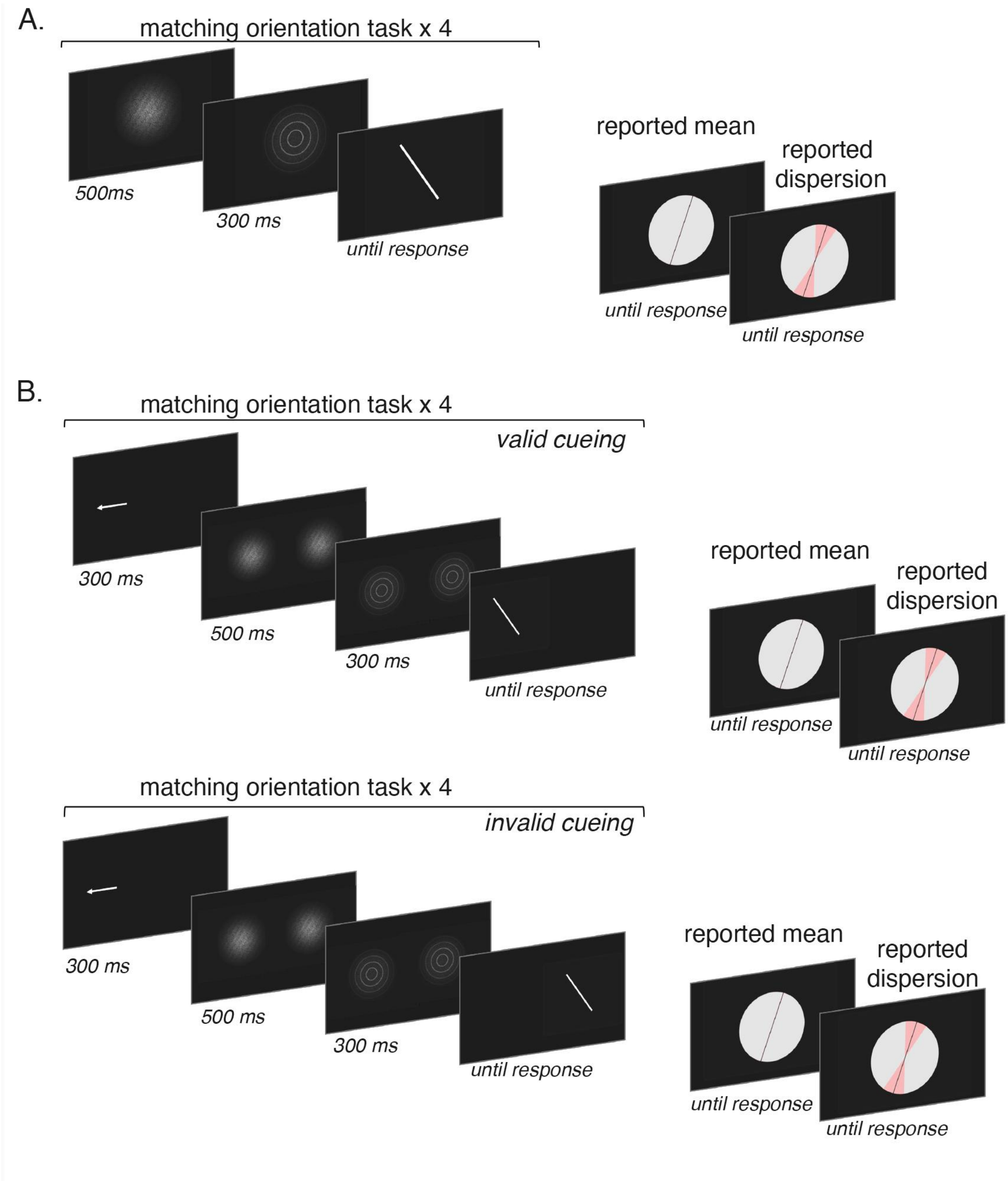

Figure 1: Summary experiments' procedure. A. Bar orientation task: a Gabor patch was displayed at an angle of 0º, 45º, 90º, or 135º (+/- jitter [-5º, -2.5º, 2.5º, 5º]) from the horizontal, followed by a mask made of concentric circles. Then, the response bar appeared, and participants were asked to rotate it until its orientation matched the orientation of the Gabor patch they saw. This sequence of stimuli was used both in the initial calibration and the main phase of the experiment. In the main

phase of the experiment, each target angle (+/- jitter) was presented four times in a row. Then, the participants were asked to report (1) their mean response orientation, (2) their "zone of confidence", i.e., how precise they thought they had reported the correct orientation across the four previous trials. B. Schematic representation of experiment 2 including exogenous attentional cueing. Two Gabor patches were presented to the participants, followed by the response bar appearing on the side of the screen corresponding to the stimulus that should be reported. The exogenous cue shown before displaying the Gabor patches could be valid (same side as the Gabor patch to be reported at the end of the trial, upper panel) or invalid (opposite side to the Gabor patch to be reported at the end of the trial, lower panel). The number of valid/invalid trials varied, creating four different conditions of exogenous cueing.

Experiment 2

Experiment 2 attempted to replicate our main findings from Experiment 1 while introducing another exogenous manipulation of uncertainty by influencing the participants' level of attention through exogenous cueing (Posner 1980). This method is known to induce an additional uncertainty that observers incorporate into their perceptual decision and local confidence estimates (Denison et al. 2018). We hypothesized that this type of attentional-related uncertainty would also affect the participants' performance monitoring over each set of four trials.

To test this hypothesis, we used the same experimental procedure as described above with the following modifications. Participants were no longer presented with just one but two Gabor patches with different orientations (cardinal orientations only, +/- jitter), one on the right side of the screen, the other on the left side of the screen (Figure 1. B). Participants were instructed to remember both orientations. Both Gabor patches were then masked, and a response bar appeared on the side of the screen corresponding to the Gabor's orientation that had to be reported. On each stimulus presentation, before the appearance of the two Gabor patches, an arrow appeared for 300 ms pointing either toward the stimulus to be reported (valid cue) or towards the opposite side (invalid cue). Like in Experiment 1, participants were asked to match the target stimulus' orientation and to assess their performance (mean target orientation and zone of confidence angle) after four trials. Within a set of four trials, the number of valid cues was parametrically varied. A set could contain four (*full* set), three (*valid* set), two (*neutral* set), or one valid cues (invalid set) — but there were no sets with no valid cues —, resulting in four different conditions of cueing at the level of a set. Each cardinal orientation ([0º, 90º]) was repeated 72 times in a pseudorandomized order, divided in 4 blocks. This part of the experiment lasted approximately 60 min for a total of 576 trials (144 sets, including 47 full, 47 valid, 25 neural, and 25 invalid sets). This imbalance in favor of valid cueing was chosen to implement our attentional manipulation, as it meant that participants had a strategic advantage if they took into account the cue instead of simply discarding it.

Data analysis

First-order analyses

We estimated first-order performance as the difference between the target angle (Gabor patch orientation) and the participants' response (response bar orientation) on each trial. We analyzed it as a function of stimulus orientation using Bayesian mixed-effects linear regressions.

- **M_orientation**: $\text{Error}_{\text{Trial}}$ ~ Orientation + (Orientation| participant)

Second-order analyses

We expected participants to take into account their actual performance and the different sources of uncertainty when monitoring their performance over a set of four consecutive trials. Thus, we calculated the mean error ($MeanError_{Set}$) and response standard deviation (SD_{set}) for each set of four trials, and evaluated their contribution to two metacognitive variables (1) *Reported_Mean*, i.e., the absolute difference between the reported mean response orientation and the actual orientation presented across the four trials. (2) *Reported_Dispersion*, i.e., the angle of the confidence area. These two metacognitive variables decrease when participants evaluate that their performance is high. Since we expected participants' metacognitive reports to appropriately reflect their performance and the level of exogenous uncertainty, we predicted accurate performance monitoring i) when the participants' responses within a set were more accurate on average, hence closer to the orientation they had to reproduce ($MeanError_{Set}$ closer to 0), ii) when the participants' responses in a set were less variable and therefore endogenous uncertainty was lower (decreased SD_{set}), iii) when external uncertainty was low, i.e., for cardinal orientations compared to oblique orientations.

To test these predictions, we used Bayesian mixed-effects linear regressions with the following formulae to examine each variable:

> **M1**: Reported_Mean ~ poly(MeanError$_{Set}$,2) * SD$_{set}$ * Orientation + (poly(MeanError$_{Set}$,2) + SD$_{set}$ + Orientation| participant)
>
> **M2**: Reported_Dispersion ~ poly(MeanError$_{Set}$,2) * SD$_{set}$ * Orientation + (poly(MeanError$_{Set}$,2) + SD$_{set}$ + Orientation| participant)

We added a quadratic expansion to $MeanError_{Set}$ using poly(MeanError$_{Set}$,2) to account for the expected U-shape relationship it had with our dependent variables. Indeed, worse performance on a set corresponded to a larger $MeanError_{Set}$, involving either an overshoot (negative $MeanError_{Set}$) or undershoot (positive $MeanError_{Set}$). In both cases, we expected Reported_Mean and Reported_Dispersion to increase with the absolute value of $MeanError_{Set}$ and to decrease when $MeanError_{Set}$ gets closer to 0. In the results section, effects involving the linear component will be noted as $MeanError_{Set}^1$, while the quadratic component will be written as $MeanError_{Set}^2$.

All models were fitted once using an uninformed, neutral prior (Gaussian distribution with mean = 0 and SD = 2) and a second time with a prior informed by the result of a pilot study (N = 19 participants) using the same experimental procedure as in Experiment 1. Those informed priors had the same means as the pilot posterior means, and SDs equal to 1.5 times the pilot posterior SDs. We reasoned that taking into account the results with both types of priors is of interest since using an informed prior leads to more precise but more biased estimates (Morris, Vesk, and McCarthy 2013; Zampieri et al. 2021). Uninformed and informed priors provided qualitatively similar results. The results obtained with informed priors can be found in the Supplementary Materials.

Transformations

To account for between-participants variation, we *z*-scored all ratings separately for each participant and used z-scored values to fit models, except for $MeanError_{Set}$: $MeanError_{Set}$

could be a positive or negative angle and participants were not expected to equally over and undershoot when matching the target orientation. Therefore, we did not expect $MeanError_{Set}$ to be centered on 0. In an hypothetical case where a participant only overshot, resulting in only negative $MeanError_{Set}$, after a Z-transformation the largest $MeanError_{Set}\_z$ values (most positive z-values) would be the ones closest to 0 in the raw data. Given our models' syntax, for this theoretical participant, the model would test an increase in our metacognitive variables when $MeanError_{Set}$ gets closer to 0 while we were actually predicting an increase in our metacognitive variables when the magnitude of $MeanError_{Set}$ increased. Note that we could not simply use absolute values of $MeanError_{Set}$ in our model because of collinearity between $|MeanError_{Set}|$ and SD_{set}.

Inference criteria
We used a criterion of BF10 = 5 or BF01 = 0.02.

Data exclusion
Participants were excluded in case they did not use the confidence zone properly (i.e., no significant difference in the zone of confidence between conditions) or made an orienting error superior to 45º in more than half trials (four participants in Experiment 1, two participants in Experiment 2). Finally, response time distributions were inspected to ensure the good quality of the collected data, however no participants were excluded based on this criterion.
When a trial was flagged as missed (see *procedure*), the whole set of trials was removed from further analysis to avoid having “random guess” trials included in the metacognitive assessment of their performance. Moreover, trials were excluded if participants took an unusually long time to respond (trial duration above 10 seconds). After applying these criteria, a total of 130 trials (out of 13 328, all participants included) were excluded from Experiment 1 and 435 (out of 20 532, all participants included) from Experiment 2.

Exploratory analyses
We asked participants not only for a point estimate of their global performance via the report of their mean response orientation, but also for a dispersion around it. Thus, as an exploratory analysis, we also examined the relationship between these two metacognitive reports by adding *Reported_Mean* as a factor to the model M2:

M2_Extended: Reported_Dispersion ~ Reported_Mean * poly($MeanError_{Set}$,2) * SD_{set} * Orientation + (Reported_Mean + poly($MeanError_{Set}$,2) + SD_{set} + Orientation| participant)

Results

Experiment 1

Concerning first-order performance, the mean absolute error (+/- SD) on a single matching orientation was 7.54° (+/- 3.42) and we did not observe clear evidence of an oblique effect, i.e. smaller errors for cardinal than oblique orientations (effect of orientation: M = 0.22, 95% CI = [0.05, 0.38], BF10 = 1.14).

We now turn to analyses of the *Reported_Mean*, which reflects the magnitude of the difference between the actual mean target orientation of the Gabor patch in a set and the mean response orientation reported by the participant. Table 1 gathers the full outcome table for the Bayesian mixed-effects regression M1. This regression revealed a main effect of endogenous uncertainty: *Reported_Mean* increased with participants' response variability, (SD_{Set}: posterior distribution of the model estimate mean: M = 0.22, 95% CI = [0.16, 0.28], BF10 > 1000). However, this effect was reduced when both sources of noise increased: the increase of *Reported_Mean* with SD_{Set} was weaker for oblique than for cardinal stimuli (M = -0.15, 95% CI = [-0.22, -0.08], BF10 > 1000, Fig. 2.A).

The model also revealed moderate evidence for a triple interaction between $MeanError_{Set}^2$, SD_{set_z}, and *Orientation* (M = -2.33, 95% CI = [-4.62, -0.06], BF10 = 3.73, Fig. 2.B and C). This interaction reflects that *Reported_Mean* increased in sets where participants were less accurate on average ($MeanError_{Set}^2$ x SD_{set_z}: M = -3.63, 95% CI = [-5.51, -1.78], BF10 > 1000), but this interaction between accuracy and endogenous uncertainty decreased in conditions of higher exogenous uncertainty (i.e., for oblique orientations). In other words, both endogenous and exogenous sources of uncertainty decreased the capacity of participants to monitor the result of their actual performance in a set of trials.

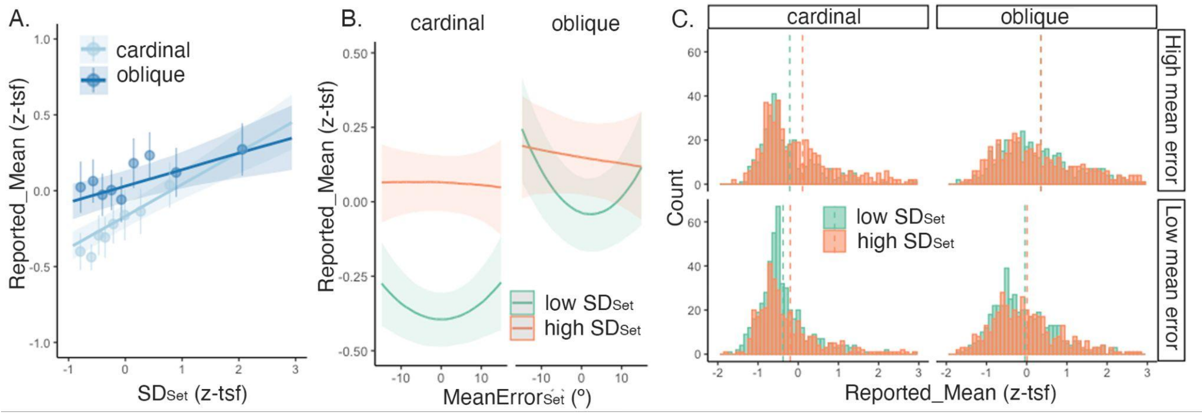

Figure 2: Variation of the error between the actual mean target orientation and the reported mean response orientation (*Reported_Mean*) reported by participants after a set of trials depending on first-order performance and stimulus orientation. A. The magnitude of *Reported_Mean* changed with the participant's variability on a set (SD_{Set}) in a steeper manner for cardinal (low sensory uncertainty in light blue) than for oblique stimuli (high sensory uncertainty in dark blue). Although the model took continuous variables as input, for illustrative purposes we plotted dots and error bars that represent the mean ± 95% confidence interval over participants after grouping the values SD_{Set} in deciles. **B.** Model prediction reflecting the triple interaction between $MeanError_{Set}^2$, SD_{set_z}, and *Orientation*. Lines and shaded areas correspond to the regression model predictions and 95% confidence interval. **C.** The histograms represent the distribution of *Reported_Mean* values and illustrate the triple interaction. For illustrative purposes we divided the sets into two levels of absolute values of $MeanError_{Set}$ (high on the upper panel, low in the lower panel) and two levels of SD_{Set} (green/orange: low/high endogenous uncertainty, i.e., less/more variable responses within the set). We observe more low values of *Reported_Mean* for sets corresponding to more accurate performance (low mean error) and low exogenous uncertainty, especially for low endogenous uncertainty sets (left lower panel). A degradation in accuracy of the first order performance (upper panels) and/or an increase in exogenous uncertainty (right panels) lead to higher *Reported_Mean* and reduces the difference between low and high endogenous sets.

Table 1: Full outcome table for the M1 and M2 Bayesian mixed-effects regressions to assess the contribution of *$MeanError_{Set}$*, *SD_{set}*, and *orientation* to *Reported_Mean* and *Reported_Dispersion*. _z indicates that the variable has been z transformed. ^2 indicates that the quadratic component of the corresponding variable is considered.

Reported_Mean_z (M1)	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**BF10**
Intercept	-0.15	0.05	-0.25	-0.06	
$MeanError_{Set}$	-0.96	1.71	-4.33	2.43	0.96
$MeanError_{Set}$^2	3.21	2.02	-0.62	7.13	3.45
SD_{set}_z	0.22	0.03	0.16	0.28	1000
Orientation	0.23	0.11	0.03	0.45	0.7
$MeanError_{Set}$ x SD_{set}_z	0.74	0.83	-0.87	2.34	0.62
$MeanError_{Set}$ ^2 x SD_{set}_z	-3.63	0.92	-5.46	-1.84	1000
$MeanError_{Set}$ x Orientation	-2.05	1.47	-4.93	0.85	2.08
$MeanError_{Set}$ ^2 x Orientation	2.99	1.67	-0.37	6.26	3.85
SD_{set} _z x Orientation	-0.15	0.04	-0.22	-0.08	1000
$MeanError_{Set}$ x SD_{set} _z x Orientation	1.18	1.02	-0.86	3.22	0.98
$MeanError_{Set}$ ^2 x SD_{set} _z x Orientation	-2.33	1.19	-4.62	-0.06	3.7
Reported_Dispersion_z (M2)	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**BF10**
Intercept	-0.11	0.04	-0.18	-0.04	
$MeanError_{Set}$	-0.11	1.23	-2.49	2.28	0.62
$MeanError_{Set}$^2	1.92	1.50	-1.11	4.68	2.04
SD_{set}_z	0.28	0.04	0.20	0.36	1000
Orientation	0.21	0.07	0.08	0.35	3.15
$MeanError_{Set}$ x SD_{set}_z	-0.00	0.80	-1.56	1.57	0.39
$MeanError_{Set}$ ^2 x SD_{set}_z	2.21	0.80	0.60	3.81	15.81
$MeanError_{Set}$ x Orientation	0.82	1.37	-1.84	3.59	0.83
$MeanError_{Set}$ ^2 x Orientation	0.02	1.56	-2.93	3.01	0.76
SD_{set} _z x Orientation	-0.14	0.04	-0.21	-0.07	20.16
$MeanError_{Set}$ x SD_{set} _z x Orientation	-0.47	1.05	-2.54	1.56	0.57
$MeanError_{Set}$ ^2 x SD_{set} _z x Orientation	-1.04	1.13	-3.31	1.16	0.88

The same Bayesian mixed-effects regression applied to the confidence zone revealed the same pattern of results as we found for the reported mean (Table 1). Participants reported a larger dispersion of their responses (i.e., reported a larger confidence zone) in noisier conditions, whether the noise corresponded to high endogenous uncertainty (*SD_{Set}*: M = 0.28, 95% CI = [0.20, 0.36], BF10 > 1000) or exogenous uncertainty (cardinal versus oblique: M = 0.21, 95% CI = [0.08, 0.35], BF10 = 3.15). However, once again this effect was reduced when both sources of noise increased: the increase of *Reported_Dispersion* with *SD_{Set}* was weaker for oblique than for cardinal stimuli (M = -0.14, 95% CI = [-0.21, -0.06], BF10 = 20.16, Fig. 3.A).

Moreover, the increase of *Reported_Dispersion* with increasing endogenous uncertainty was amplified when the magnitude of the participant's mean error also increased ($MeanError_{Set}^2$ x SD_{set_z}: M = 2.21, 95% CI = [0.68, 3.78], BF10 = 15.81). In other words, *Reported_Dispersion* reflected the degradation of both aspects of the first-order performance ($MeanError_{Set}$ and SD_{Set}, Fig. 3.B and C).

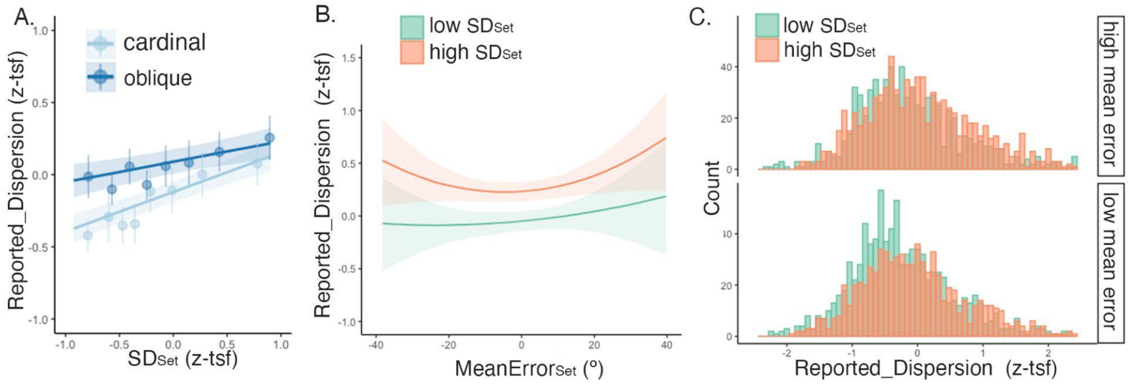

Figure 3: Variation of the angle of the confidence zone (*Reported_Dispersion*) depending on first-order performance and stimulus orientation. A. *Reported_Dispersion* increased with the variability of the participants' responses on a set (SD_{Set}), in a steeper manner for cardinal (low sensory uncertainty in light blue) than for oblique stimuli (high sensory uncertainty in dark blue). Although the model took continuous variables as input, for illustrative purposes we plotted dots and error bars that represent the mean ± 95% confidence interval over participants after grouping the values SD_{set} in deciles. **B.** Model predictions corresponding to the interaction between $MeanError_{Set}^2$ and SD_{set_z}. Continuous lines and shaded areas correspond to the regression model predictions and 95% confidence interval. **C.** Histograms represent the distribution of *Reported_Dispersion* values and illustrate the double interaction, revealing that participants reported smaller *Reported_Dispersion* when $MeanError_{Set}$ and SD_{Set} were both smaller (lower panel). For illustrative purposes we divided the sets into two levels of absolute values of $MeanError_{Set}$ (high on the upper panel, low in the lower panel) and two levels of SD_{Set} (green/orange: low/high endogenous uncertainty, i.e., less/more variable responses within the set).

Finally, we examined the relation between *Reported_Dispersion* and *Reported_Mean* (M2_extended, Table 2). This model indicates that *Reported_Dispersion* increased with *Reported_Mean* (M = 0.22, 95% CI = [0.15, 0.29], BF > 1000), i.e., participants reported a larger dispersion of their performance when their mean response orientation was less accurate. This correlation between the two metacognitive reports tended to be more pronounced under low exogenous uncertainty (*Reported_Mean* x *Orientation:* M= -0.11, 95% CI = [-0.19, -0.04], BF10 = 1.43, Fig. 4. A).

We also observed another interaction between *Reported_Mean* and $MeanError_{Set}^2$ (M= -2.10, 95% CI = [-3.86, -0.44], BF10 > 5;), indicating that relationship between *Reported_Dispersion* and $MeanError_{Set}$ was weaker when *Reported_Mean* decreased (i.e., when the mean response orientation was more accurate, Fig. 4. B-C). Thus, the perceived mean response and the actual performance accuracy interacted in influencing how precise the participants perceived their performance to be.

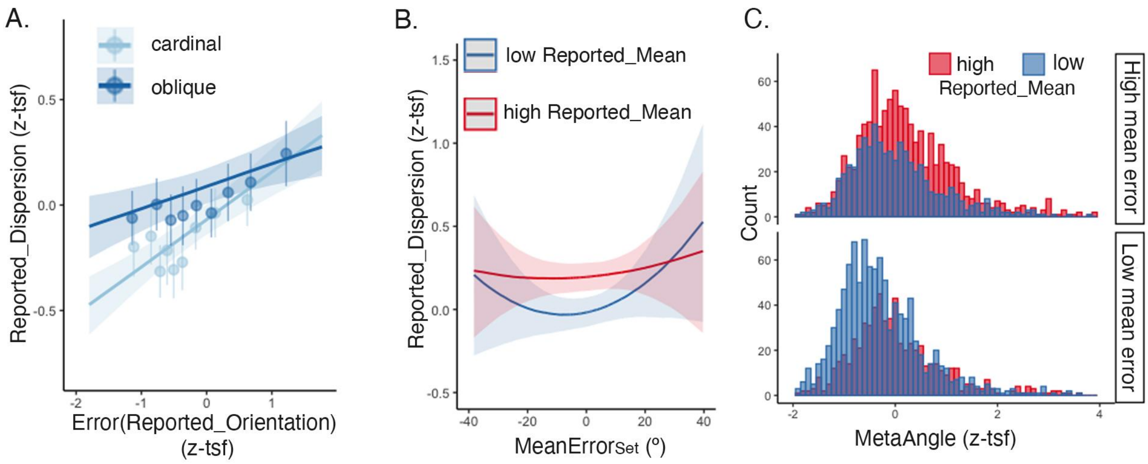

Figure 4: Variation of the angle of the confidence zone (Reported_Dispersion) depending on the error between the actual mean target orientation and the reported mean response orientation, first-order performance and stimulus orientation. A. *Reported_Dispersion* increased when the accuracy of the reported mean response orientation decreased (higher *Reported_Mean*), in a steeper manner for cardinal (low sensory uncertainty in light blue) than for oblique stimuli (high sensory uncertainty in dark blue). **B.** Model (M2_extended) predictions corresponding to the interaction between $MeanError_{Set}^2$ and *Reported_Mean*, **C**. The distribution of *Reported_Dispersion* values illustrates this interaction revealing that participants reported a more precise performance (smaller *Reported_Dispersion*) when both the accuracy of the actual and the accuracy of the reported performance were better (smaller $MeanError_{Set}$ and *Reported_Mean*, lower panel, blue distribution). *Continuous lines and shaded areas correspond to the regression model predictions and 95% confidence interval. Although the model took continuous variables as input, for illustrative purposes we plotted dots and error bars that represent the mean ± 95% confidence interval over participants after grouping the values of the independent variables in deciles in the panel A. Similarly, histograms in panels C and E were plotted after dividing the sets into two levels of each independent variable.*

Table 2: Full outcome table for the M2_extended regression to assess the link between Reported_Mean and Reported_Dispersion. _z indicates that the variable has been z transformed. ^2 indicates that the quadratic component of the variable is considered**.**

Reported_Dispersion_z	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**BF**
Intercept	-0.06	0.04	-0.13	0.01	
Reported_Mean_z	0.22	0.04	0.15	0.29	1000
MeanErrorSet	-0.24	1.30	-2.76	2.27	0.63
MeanErrorSet^2	1.66	1.66	-1.58	4.84	1.39
SDset_z	0.24	0.04	0.16	0.32	1000
Orientation	0.16	0.07	0.03	0.29	0.62
Reported_Mean_z x MeanErrorSet	-0.12	0.88	-1.86	1.64	0.43
Reported_Mean_z x MeanErrorSet^2	-2.11	0.84	-3.86	-0.44	8.35
Reported_Mean_z x SDset_z	-0.02	0.02	-0.07	0.01	0.02
MeanErrorSet x SDset_z	0.78	0.93	-1.07	2.57	0.70
MeanErrorSet^2 x SDset_z	2.60	1.17	0.23	4.86	6.67
Reported_Mean_z x Orientation	-0.11	0.04	-0.19	-0.04	1.43
MeanErrorSet x Orientation	0.99	1.40	-1.77	3.67	0.91
MeanErrorSet^2 x Orientation	0.20	1.75	-3.39	3.58	0.81
SDset_z x Orientation	-0.09	0.04	-0.17	-0.02	0.39
Reported_Mean_z x MeanErrorSet x SDset_z	-1.04	0.50	-2.06	-0.09	2.04
Reported_Mean_z x MeanErrorSet^2 x SDset_z	-0.09	0.46	-1.00	0.85	0.22
Reported_Mean_z x MeanErrorSet x Orientation	-0.14	1.01	-2.15	1.84	0.48
Reported_Mean_z x MeanErrorSet^2 x Orientation	0.96	0.92	-0.84	2.77	0.68
Reported_Mean_z x SDset_z x Orientation	-0.02	0.03	-0.07	0.04	0.02
MeanErrorSet x SDset_z x Orientation	-0.68	1.10	-2.78	1.50	0.71
MeanErrorSet^2 x SDset_z x Orientation	-2.12	1.34	-4.82	0.52	2.13
Reported_Mean_z x MeanErrorSet x SDset_z x Orientation	-0.18	0.70	-1.54	1.24	0.37
Reported_Mean_z x MeanErrorSet^2 x SDset_z x Orientation	1.20	0.65	-0.09	2.44	1.56

Experiment 2

Overall, Experiment 1 revealed that observers considered their accuracy as well as both endogenous (response variability) and exogenous (sensory uncertainty) sources of uncertainty when monitoring their own performance in an orientation reproduction task. In Experiment 2, we sought to replicate these results and examine another source of uncertainty modulation via exogenous attentional cueing. To manipulate participants' attentional focus, we presented two instead of one gabor patch on each trial, and preceded them with a valid or an invalid exogenous cue. Within a set of four trials, all cues (*full* set), three cues (*valid* set), two cues (*neutral* set), or only one cue could be valid (invalid set), resulting in four different conditions of cueing at the level of a set. Because the change in the participants' attentional focus came from an experimental, stimulus-related, manipulation, we considered that these conditions resulted in different levels of exogenous uncertainty.

Mean (+/- SD) first-order performance pooled across all sets, was 7.43° +/- 4.14 Participants made larger errors on single trials that were invalidly cued, as compared to validly cued trials (M = -0.22, 95% CI = [-0.29, -0.14], BF10 > 1000). This confirms that our attentional manipulation was effective.

Regarding global performance monitoring, we replicated the effect of $MeanError_{Set}$ and SD_{Set} on both *Reported_Mean* and *Reported_Dispersion* (Fig. 5, Table 3). *Reported_Mean* increased with SD_{Set} (SD_{Set}: M = 0.18, 95% CI = [0.09, 0.27], BF10 > 5) and this effect was amplified by the magnitude of $MeanError_{Set}$ ($MeanError_{Set}^2$ x SD_{Set}: M = 2.42, 95% CI = [0.64, 4.28], BF10 > 5). Similarly, *Reported_Dispersion* increased with SD_{Set} (M = 0.23, 95% CI = [0.15, 0.32], BF10 > 5) and this effect was also modulated by the magnitude of $MeanError_{Set}$ ($MeanError_{Set}^2$x SD_{Set}: M = -4.28, 95% CI = [-6.19, -2.44], BF10 > 5). Once again, these results suggest that participants correctly accounted for changes in the performance accuracy and endogenous uncertainty when monitoring different aspects of their performance over a series of trials.

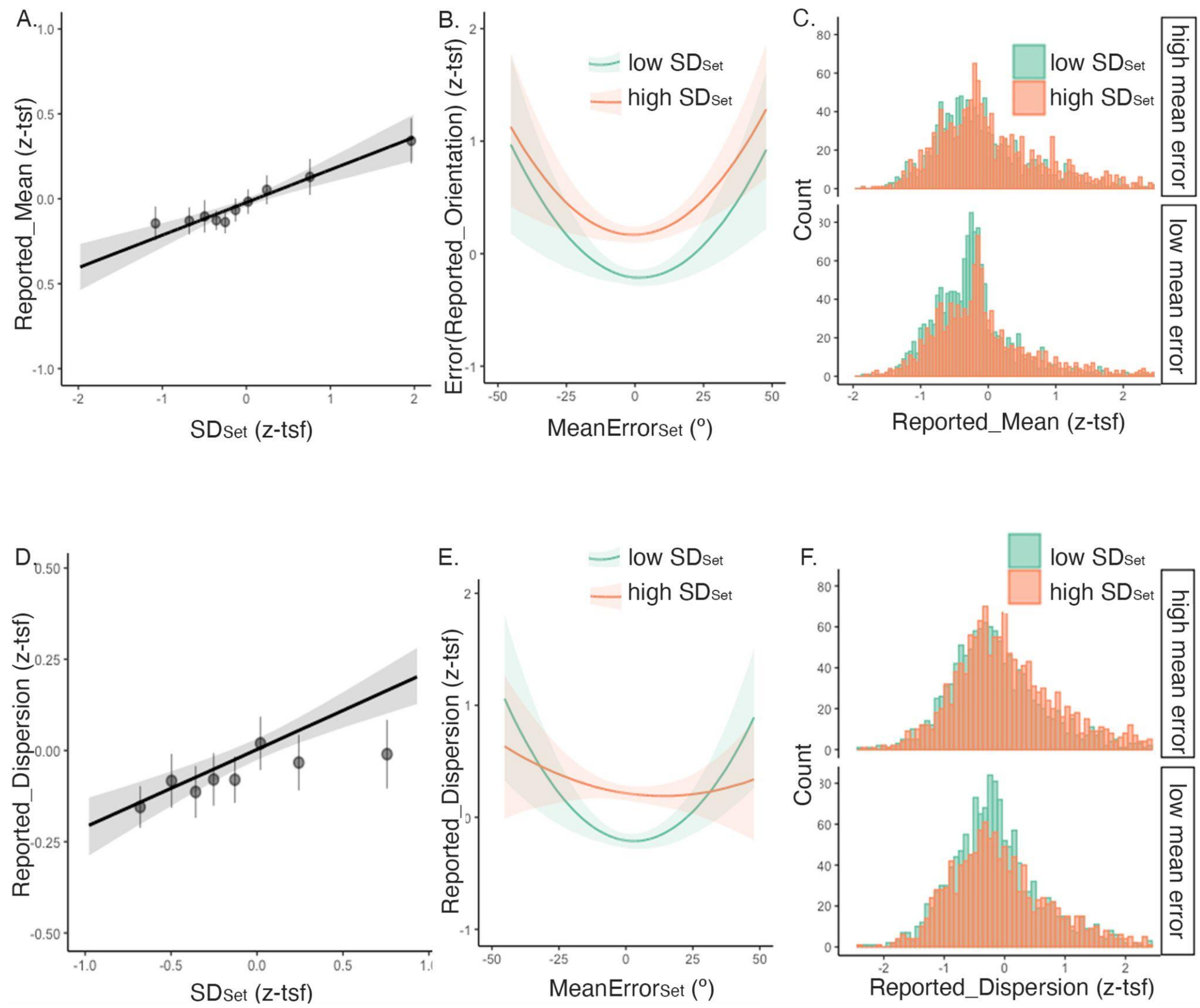

Figure 5: Variations of the error between the actual mean target orientation and the mean target orientation (*Reported_Mean*, A, B, C) and the confidence zone (*Reported_Dispersion*, D, E, F) depending on first-order performance.

A. Reported_Mean increased with the variability of the participants' responses on a set (SD_{Set}). **B.** Model predictions corresponding to the interaction between $MeanError_{Set}^2$ and SD_{set_z}. **C**.The distribution of Reported_Mean values illustrates this double interaction revealing that participants reported smaller Reported_Mean when they were more accurate (smaller absolute $MeanError_{Set}$) and more precise (smaller SD_{Set}) in the orientation reproduction task. Thus, we observe more low values of Reported_Mean for low mean error (lower panel) and especially for sets with low endogenous uncertainty (in green). **D**. Similarly, Reported_Dispersion increased with the variability of the participants' responses on a set (SD_{Set}). **E.** Model predictions corresponding to the interaction between $MeanError_{Set}^2$ and SD_{set_z}. **F**. The distribution of Reported_Dispersion values illustrates this double interaction revealing that participants felt that the responses were more precise (smaller Reported_Dispersion) when their responses were more accurate and precise (smaller $MeanError_{Set}$ and SD_{Set}, respectively). *Continuous lines and shaded areas correspond to the regression model predictions and 95% confidence interval. Although the model took continuous variables as input, for illustrative purposes we plotted dots and error bars that represent the mean ± 95% confidence interval over participants after grouping the values of the independent variables in deciles in panels A and D. Similarly, histograms in panels C and F were plotted after dividing the sets into two levels of each independent variable.*

The cueing condition modulated the effect of the participants' first-order performance on the report of their mean response orientation: the relationship between *Reported_Mean* and $MeanError_{Set}$ changed with the proportion of invalid cues in a set ($MeanError_{Set}$^2 x *Cueing*: M = 4.70, 95% CI = [2.01, 7.45], BF10 > 5; $MeanError_{Set}$^2 x SD_{set} x *Cueing*: M = -2.40, 95% CI = [-3.66, -1.13], BF10 > 5, Fig. 6.). *Reported_Mean* increased with lower set accuracy (increasing of the absolute value of $MeanError_{Set}$) especially under low attentional focus; and this effect of reduced attention towards the target stimulus was less pronounced with increasing endogenous uncertainty. This last result suggests once again a difficulty in accounting for both sources of uncertainty (endogenous and exogenous) when monitoring the performance over several trials.

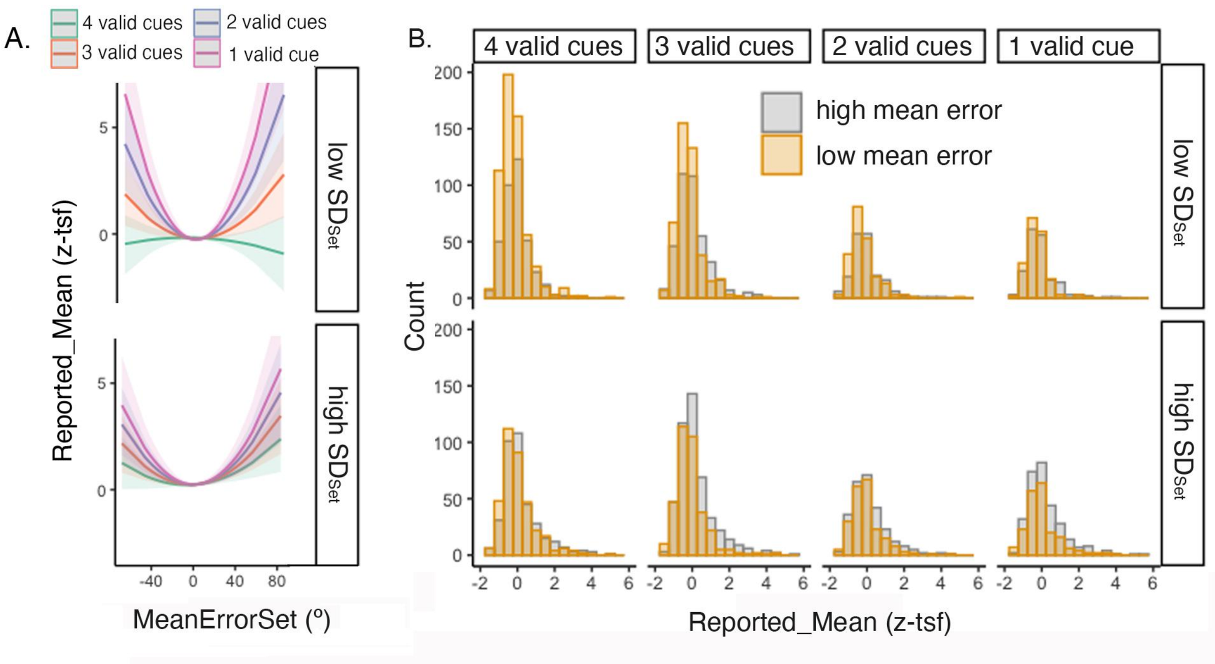

Figure 6: Variations of the angle of Reported_Mean depending on the participant's performance and the cueing condition. For illustrative purposes we divided the sets into two levels of absolute values of $MeanError_{Set}$ (grey: high mean error, yellow: low mean error on the set) and two levels of SD_{Set} (higher panel: low endogenous uncertainty, i.e., less variable responses with a set, lower panel: high endogenous uncertainty, i.e., more variable responses with a set). A. Regression model predicted that *Reported_Mean* increased when the actual performance on the set worsened (increasing $MeanError_{Set}$^2) especially when participants were distracted by invalid cues. This effect of cueing was more pronounced in low endogenous uncertainty sets (for less variable responses, upper panel). B. In agreement with these model predictions, data distribution revealed a higher frequency of low values of *Reported_Mean* when the first order performance was more accurate (low mean error in yellow) when the exogenous and endogenous uncertainty were the lowest (left upper panel). Values of *Reported_Mean* increased when the first order performance worsened and the ability to correctly monitor this change in performance was reduced when both sources of uncertainty are combined.

Table 3: Full outcome table for the M1 and M2 regressions to assess the contribution of MeanErrorSet, SD_{set}, and Cueing to Reported_Mean and Reported_Dispersion. _z means the variable has been z transformed. ^2 means the quadratic component of the corresponding variable is considered.

Reported_Mean_z (M1)	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**BF10**
Intercept	-0.00	0.02	-0.04	0.04	
MeanErrorSet	0.58	1.57	-2.59	3.70	0.83
MeanErrorSet^2	1.21	1.88	-2.53	4.89	1.17
SDset_z	0.18	0.05	0.09	0.27	16.5
Cueing	0.00	0.01	-0.03	0.03	0.01
MeanErrorSet x SDset_z	1.48	0.74	0.04	2.97	2.92
MeanErrorSet^2 x SDset_z	2.42	0.93	0.64	4.28	20.85
MeanErrorSet x Cueing	0.62	0.94	-1.22	2.44	0.59
MeanErrorSet^2 x Cueing	4.70	1.39	2.01	7.45	292.74
SDset_z x Cueing	0.01	0.02	-0.02	0.04	0.01
MeanErrorSet x SDset_z x Cueing	-0.89	0.49	-1.87	0.07	1.28
MeanErrorSet^2 x SDset_z x Cueing	-2.40	0.64	-3.66	-1.13	1000
Reported_Dispersion_z (M2)	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**BF10**
Intercept	-0.01	0.02	-0.05	0.03	
MeanErrorSet	-0.19	1.60	-3.37	3.00	0.79
MeanErrorSet^2	1.62	1.83	-1.92	5.12	1.3
SDset_z	0.23	0.04	0.15	0.32	1000
Cueing	0.02	0.01	-0.01	0.04	0.02
MeanErrorSet x SDset_z	-0.27	0.75	-1.74	1.22	0.39
MeanErrorSet ^2 x SDset_z	-4.28	0.96	-6.19	-2.44	1000
MeanErrorSet x Cueing	-0.16	0.94	-2.00	1.65	0.48
MeanErrorSet ^2 x Cueing	2.52	1.35	-0.11	5.15	3.81
SDset _z x Cueing	-0.02	0.02	-0.05	0.01	0.02
MeanErrorSet x SDset _z x Cueing	-0.38	0.49	-1.35	0.60	0.34
MeanErrorSet ^2 x SDset _z x Cueing	1.20	0.65	-0.08	2.45	1.99

Finally, we examined the relation between *Reported_Dispersion* and *Reported_Mean* (M2_extended, Table 4, Figure S4). The regression revealed one triple and one quadruple interaction involving *Reported_Mean* (*Reported_Mean* x $MeanError_{Set}$^2 x SD_{Set}; *Reported_Mean* x $MeanError_{Set}$^2 x SD_{Set} x *Cueing*; Figure 7), showing that the adjustment of *Reported_Dispersion* to *Reported_Mean* decreased when the first-order performance accuracy decreased (higher $MeanError_{Set}$^2) and/or when the uncertainty increased (higher SD_{Set} or decreased attentional focus).

<u>Table 4</u>: Full outcome table for the M2_extended regression to assess the link between Reported_Mean and Reported_Dispersion. _z means the variable has been z transformed. ^2 means the quadratic component of the variable is considered.

Reported_Dispersion_z (M2_extended)	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**BF**
Intercept	-0.02	0.02	-0.06	0.02	
Reported_Mean_z	0.11	0.04	0.03	0.19	0.73
MeanErrorSet	-1.10	1.57	-4.16	1.98	1.01
MeanErrorSet^2	1.00	1.79	-2.44	4.58	1.00
SDset_z	0.18	0.04	0.10	0.27	288.90
Cueing	0.02	0.01	-0.01	0.05	0.02
Reported_Mean_z x MeanErrorSet	0.75	1.23	-1.62	3.20	0.74
Reported_Mean_z x MeanErrorSet^2	-1.22	1.20	-3.59	1.16	1.00
Reported_Mean_z x SDset_z	0.03	0.02	-0.01	0.06	0.03
MeanErrorSet x SDset_z	0.72	0.80	-0.89	2.29	0.57
MeanErrorSet^2 x SDset_z	0.66	1.09	-1.49	2.84	0.64
Reported_Mean_z x Cueing	0.00	0.01	-0.03	0.03	0.01
MeanErrorSet x Cueing	-0.10	0.93	-1.91	1.65	0.47
MeanErrorSet^2 x Cueing	0.31	1.39	-2.43	3.00	0.71
SDset_z x Cueing	-0.01	0.02	-0.04	0.02	0.01
Reported_Mean_z x MeanErrorSet x SDset_z	0.14	0.42	-0.69	0.97	0.23
Reported_Mean_z x MeanErrorSet^2 x SDset_z	-1.78	0.42	-2.60	-0.94	1000
Reported_Mean_z x MeanErrorSet x Cueing	-0.01	0.80	-1.62	1.52	0.39
Reported_Mean_z x MeanErrorSet^2 x Cueing	0.51	0.81	-1.10	2.13	0.49
Reported_Mean_z x SDset_z x Cueing	-0.02	0.01	-0.04	0.00	0.02
MeanErrorSet x SDset_z x Cueing	-0.53	0.54	-1.58	0.54	0.44
MeanErrorSet^2 x SDset_z x Cueing	-0.56	0.74	-2.07	0.85	0.53
Reported_Mean_z x MeanErrorSet x SDset_z x Cueing	-0.38	0.39	-1.14	0.37	0.31
Reported_Mean_z x MeanErrorSet^2 x SDset_z x Cueing	1.09	0.40	0.31	1.86	9.16

Discussion

We investigated the contribution of first-order accuracy and uncertainty to global performance monitoring. In a novel task, we asked participants to provide two kinds of judgments about their own performance. After a set of four trials of an orientation matching task, participants first reported their perceived mean response, and then a region around that mean corresponding to their estimation of their responses' dispersion. We could assess how first-order performance, endogenous uncertainty, and exogenous uncertainty impacted performance monitoring by comparing the difference between mean performance and actual orientation target (*Reported_Mean*) as well the estimated dispersion of performance (*Reported_Dispersion*). Endogenous uncertainty was estimated from the variability of first-order performance across the four repetitions of the orientation task. Exogenous uncertainty was operationalized as the amount of sensory noise affecting the stimuli upon which the global performance monitoring was made (i.e., stimulus-related oblique effect in Experiment 1) and was also manipulated by changing the attentional focus manipulated through exogenous cueing (attention-related uncertainty in Experiment 2).

In these two experiments, we showed that participants arguably used the average and dispersion of their first-order performance to monitor global performance. The calibration of metacognitive judgments to first-order performance was better when participants' responses were less variable i.e., when the endogenous uncertainty was lower. Similarly, other sources of uncertainty also modulated the calibration between global metacognitive judgments and first-order performance. First, higher levels of exogenous uncertainty (oblique vs cardinal orientations) led to a weaker relationship between metacognitive reports and first-order responses variability. Second, when manipulating the participants' attentional focus, increased attentional-related uncertainty resulting from a higher number of invalid cues also impacted global metacognitive judgements: participants correctly accounted for changes in task difficulty with the proportion of invalid cueing when monitoring their global mean performance, but this effect was less pronounced for sets during which participants' responses were more variable, i.e. when higher endogenous uncertainty was combined to the increase in exogenous uncertainty. Two key points can be made from these results: first, people appeared to reliably estimate the mean and variability of their performance and use it together with exogenous uncertainty to inform their global performance monitoring. Second, this capacity decreases in the presence of both endogenous and exogenous uncertainty.

The importance of uncertainty, regardless of its origin, as a contributing factor to local metacognitive judgments has been highlighted before (Atiya et al. 2021; Denison et al. 2018; Geurts et al. 2022; Honig, Ma, and Fougnie 2020; Mole et al. 2018; Rahnev 2021). Our results suggest that uncertainty contributes not only to local metacognitive judgments, i.e., judgment of an isolated performance, but also to global performance monitoring, i.e., over a series of performances and events. If individuals have the ability to track their performance while performing a visuomotor task (Locke et al. 2020) and to extract summary statistics like average and variance from a group of stimuli (de Gardelle and Summerfield 2011; Ji and Hayward 2021), our study highlights that they are also able to use such statistics about their own performance on a series of trials, in addition to exogenous sensory uncertainty, to evaluate this performance. Our results also indicate that attention-related uncertainty is also taken into account to form global judgements, even though the mechanisms connecting attention to variation in external uncertainty are still not fully defined (Carrasco 2011). For local metacognitive judgments, the contribution of attention to confidence has only recently been highlighted with an experimental paradigm where participants were asked to categorize visual stimuli in two embedded categories (Denison et al. 2018). Using a spatial cueing method similar to ours, this computational study revealed a Bayesian-like link between attention and confidence. Our results suggest that this finding about local metacognitive judgments, which has since then been replicated (Recht, Mamassian, and de Gardelle 2022), holds true for global metacognitive judgments as well.

Interestingly, when exogenous uncertainty was experimentally increased, for example in sets of trials including noisy (oblique orientation) or unattended stimuli (invalid cues), the increase in variability in first-order responses led to a reduced calibration of global metacognitive judgements. This effect seems to be an example of metacognitive inefficiency, i.e., a failure to monitor behavioral accuracy. Identifying and understanding the emergence of metacognitive inefficiency is an important goal of the research in metacognition (Shekhar and Rahnev 2021; Rahnev et al., 2022). Our study reveals a potential source of failure at the level of global performance monitoring; however, the cause of this failure remains an open question: Do participants show a reduced capacity to assess and use uncertainty in their global metacognitive judgments when this uncertainty becomes too important (similar to a ceiling effect where uncertainty cannot be encoded anymore); or does the limited ability come from a difficulty in combining uncertainty emerging from different sources (endogenous versus exogenous source of uncertainty)? We argue that approaches like ours are relevant to address the question of suboptimality and inefficiency at the metacognitive level.

As already mentioned, the present study shows a pattern of results for global performance monitoring that is similar to a recent study highlighting how uncertainty shaped local confidence according to Bayesian principles regardless of its exogenous or endogenous origin (Geurts et al. 2022). The question remains about the mechanisms governing this contribution of exogenous and endogenous uncertainty to global metacognitive judgments: First-order performance could influence directly the global metacognitive estimates, or indirectly by influencing only local metacognitive estimates that are then combined to form global estimates. Answering this question would require collecting both local and global metacognitive judgements in the same continuous task, which no one has done so far. As mentioned before, continuous tasks are ideally suited to finely quantify the formation of local and global metacognitive judgments as direct functions of exogenous and endogenous uncertainty. To the best of our knowledge, only two studies have directly examined how global judgements derive from local confidence estimates (Lee et al. 2021; Rouault et al. 2019). Both studies used binary decisions as metacognitive reports, i.e., participants had to choose between two sets of trials of the same task (Lee et al. 2021) or two tasks (Rouault et al. 2019) the one for which they performed better. In both studies, global metacognitive judgments appeared to integrate information across multiple perceptual decisions and to be formed from local confidence reports. However, these studies disagree on the question of an equal contribution of each local estimate to global metacognition (e.g., recency effect found in Lee et al. but not in Rouault et al.) and on which components of first-order performance contribute to global metacognitive judgements (accuracy, response time, number of events to be considered).

We mentioned that the two different global dependent variables collected during our experiment captured two different aspects of performance monitoring. Our results revealed that the two variables were not independent. The ability of the participant to track the overall accuracy of the performance across four trials, reflected by the difference between the actual target on a set and the reported mean response orientation (*Reported_Mean*) influenced how participants monitored their response's dispersion (*Reported_Dispersion*). First, the global performance accuracy estimate changed how well the participants accounted for their actual accuracy (first-order performance accuracy) when reporting their global performance precision. Second, the magnitude of the correlation between these two types of metacognitive reports also depended on the level of uncertainty, suggesting once again a weakening of the calibration mechanisms for global performance precision monitoring when both endogenous and exogenous uncertainty increased. It is tempting (but unjustified, as we argue below) to consider that the dispersion estimate reported by the participants constitutes a measure of metacognitive uncertainty around the global performance mean estimate. Together, our two measures could then reflect the mean and dispersion distribution of the global performance monitoring estimate. However, our results suggested that performance dispersion monitoring was still shaped by the actual first-order performance, thus does not seem to solely reflect metacognitive uncertainty (i.e., another type of noise corrupting the estimate formed about the global performance, independently of first-order processes). Future work is needed to probe further the mechanisms linking different aspects of performance monitoring and metacognitive judgments in general.

Finally, in regards to the connexion between local and global metacognition, we would like to emphasize a key interest of our approach: by asking participants to monitor their performance across four trials on a continuous orientation task, we focus on a metacognitive process more global than the classical metacognitive judgment reported about an isolated event or task; yet these collected metacognitive estimates are still anchored in a perceptual task and are to be considered at a lower conceptual level than global beliefs about self-performance (i.e., information-based metacognition, Koriat 2007). Thus, our study opens a new window of observation on the relationship between local and global metacognitive judgments. Moreover, tasks such as the present one are also useful to finely quantify metacognition with greater ecological validity, since we are more likely to evaluate our global performance after several repetitions of the same task in daily life rather than making a judgment on a visual scale following a single forced choice. This task could therefore help to capture global metacognitive deficits in neurological or psychiatric disorders (Seow et al. 2021), which is already the subject of intense research at present via local judgements (Hoven et al. 2019, 2022; Rouy et al. 2021). Uniquely, we also asked participants not only for a point estimate of their performance, but also for a range around it, stepping away from the classical reports of confidence. The need for such new paradigms aiming to expand both the scale and the scope of metacognitive reports has been highlighted as a crucial development for future metacognitive neuroscience (Katyal and Fleming 2023).

Acknowledgements

EF is supported by a Freigeist Fellowship from the VolkswagenStiftung (Grant number 91620). NF has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 803122).

Data availability

Data and custom scripts are available on https://gitlab.com/nfaivre/meta_angle_public.

References

Appelle, Stuart. 1972. “Perception and Discrimination as a Function of Stimulus Orientation: The ‘Oblique Effect’ in Man and Animals.” *Psychological Bulletin* 78:266–78. doi: 10.1037/h0033117.

Atiya, Nadim A. A., Quentin J. M. Huys, Raymond J. Dolan, and Stephen M. Fleming. 2021. “Explaining Distortions in Metacognition with an Attractor Network Model of Decision Uncertainty.” *PLoS Computational Biology* 17(7):e1009201. doi: 10.1371/journal.pcbi.1009201.

Balsdon, Tarryn, Pascal Mamassian, and Valentin Wyart. 2021. “Separable Neural Signatures of Confidence during Perceptual Decisions” edited by R. G. O’Connell, M. J. Frank, and R. G. O’Connell. *ELife* 10:e68491. doi: 10.7554/eLife.68491.

Bandura, Albert. 1977. “Self-Efficacy: Toward a Unifying Theory of Behavioral Change.” *Psychological Review* 84(2):191–215. doi: 10.1037/0033-295X.84.2.191.

Barbosa, Leonardo S., and Sid Kouider. 2018. “Prior Expectation Modulates Repetition Suppression without Perceptual Awareness.” *Scientific Reports* 8(1):5055. doi: 10.1038/s41598-018-23467-3.

van den Berg, Ronald, Kavitha Anandalingam, Ariel Zylberberg, Roozbeh Kiani, Michael N. Shadlen, and Daniel M. Wolpert. 2016. “A Common Mechanism Underlies Changes of Mind about Decisions and Confidence” edited by T. E. Behrens. *ELife* 5:e12192. doi: 10.7554/eLife.12192.

Carrasco, Marisa. 2011. “Visual Attention: The Past 25 Years.” *Vision Research* 51(13):1484–1525. doi: 10.1016/j.visres.2011.04.012.

Constant, Marika, Michael Pereira, Nathan Faivre, and Elisa Filevich. 2022. *Prior Information Differentially Affects Discrimination Decisions and Subjective Confidence Reports*. *preprint*. Neuroscience. doi: 10.1101/2022.10.26.513829.

Denison, Rachel N., William T. Adler, Marisa Carrasco, and Wei Ji Ma. 2018. “Humans Incorporate Attention-Dependent Uncertainty into Perceptual Decisions and Confidence.” *Proceedings of the National Academy of Sciences* 115(43):11090–95. doi: 10.1073/pnas.1717720115.

Desender, Kobe, Annika Boldt, and Nick Yeung. 2018. “Subjective Confidence Predicts Information Seeking in Decision Making.” *Psychological Science* 29(5):761–78. doi: 10.1177/0956797617744771.

Dunlosky, John. 2008. *Metacognition*. 1er édition. Los Angeles: SAGE Publications, Inc.

Elliott, R., B. J. Sahakian, A. P. McKay, J. J. Herrod, T. W. Robbins, and E. S. Paykel. 1996. “Neuropsychological Impairments in Unipolar Depression: The Influence of Perceived Failure on Subsequent Performance.” *Psychological Medicine* 26(5):975–89. doi: 10.1017/s0033291700035303.

Faivre, Nathan, Elisa Filevich, Guillermo Solovey, Simone Kühn, and Olaf Blanke. 2018. “Behavioral, Modeling, and Electrophysiological Evidence for Supramodality in Human Metacognition.” *The Journal of Neuroscience* 38(2):263–77. doi: 10.1523/JNEUROSCI.0322-17.2017.

Faivre, Nathan, Laurène Vuillaume, Fosco Bernasconi, Roy Salomon, Olaf Blanke, and Axel Cleeremans. 2020. “Sensorimotor Conflicts Alter Metacognitive and Action Monitoring.” *Cortex* 124:224–34. doi: 10.1016/j.cortex.2019.12.001.

Filevich, Elisa, Christina Koß, and Nathan Faivre. 2020. “Response-Related Signals Increase Confidence But Not Metacognitive Performance.” *ENeuro* 7(3):ENEURO.0326-19.2020. doi: 10.1523/ENEURO.0326-19.2020.

Fleming, Stephen M., Raymond J. Dolan, and Christopher D. Frith. 2012. “Metacognition: Computation, Biology and Function.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1594):1280–86. doi: 10.1098/rstb.2012.0021.

Gajdos, Thibault, Stephen M. Fleming, Marta Saez Garcia, Gabriel Weindel, and Karen Davranche. 2019. “Revealing Subthreshold Motor Contributions to Perceptual

Confidence." *Neuroscience of Consciousness* 2019(1). doi: 10.1093/nc/niz001.
de Gardelle, Vincent, and Christopher Summerfield. 2011. "Robust Averaging during Perceptual Judgment." *Proceedings of the National Academy of Sciences* 108(32):13341–46. doi: 10.1073/pnas.1104517108.
Geurts, Laura S., James R. H. Cooke, Ruben S. van Bergen, and Janneke F. M. Jehee. 2022. "Subjective Confidence Reflects Representation of Bayesian Probability in Cortex." *Nature Human Behaviour*. doi: 10.1038/s41562-021-01247-w.
Girshick, Ahna R., Michael S. Landy, and Eero P. Simoncelli. 2011. "Cardinal Rules: Visual Orientation Perception Reflects Knowledge of Environmental Statistics." *Nature Neuroscience* 14(7):926–32. doi: 10.1038/nn.2831.
Honig, Maija, Wei Ji Ma, and Daryl Fougnie. 2020. "Humans Incorporate Trial-to-Trial Working Memory Uncertainty into Rewarded Decisions." *Proceedings of the National Academy of Sciences* 117(15):8391–97. doi: 10.1073/pnas.1918143117.
Hoven, Monja, Damiaan Denys, Marion Rouault, Judy Luigjes, and Ruth van Holst. 2022. "How Do Confidence and Self-Beliefs Relate in Psychopathology: A Transdiagnostic Approach."
Hoven, Monja, Maël Lebreton, Jan B. Engelmann, Damiaan Denys, Judy Luigjes, and Ruth J. van Holst. 2019. "Abnormalities of Confidence in Psychiatry: An Overview and Future Perspectives." *Translational Psychiatry* 9(1):1–18. doi: 10.1038/s41398-019-0602-7.
Ji, Luyan, and William G. Hayward. 2021. "Metacognition of Average Face Perception." *Attention, Perception, & Psychophysics* 83(3):1036–48. doi: 10.3758/s13414-020-02189-7.
Katyal, Sucharit, and Stephen Fleming. 2023. "Construct Validity in Metacognition Research: Balancing the Tightrope between Rigor of Measurement and Breadth of Construct."
Kepecs, Adam, and Zachary F. Mainen. 2012. "A Computational Framework for the Study of Confidence in Humans and Animals." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1594):1322–37. doi: 10.1098/rstb.2012.0037.
Kiani, Roozbeh, and Michael N. Shadlen. 2009. "Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex." *Science (New York, N.Y.)* 324(5928):759–64. doi: 10.1126/science.1169405.
Koriat, Asher. 2007. "Metacognition and Consciousness." in *The Cambridge Handbook of Consciousness*, *Cambridge Handbooks in Psychology*, edited by E. Thompson, M. Moscovitch, and P. D. Zelazo. Cambridge: Cambridge University Press.
Lee, Alan L. F., Vincent de Gardelle, and Pascal Mamassian. 2021. "Global Visual Confidence." *Psychonomic Bulletin & Review* 28(4):1233–42. doi: 10.3758/s13423-020-01869-7.
Levitt, H. 1971. "Transformed Up‐Down Methods in Psychoacoustics." *The Journal of the Acoustical Society of America* 49(2B):467–77. doi: 10.1121/1.1912375.
Locke, Shannon M., Pascal Mamassian, and Michael S. Landy. 2020. "Performance Monitoring for Sensorimotor Confidence: A Visuomotor Tracking Study." *Cognition* 205:104396. doi: 10.1016/j.cognition.2020.104396.
Mamassian, Pascal. 2016. "Visual Confidence." *Annual Review of Vision Science* 2(1):459–81. doi: 10.1146/annurev-vision-111815-114630.
Michael, Elizabeth, Vincent de Gardelle, and Christopher Summerfield. 2014. "Priming by the Variability of Visual Information." *Proceedings of the National Academy of Sciences* 111(21):7873–78. doi: 10.1073/pnas.1308674111.
Mole, Callum D., Radka Jersakova, Georgios K. Kountouriotis, Chris JA Moulin, and Richard M. Wilkie. 2018. "Metacognitive Judgements of Perceptual-Motor Steering Performance." *Quarterly Journal of Experimental Psychology* 71(10):2223–34. doi: 10.1177/1747021817737496.
Murphy, Peter R., Ian H. Robertson, Siobhán Harty, and Redmond G. O'Connell. 2015. "Neural Evidence Accumulation Persists after Choice to Inform Metacognitive Judgments" edited by M. J. Frank. *ELife* 4:e11946. doi: 10.7554/eLife.11946.
Pereira, Michael, Nathan Faivre, Iñaki Iturrate, Marco Wirthlin, Luana Serafini, Stéphanie

Martin, Arnaud Desvachez, Olaf Blanke, Dimitri Van De Ville, and José del R. Millán. 2020. “Disentangling the Origins of Confidence in Speeded Perceptual Judgments through Multimodal Imaging.” *Proceedings of the National Academy of Sciences* 117(15):8382–90. doi: 10.1073/pnas.1918335117.

Pereira, Michael, Denis Perrin, and Nathan Faivre. 2022. “A Leaky Evidence Accumulation Process for Perceptual Experience.” *Trends in Cognitive Sciences* 26(6):451–61. doi: 10.1016/j.tics.2022.03.003.

Pleskac, Timothy J., and Jerome R. Busemeyer. 2010. “Two-Stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence.” *Psychological Review* 117:864–901. doi: 10.1037/a0019737.

Posner, Michael I. 1980. “Orienting of Attention.” *Quarterly Journal of Experimental Psychology* 32(1):3–25. doi: 10.1080/00335558008248231.

Rahnev, Dobromir. 2021. “Visual Metacognition: Measures, Models, and Neural Correlates.” *The American Psychologist* 76(9):1445–53. doi: 10.1037/amp0000937.

Rahnev, Dobromir, Tarryn Balsdon, Lucie Charles, Vincent de Gardelle, Rachel Denison, Kobe Desender, Nathan Faivre, Elisa Filevich, Stephen M. Fleming, Janneke Jehee, Hakwan Lau, Alan L. F. Lee, Shannon M. Locke, Pascal Mamassian, Brian Odegaard, Megan Peters, Gabriel Reyes, Marion Rouault, Jerome Sackur, Jason Samaha, Claire Sergent, Maxine T. Sherman, Marta Siedlecka, David Soto, Alexandra Vlassova, and Ariel Zylberberg. 2022. “Consensus Goals in the Field of Visual Metacognition.” *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 17(6):1746–65. doi: 10.1177/17456916221075615.

Recht, Samuel, Pascal Mamassian, and Vincent de Gardelle. 2022. “Metacognition Tracks Sensitivity Following Involuntary Shifts of Visual Attention.” *Psychonomic Bulletin & Review* 1–12. doi: 10.3758/s13423-022-02212-y.

Rouault, Marion, Peter Dayan, and Stephen M. Fleming. 2019. “Forming Global Estimates of Self-Performance from Local Confidence.” *Nature Communications* 10(1):1141. doi: 10.1038/s41467-019-09075-3.

Rouy, Martin, Pauline Saliou, Ladislas Nalborczyk, Michael Pereira, Paul Roux, and Nathan Faivre. 2021. “Systematic Review and Meta-Analysis of Metacognitive Abilities in Individuals with Schizophrenia Spectrum Disorders.” *Neuroscience & Biobehavioral Reviews* 126:329–37. doi: 10.1016/j.neubiorev.2021.03.017.

Seow, Tricia X. F., Marion Rouault, Claire M. Gillan, and Stephen M. Fleming. 2021. “How Local and Global Metacognition Shape Mental Health.” *Biological Psychiatry* 90(7):436–46. doi: 10.1016/j.biopsych.2021.05.013.

Zacharopoulos, George, Nicola Binetti, Vincent Walsh, and Ryota Kanai. 2014. “The Effect of Self-Efficacy on Visual Discrimination Sensitivity.” *PLOS ONE* 9(10):e109392. doi: 10.1371/journal.pone.0109392.