

Deploying voluntary spatial attention weakens perceptual metacognition

Samuel Recht ^{1,2*}, Vincent de Gardelle ³⁺, Pascal Mamassian ¹⁺

¹Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France

²Department of Experimental Psychology, University of Oxford, Oxford, UK

³ CNRS and Paris School of Economics, Paris, France

* Corresponding author: samuel.recht@gmail.com

+ these authors share senior authorship

ABSTRACT

How does orienting attention in space affect the quality of our confidence judgments? Orienting attention to a particular location is known to boost visual performance, but the deployment of attention is far from being instantaneous. Whether observers are able to monitor the time needed for attention to deploy remains largely unknown. To address this question, we adapted a “Wundt clocks” paradigm, asking observers to reproduce the phase of a rotating clock at the time of an attentional cue, and to evaluate the confidence about their responses. As expected, attention altered the latency between objective and perceived events: the average reported phase was delayed in accordance with the known latencies of (in)voluntary attention. Yet, confidence remained oblivious to these attention-induced perceptual delays, a form of ‘metacognitive blind spot’. Besides, metacognition was degraded during the deployment of voluntary attention, suggesting a tight relationship between attentional and metacognitive systems. While previous work considered how visual confidence adjusts to fully attended versus unattended locations, our study demonstrates that the very process of orienting voluntary attention in space weakens metacognitive ability.

1. INTRODUCTION

Visual confidence is the subjective reliability of a preceding decision in the visual domain; broadly, it can be defined as the self-evaluation of perceptual performance (Mamassian, 2016). Confidence is an important second-order judgement that allows to objectively evaluate the quality of a first-order judgment (e.g. “the traffic light is green”): a confidence estimate can be directly matched to first-order accuracy. One’s ability to reliably track self-performance through confidence judgments has been coined ‘metacognitive ability’, or simply metacognition. Metacognition plays a critical role in task prioritization (Aguilar-Lleyda, Lemarchand, & de Gardelle, 2020), adaptive learning (Guggenmos, Wilbertz, Hebart, & Sterzer, 2016; Hainguerlot, Vergnaud, & De Gardelle, 2018), evidence accumulation (Balsdon, Wyart, & Mamassian, 2020), information seeking (Desender, Boldt, & Yeung, 2018) and the integration of multiple decision stages (van den Berg, Zylberberg, Kiani, Shadlen, & Wolpert, 2016). It can also be used to compare performance between different tasks and modalities (de Gardelle, Le Corre, & Mamassian, 2016; de Gardelle & Mamassian, 2014).

Performance variability is usually well tracked by confidence judgments, despite multiple dissociations being documented in the literature (for a review, see Fleming & Daw, 2017). For instance, in a recent study we found visual confidence to neglect that a single attentional episode can benefit two consecutive targets. When two targets are presented in close succession within a stream of distractors, observers’ discrimination of the second target is degraded when it appears around 300 ms after the first target (the well-known ‘Attentional Blink’, e.g. Raymond, Shapiro, & Arnell, 1992). Yet, if the second target is presented immediately after the first, both are on average correctly reported: paying attention to the first item also benefits the second, a phenomenon referred to as ‘lag-1 sparing’ in the literature. However, confidence judgments remain blind to the initial sparing of performance, as if some aspects of attentional processing were concealed from metacognition (Recht, Mamassian, & de Gardelle, 2019).

Given the limitations of cognitive resources and the vast amount of peripheral sensory information, visual attention is thought to play the critical

role of a filter: it selects, prioritizes and amplifies some specific sensory information for further processing (Buschman & Kastner, 2015; Carrasco, 2011). Such a selective mechanism provides an organism with an important tool to optimize resources allocation in space and in time. Changes in performance induced by attention should, ideally, be taken into account during decision-making. Intuitively, knowing whether one was paying attention at a given moment should help to estimate the reliability of one's perception, a subjective reliability that might be used later when trying to optimize one's behavior. This reasoning can be extended to a situation where attention is on the verge of orienting, but is not yet fully deployed: such a situation should be considered as less reliable than a condition where attention had enough time to orient before target onset.

Of particular relevance to this question is the classical taxonomy in the literature that differentiates exogenous attention from endogenous attention through their respective temporal dynamics. Exogenous stands for an involuntary, early and short-lasting orienting of attention, while endogenous relates to a voluntary, late and long-lasting allocation (Carrasco, 2011). The nature of an attentional episode is therefore defined primarily by the time it takes to emerge, with exogenous attention taking roughly 100ms to be effective while approximately 300ms are necessary for endogenous attention to be allocated. Whereas the links between metacognition and the temporal aspects of attention have so far been little studied, there are several reports on its spatial aspects. Some studies have found a dissociation between the two (Rahnev et al., 2011; Wilimzig, Tsuchiya, Fahle, Einhäuser, & Koch, 2008), while others show a tight positive association between confidence and the effects of spatial attention (Denison, Adler, Carrasco, & Ma, 2018; Zizlsperger, Sauvigny, & Haarmeier, 2012; Zizlsperger, Sauvigny, Händel, & Haarmeier, 2014). The distinction between exogenous and endogenous attention regarding confidence judgments is also far from being settled. One study found that only voluntary, but not involuntary attention is reflected in confidence judgments (Kurtz, Shapcott, Kaiser, Schmiedt, & Schmid, 2017). Yet, the experimental design used in this study cannot fully guarantee the strict exogenous nature of the cueing procedure. Moreover, we note that most of the aforementioned studies on spatial attention considered how perceptual and metacognitive judgments vary between attended and unattended locations, using a typical cueing design

known as the 'Posner paradigm' (Carrasco, 2011; Posner, 1980). They measured accuracy and confidence after a valid or invalid pre-cue, and evaluated whether cued-induced changes in accuracy are reflected in confidence judgments. There is, however, an aspect of attentional deployment that has been overlooked so far: do observers properly evaluate the time it takes for spatial attention to be deployed?

From the moment the visual system 'decides' to orient spatial attention, to the moment attention is actually deployed, some period of time should inevitably pass, a period which may vary from one trial to the next. In the present work, we investigated whether metacognition considers such delay when estimating perceptual performance. On one hand, confidence could be sensitive to the average delay: observers could be more confident for shorter than longer attentional latencies. On the other hand, confidence could also be sensitive to the trial-by-trial variability, observers discerning that on a given trial attention was triggered in an unusual way (either too early or too late) compared to other trials. In other words, confidence could be implementing both the mean and the variance of attention timing to various - potentially distinct - degrees, or neither of these two dimensions.

Interestingly, in a recent work we found confidence to be oblivious to the average latency of temporal attention, despite reflecting some of its variability (Recht et al., 2019), but whether such results hold for the deployment latency of spatial attention remains an open question. Moreover, depending on the (in)voluntary nature of attention orienting, the sensitivity of confidence to the average delay and variability in timing can differ between endogenous and exogenous cueing. Intuitively, it can be argued that both endogenous attention and metacognition involving a central stage processing, one might expect a better metacognitive access to fluctuation in performance induced by the timing of endogenous - compared to exogenous - attention. Hence, to which extent metacognition includes the timing of spatial (in)voluntary attention in confidence judgments is the subject of the present work.

Here, we adapted a 'Wundt clocks' paradigm where participants have to estimate when an event occurred by reproducing the phase of a clock at the onset of a cue. Critically, the cue triggers the deployment of spatial attention towards one of multiple clocks presented in the periphery of the visual field.

Such continuous report has been proposed to be a proxy for attentional timing (Carlson, Hogendoorn, & Verstraten, 2006; Chakravarthi & VanRullen, 2011; Hogendoorn, Carlson, VanRullen, & Verstraten, 2010). The ‘Wundt clocks’ design, contrary to the canonical Posner paradigms discussed earlier, also permits to decouple attention from performance: the collected responses both involve an average delay (as a function of attention) and a variability (as a function of encoding precision, or ‘performance’). In the present study, we combine such a paradigm with a confidence forced-choice procedure (Barthelmé & Mamassian, 2009, 2010; de Gardelle & Mamassian, 2014, 2015) to assess metacognition in different attentional conditions (pre-cue, exogenous, and endogenous). Specifically, after completing two trials, participants were requested to indicate which of the two previous responses they felt more confident about being correct. To investigate the extent to which confidence in the task might be related to confidence in a distinct cognitive timing process, we compared these results to a simple detection task in which participants had to estimate their own response times (e.g. Gorea, Mamassian, & Cardoso-Leite, 2010).

Our study revealed three main findings. First, visual confidence was oblivious to the average latency of both exogenous and endogenous spatial attention. Secondly, metacognition was sensitive to the variability of such delay, but was transiently altered during endogenous/voluntary orienting of attention to a particular location. Finally, metacognitive ability in the clock task was not correlated to metacognition of response times, suggesting that metacognition of the temporal variability of attention is at least partially distinct from the introspection of internal sensorimotor variability.

2. MATERIAL & METHODS

2.1. PARTICIPANTS

20 adult volunteers were recruited from the French RISC pool of participants (age $M \pm SD = 25.85 \pm 2.30$, 14 females). They all provided informed written consent prior to the experiment. Participants were compensated for their time at a rate of 10€ per hour. The experiment consisted in two 1-hour sessions. The

experimental procedure was approved by the ethics review board of the Paris School of Economics (PSE).

2.2. OVERVIEW

Participants were engaged in two identical sessions. In both sessions, there were 3 blocks of the main task, and 2 blocks of the response time (RT) task. The order of the two tasks was counterbalanced across sessions and participants. The full experiment consisted in 432 trials for the main task and 288 trials for the RT task.

For both tasks, participants were instructed to fixate the centre of the screen during the whole trial period and their gaze was monitored online using an eye-tracker (EyeLink 1000 Hz, SR Research). Any trial during which participants blinked or moved their gaze away from the fixation dot were automatically aborted, and a new sample of the trials pair was added at the end of the block. Fixation was enforced from 200ms before cue onset.

Participants sat 60cm from the monitor (11 in front of a 1280x1024 pixels CRT monitor, 85Hz refresh rate, and 9 in front of a 1920x1080 pixels monitor, 60Hz refresh rate), their head maintained with a chin-rest. The tasks were coded using the Python programming language and the PsychoPy library (Peirce, 2007), on a Mac OS computer.

2.3. STIMULI

Each trial started with a central fixation dot presented on a grey background for 1000ms (fig. 1). Then six clocks (black outline, 4° diameter, 0.2° thick) were presented on the right and left side of the screen, with 4 clocks on the diagonals (4° eccentricity from fixation), and 2 on the horizontal midline (6° from fixation). This small difference in eccentricity did not produce observable effects on attention or confidence (see Supplementary material). Each clock had a black dot (diameter 0.2°) at the centre. The hand was a line starting 0.2° from the clock's center (length: 1.4°). Each clock rotated for 4s, at a fixed speed of 1 revolution per second, starting from a random onset phase.

At a random time uniformly drawn between 1000ms and 2000ms after clocks' onset, the probe was presented for about 20ms (1 frame for the 60Hz monitor and 2 frames for the 85 Hz monitor). In the pre-cue and exogenous conditions, this probe was a red annulus surrounding the clock (diameter: 4.1°),

whereas in the endogenous condition, the cue was a central black line (length: 1°) pointing towards the clock. In addition, in the pre-cue condition, a green line (length: 1°) was displayed centrally for the whole trial duration, indicating with 100% validity the clock about to be probed. Each clock was probed with probability $1/6$.

2.4. MAIN TASK

In each session, the 3 attention conditions were presented in 3 separate blocks of 72 trials each, in a randomized order. Each block started with 10 trials of practice. Participants were instructed to fixate the centre of the screen, to monitor all clocks (exogenous/endogenous conditions) or only the pre-cued clock (pre-cue condition), and to register the phase of the relevant clock at probe onset. They had to report the phase at the end of the trial using the mouse cursor (fig. 1). The display of the clock's hand during the report period was initiated by the participant's mouse click. Every two trials, participants were asked to select which of the two previous responses they felt more confident about being correct, by clicking on one of two squares ($2^\circ \times 2^\circ$) displayed at 6° eccentricity on each side below the fixation cross, flanked by '1' and '2' (for first or second trial in the pair). Based on this report, for each pair of trials, one trial was labelled as 'higher confidence' and the other one as 'lower confidence'. Participants were not instructed to make speeded responses. At the end of each ~15min block, participants had the opportunity to take a break. Our design did not involve between-conditions comparisons (i.e., one exogenous and one endogenous trial in the same pair) because it would have concerned a distinct experimental question, namely, the relative effect of cue type on confidence. Moreover, it would have added a significant cognitive switching cost for the participant. Yet, testing how observers are able to compare the differential effect of exogenous and endogenous attention is a question that could be of interest for further studies.

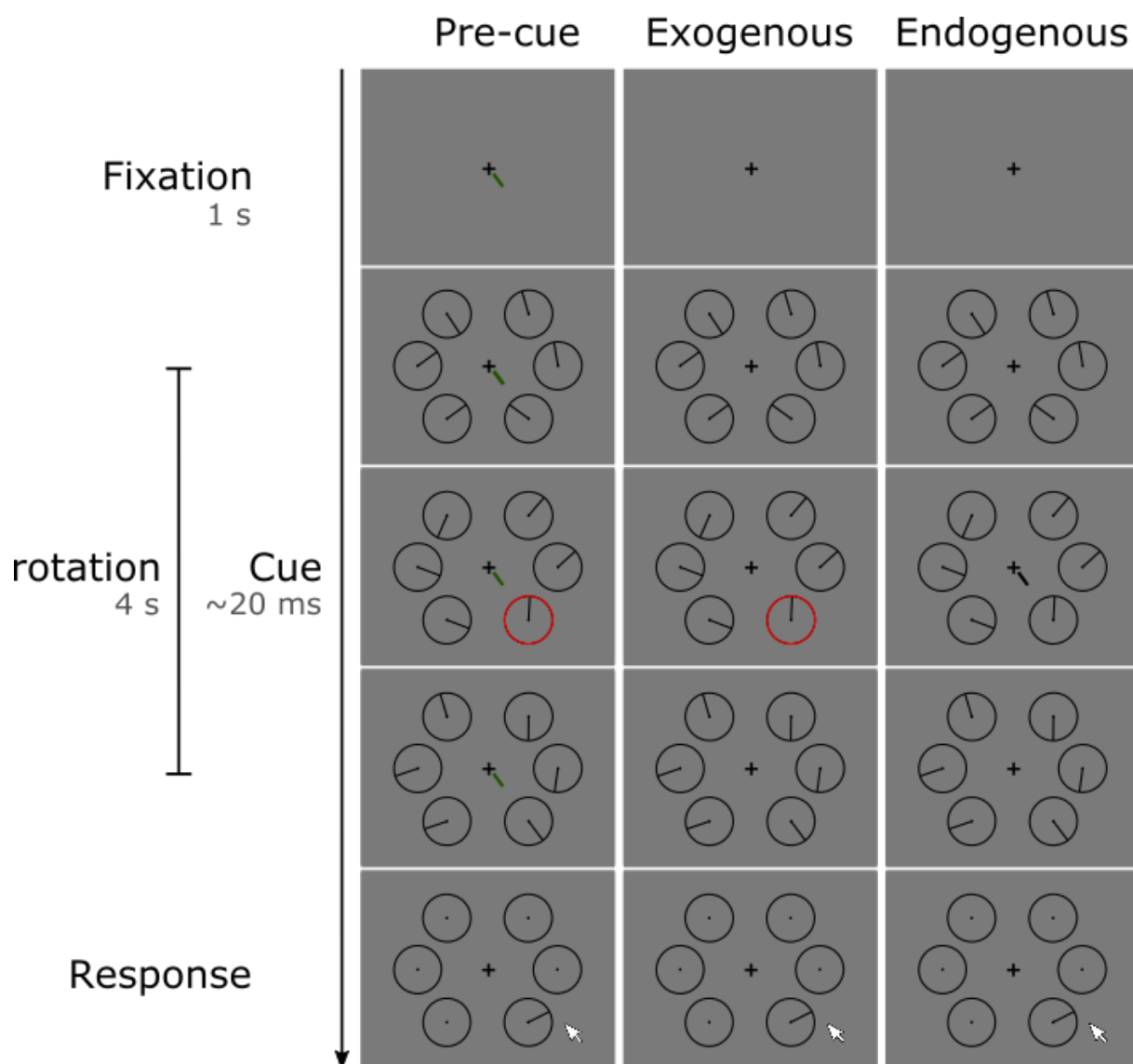


Figure 1: Experimental protocol. Main task: On each trial, the stimuli consisted in 6 clocks rotating for 4 seconds, at 1 revolution per second but with random starting phases. After a variable delay, one of the clocks was probed, either peripherally ('exogenous' condition) or centrally ('endogenous' condition). A third ('pre-cue') condition included a central pre-cue through the whole trial, indicating with 100% validity which clock will be probed. At the end of the trial, participants had to reproduce the phase of the clock at probe onset. Every two trials, participants had to indicate the best of their two preceding responses.

2.5. REACTION TIMES (RT) TASK

Each session consisted of one block of 72 trials for the pre-cue condition and one block of 72 trials for the exogenous condition, in a random order. We did not test the endogenous condition in the RT task. The stimuli were identical to

the main task, and participants had to make a speeded response by pressing a key at probe onset. Every two trials, participants were asked to select which of the two previous responses had the shortest reaction time (i.e. confidence forced-choice).

2.6. ANALYSES

Two types of analyses were applied: (a) an analysis focusing on the two parameters of the von Mises distribution - typically used for circular data in place of the normal distribution - to probe attentional latency and precision together with confidence (Section 2.6.1.); and (b) a logistic regression approach, to study how continuous responses were predictive of confidence judgment on a trial-by-trial basis (Section 2.6.2.). The logistic regression analysis considered the difference in the absolute angular error between each trial in the pair. The rationale was that if an error is lower for one trial, the participant should select this trial during confidence judgement.

2.6.1. Estimating attentional latency and precision

For each trial, the angular error between the phase of the clock at cue onset and the phase reported by the participant was calculated. We fitted von Mises distributions on angular values using maximum likelihood estimation (MLE) separately for each participant and condition (see Supplementary Material). The location parameter of the distribution (equivalent to the mean of a normal distribution) is an estimate for the latency of response errors, since it relates to the average time difference between the objective event and the perceived event. The concentration parameter (or 'kappa') is an estimate for the precision of the responses, and its inverse is analogous to the variance of a normal distribution.

To ensure that using a von Mises distribution to describe our data was sound, we used the Watson test, a circular adaptation of the Cramér-von Mises goodness-of-fit test. A bootstrapped Watson test was applied at the same granularity level as the von Mises individual fits. We found no evidence for rejecting the von Mises distribution (see Supplementary material). We therefore chose a simple von Mises distribution to account for our data. Yet, there are other possible models in the literature to account for subjective reports of

circular continuous variables such as orientation or colour. Two prominent alternatives are a model including a guess rate (Zhang & Luck, 2008) and a model with variable precision (e.g. Van Den Berg, Shin, Chou, George, & Ma, 2012). The former involves a mixture of a von Mises and a uniform distribution, in order to account for possible guesses in certain trials (Zhang & Luck, 2008). However, the guess rate model did not improve our fit (it was even worse in terms of BIC, see Supplementary Material), and the estimation of precision between conditions remained unaltered. The second, variable-precision model (Van Den Berg et al., 2012), did not change the pattern of estimated precision either (see Supplementary Material for a BIC and AIC model comparison).

2.6.2. Estimating trial-by-trial metacognition

To predict confidence from trial-by-trial error, we used logistic regression models. One potential challenge to such an approach comes from the circularity of our decision space: projected on a $[-180^\circ, 180^\circ)$ axis, a few, large-magnitude errors might be categorised as negative values, despite being generated by a positive delay. To minimise the occurrence of such cases, we centred the error space on the circular mean (i.e. the average latency) of the distribution per participant, condition and confidence level. The space was then cut at the exact opposite of the circular mean, and the axis was then moved back to the original range. This approach allowed to recalibrate the sign of extreme errors by moving to a near-linear space, on the realistic assumption that the underlying generative process obeys a von Mises distribution (see Supplementary Material).

Simple model

$$\Delta\varepsilon = |\varepsilon_A| - |\varepsilon_B| \quad (\text{eq. 1})$$

where ε_A and ε_B are the error in first and second trial in the confidence pair. A negative value of $\Delta\varepsilon$ would indicate a greater error for trial B, and a positive value a greater error for trial A.

Centred model

$$\Delta\varepsilon = |\varepsilon_A - \mu| - |\varepsilon_B - \mu| \quad (\text{eq. 2})$$

where μ is the average error (or latency) for the considered participant/condition.

Scaling model

$$\Delta\varepsilon = \frac{|\varepsilon_A - \mu| - |\varepsilon_B - \mu|}{|\varepsilon_A - \mu| + |\varepsilon_B - \mu|} \quad (\text{eq. 3})$$

For all models, we used the relative error ($\Delta\varepsilon$) as a predictor of confidence in a logistic regression model (logit), estimated per participant and condition separately

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \alpha + \beta\Delta\varepsilon,$$

where

$$p = p(\text{higher confidence} | \text{High Confidence Trial} = \text{Trial}_B)$$

Given the symmetry of the confidence forced-choice paradigm, there should not be an advantage to choose the first or second decision of a confidence pair, and therefore the parameter α should theoretically be zero. Non-zero positive values of this parameter represent biases to choose the second decision as more confident.

2.6.3. Statistical results

All the analyses were carried out using the R programming language (R Core Team, 2013). When necessary, ANOVAs were corrected using the Greenhouse-Geisser adjustment and t-tests were corrected using the Welch-Satterthwaite adjustment. We report Wilcoxon signed rank test using uppercase T when the Shapiro-Wilk normality test failed, and Student test using lowercase t otherwise. For the Student-tests and Pearson correlations, we also systematically indicated the Bayes Factor (BF_{10}), computed using the 'ttestBF' and the 'correlationBF' functions from the 'BayesFactor' R package (Morey & Rouder, 2018).

3. RESULTS

3.1. MAIN TASK

3.1.1. Spatial orienting of attention

We compared our 3 conditions of attention (pre-cue, exogenous, endogenous) in terms of the latency and precision of participants' responses (fig. 2).

The latency profile (fig. 2B) was consistent with results from the spatial attention literature showing a faster orienting for exogenous/peripheral cues compared to endogenous/central cues (see Carlson et al., 2006 for a study using the same paradigm and Carrasco, 2011 for a review). A repeated-measures ANOVA revealed a main effect of condition ($F(1.51,28.76) = 194.10$, $MSE = 1698.53$, $p < 0.001$). Bonferroni corrected t-tests ($\alpha=0.017$, corrected for 3 tests) confirmed that latency was lower for the pre-cue condition than for both the exogenous condition ($t(19) = -6.28$, $p < 0.001$; $M = -53.70$, 95% CI [-71.59 - 35.81]; $BF_{10} = 4.16 \times 10^3$) and endogenous condition ($t(19) = -15.28$, $p < 0.001$; $M = -214.63$, 95% CI [-244.02 -185.22]; $BF_{10} = 1.90 \times 10^9$), and lower for the exogenous condition than for the endogenous condition ($t(19) = -14.98$, $p < 0.001$; $M = -160.93$, 95% CI [-183.41 -138.45]; $BF_{10} = 1.37 \times 10^9$).

Regarding precision, a second ANOVA with concentration as a dependent variable showed a main effect of condition ($F(1.91,36.23) = 4.04$, $MSE = 0.43$, $p = 0.03$). However, Bonferroni-corrected t-tests (with $\alpha=0.017$, corrected for 3 tests) showed no convincing evidence for a difference between pre-cue and exogenous condition ($t(19) = -2.19$, $p = 0.041$; $M = -0.46$, 95% CI [-0.89 -0.02], $BF_{10} = 1.62$), between pre-cue and endogenous condition ($t(19) = 0.40$, $p = 0.690$; $M = 0.07$, 95% CI [-0.30 0.45], $BF_{10} = 0.25$) or between exogenous and endogenous condition ($t(19) = 2.44$, $p = 0.024$; $M = 0.53$, 95% CI [0.08 0.98], $BF_{10} = 2.45$). Importantly, the profile of the concentration parameter suggested that pre-cue and endogenous conditions led to roughly similar performance, with some evidence for a slight gain during exogenous orienting of attention (fig. 2C).

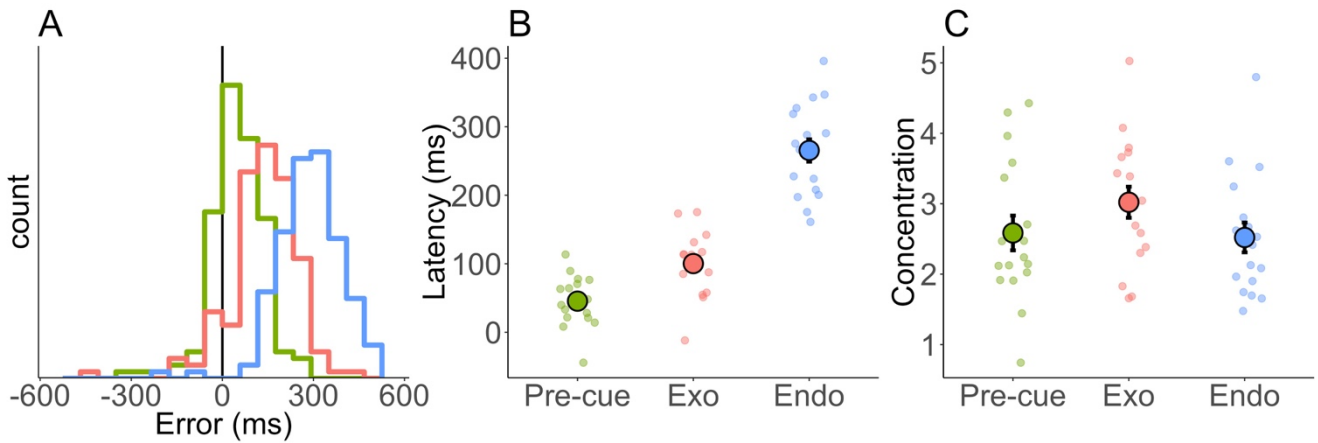


Figure 2: Latency and precision of attentional orienting. (A) Distributions of errors for a representative participant. The distribution of responses represents the angular error (reported phase minus objective phase) converted in milliseconds. Von Mises distributions were fitted to estimate the latency (location parameter) and the precision (concentration parameter) of attentional selection. The pre-cue, exogenous and endogenous conditions are represented in green, red and blue, respectively. (B) The average latency for each condition, reproducing the delays generally observed in the literature. (C) The average concentration, a measure of precision, for each condition, showing no difference between pre-cue and endogenous condition. Coloured dots correspond to individual participants in the given condition. Large dots with a black outline represent the mean across participants. Error bars represent across participants ± 1 SEM.

3.1.2. Metacognition of attentional effects

To evaluate metacognitive ability, we estimated location and concentration parameters separately for the higher confidence trials and the lower confidence trials (fig. 3A), according to the confidence forced choice response within each condition. We then evaluated in two distinct ANOVAs how location and concentration depended on confidence, the attentional condition, and the interaction between these two factors (fig. 3B and 3C).

For latency (i.e. the location parameter, fig. 3B), the ANOVA indicated a main effect of condition ($F(1.51, 28.64) = 194.75$, $MSE = 3391.36$, $p < 0.001$) as expected from our results where the data were not split between higher and lower confidence. However, we did not find a significant effect of confidence ($F(1, 19) = 0.99$, $MSE = 341.00$, $p = 0.33$), nor a confidence x condition interaction

($F(1.73,32.92) = 1.00$, $MSE = 248.66$, $p = 0.37$). The absence of a main effect of confidence suggests that within each attentional condition, confidence was oblivious to the delays induced by spatial orienting of attention.

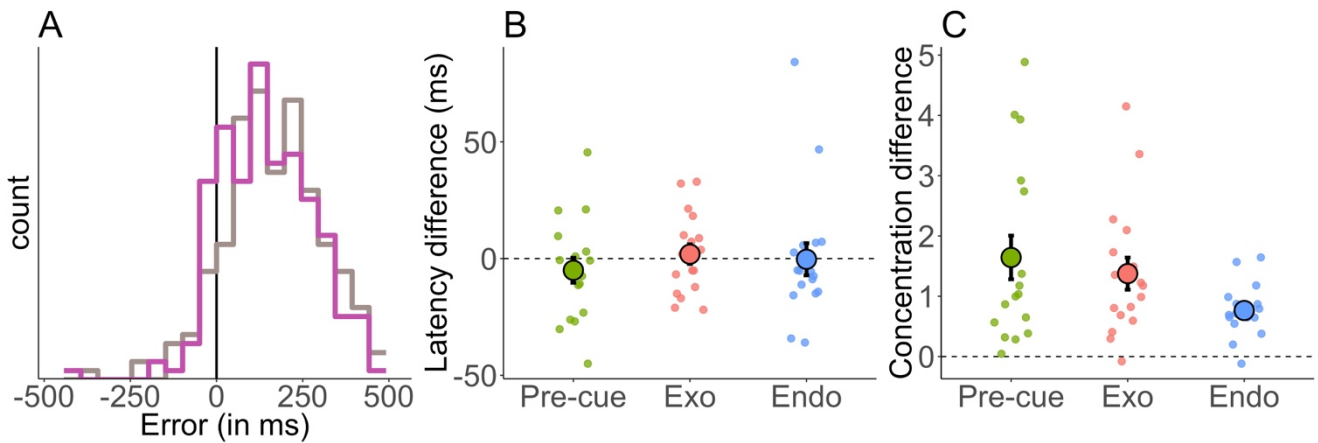


Figure 3: Metacognition of latency and precision. (A) Latency and concentration parameters were estimated by fitting von Mises distributions to the Higher (purple) and Lower (grey) confidence trials for each participant/condition. The parameter differences between Higher and Lower confidence give an estimate of metacognitive access to attentional latency and precision. The figure plots the two distributions for a representative participant. (B) The average difference in latency between Lower and Higher confidence trials for each condition. The absence of a significant difference suggests that confidence is oblivious to attentional latency. (C) The average difference in precision between Higher and Lower confidence trials. Positive values suggest that confidence has access to the variability of response precision for all conditions, despite the significant metacognitive cost for endogenous orienting compared to pre-cue. Coloured dots correspond to individual participants in the given condition. Large dots with a black outline represent the mean across participants. Error bars represent across participants ± 1 SEM.

Regarding precision (i.e. the concentration parameter, fig. 3C), as expected, we found significant main effects of condition ($F(1.95,37.02) = 3.73$, $MSE = 1.13$, $p = 0.03$) and confidence ($F(1,19) = 48.25$, $MSE = 1.0$, $p < 0.001$). Interestingly, we also observed a condition \times confidence interaction ($F(1.87,35.54)=7.05$, $MSE=0.38$, $p=0.003$). These results demonstrate that trials

labelled with higher confidence were associated with higher precision compared to lower confidence trials. Bonferroni-corrected t-tests ($\alpha=0.017$, corrected for 3 tests) confirmed a greater precision for higher confidence trials for both pre-cue ($T(19) = 210, p < 0.001$), exogenous ($T(19) = 209, p < 0.001$) and endogenous ($t(19) = 6.70, p < 0.001$; $M = 0.76, 95\% \text{ CI } [0.52 \text{ } 0.99]$; $\text{BF}_{10} = 9.1 \times 10^3$) conditions. Yet, corrected paired tests ($\alpha=0.017$) showed a significant difference between the pre-cue and the endogenous condition ($T(19) = 179, p = 0.004$). However, we found no strong evidence for a difference between pre-cue and exogenous ($t(19) = 1.85, p = 0.080$; $M = 0.46, 95\% \text{ CI } [-0.06 \text{ } 0.99]$; $\text{BF}_{10} = 0.97$) or between exogenous and endogenous condition ($t(19) = 2.17, p = 0.043$; $M = 0.54, 95\% \text{ CI } [0.02 \text{ } 1.05]$; $\text{BF}_{10} = 1.57$). Together these results show that confidence is able to access the precision of responses. Furthermore, they show that metacognitive ability is significantly greater for the pre-cue compared to the endogenous condition, pointing to a potential interaction between the deployment of voluntary attention and metacognitive ability.

3.1.3. The nature of the evidence used during metacognitive judgments

The previous analyses suggested that participants were oblivious to latency, but were accurately monitoring the precision of their response across trials. Furthermore, orienting endogenous attention appeared to induce a decrease in metacognition of precision. These analyses however give us a broad picture of attention orientation across conditions, leaving open the question of error estimation on a trial level. A straightforward approach to this question is to assume that confidence has access to some form of evidence for each trial, which can be - from the experimenter point of view - related to the difference in absolute error between trial A and trial B in each pair of trials (that is, the relative error magnitude). To investigate how metacognition tracked trial-by-trial errors, we used a logistic regression model to predict confidence judgment from the difference in error magnitude between the two trials.

Three models were compared. In the 'simple model', confidence choices were predicted from the comparison of absolute error magnitude between the two trials (eq. 1 in Section 2.6.2.). In the 'centred model', given that confidence was not able to access the overall attentional delay within each condition (Section 3.1.2. and fig. 3B), the deviation from the mean error within the condition was used, instead of the absolute error (eq. 2). Finally, in the 'scaling

model', the comparison of errors between two trials was divided by the magnitude of both errors (eq. 3), a form of scaling observed in perception (Fechner, 1964; Shepard, 1987).

The total log-likelihood across participants and conditions favoured the centred model (simple model: -843.41; centred model: -833.68; scaling model: -860.23). We compared the goodness-of-fit of each model using a Log-Likelihood Ratio (LLR), all models sharing the same number of parameters. Model comparison provided evidence for the 'centred' model compared to both the 'simple' model (LLR = 19.47) and the 'scaling' model (LLR = 53.10). We also found greater likelihood for the 'simple' model compared to the 'scaling' model (LLR = 33.64). We therefore selected the 'centred' model for all subsequent analyses. For this model, the R^2 was on average, across participants and conditions, equal to 0.08 (± 0.01 SEM).

The slope (β) of the model gives an estimate of metacognitive ability (fig. 4). After Bonferroni correction ($\alpha = 0.017$), the slope was significantly positive for the three conditions, pre-cue ($t(19) = 7.47$, $p < 0.001$; $M = 0.007$, 95% CI [0.005 0.009]; $BF_{10} = 3.67 \times 10^4$), exogenous ($t(19) = 6.34$, $p < 0.001$; $M = 0.005$, 95% CI [0.003 0.006]; $BF_{10} = 4.66 \times 10^3$) and endogenous condition ($t(19) = 5.72$, $p < 0.001$; $M = 0.003$, 95% CI [0.002 0.004] ; $BF_{10} = 1.41 \times 10^3$), showing above chance metacognitive ability for all conditions at the group level. Importantly, a comparison of the three conditions using a repeated-measures ANOVA showed a significant effect of condition ($F(1.99,37.76) = 7.29$, $MSE = 0.00$, $p = 0.02$). After Bonferroni-correction ($\alpha=0.05/3$), we found a strong evidence for a difference between the pre-cue and the endogenous conditions ($t(19) = 3.93$, $p < 0.001$; $M = 0.004$, 95% CI [0.002 0.006]; $BF_{10} = 40.40$). However, only little evidence was found for a difference between pre-cue and exogenous conditions ($t(19) = 2.17$, $p = 0.043$; $M = 0.002$, 95% CI [0 0.004]; $BF_{10} = 1.56$) and no evidence for a difference between exogenous and endogenous conditions ($t(19) = 1.56$, $p = 0.135$; $M = 0.001$, 95% CI [0 0.004]; $BF_{10} = 0.66$).

For completeness, we also analysed the intercept of the regression, which provides an estimate of metacognitive bias in favour of the second trial in the pair (independently of actual performance). We found moderate evidence for a bias in the pre-cue condition ($t(19) = -2.73$, $p = 0.013$; $M = -0.49$, 95% CI [-0.86 -0.11]; $BF_{10} = 4.02$). We found no evidence for a bias in the

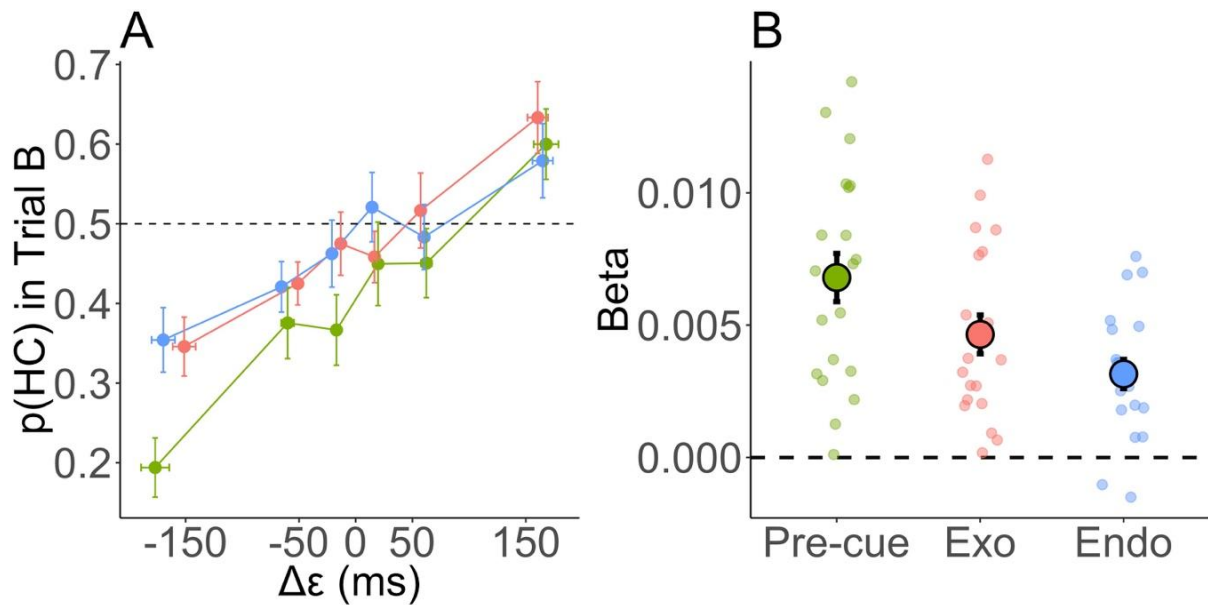


Figure 4: Deploying endogenous attention weakens metacognition. (A) The probability of higher confidence for trial B in the pair, as a function of the error difference between the two trials ($\Delta\varepsilon$) at the group level. For illustrative purposes, errors have been grouped by quantiles at the individual level, and the group average is represented with error bars for each quantile. The significant decrease in slope between the pre-cue (green) and endogenous condition (blue) confirms a metacognitive cost during voluntary orienting of spatial attention. (B) The average correlation coefficient (or slope) for each condition, used as a measure of metacognitive ability. Coloured dots correspond to individual participants in the given condition. Large dots with a black outline represent the mean across participants. Error bars represent across participants ± 1 SEM.

exogenous ($t(19) = -1$, $p = 0.330$; $M = -0.13$, 95% CI [-0.41 0.14]; $BF_{10} = 0.36$) or endogenous conditions ($t(19) = -0.99$, $p = 0.334$; $M = -0.13$, 95% CI [-0.42 0.15]; $BF_{10} = 0.36$). The ANOVA showed a significant effect of condition ($F(1.98, 37.59) = 8.99$, $MSE = 0.10$, $p < 0.001$). Bonferroni-corrected t-tests confirmed a significantly greater bias for the pre-cue condition compared to the endogenous ($t(19) = -3.86$, $p = 0.001$; $M = -0.36$, 95% CI [-0.55 -0.16]; $BF_{10} = 34.95$) and exogenous condition ($t(19) = -3.64$, $p = 0.002$; $M = -0.36$, 95% CI [-0.56 -0.15]; $BF_{10} = 22.56$), but no significant difference between exogenous and endogenous conditions ($t(19) = 0.017$, $p = 0.987$; $M = 0.002$, 95% CI [-0.21 0.21]; $BF_{10} = 0.23$). Therefore, there was evidence for a bias in the pre-cue condition only. The pre-

cue average bias log-odds was -0.49, which corresponds to a 12% increase in the probability of selecting the first trial compared to the second trial in the pair. To rule out a potential effect of the bias on the estimated metacognitive ability (or slope) found in the pre-cue condition, we tested the correlation between the value of the bias and the slope across participants. We found no evidence for a correlation ($t(18) = -0.17$, $p = 0.87$; $r = -0.04$, 95% CI [-0.47 0.41]; $BF_{10} = 0.48$). We extended such analysis across conditions and participants, with a similar outcome ($t(58) = -0.1$, $p = 0.92$; $r = -0.01$, 95% CI [-0.27 0.24]; $BF_{10} = 0.29$).

3.2. RT TASK

The RT task was designed to probe the metacognitive ability regarding the timing of a distinct, sensorimotor process, using the same stimuli as in the clock task. This task included only the pre-cue and exogenous conditions, which we did not expect to differ. Two participants were excluded from the analyses due to a technical error ($N = 18$). A repeated-measures ANOVA with median response time as a dependent variable and condition/confidence as independent variables showed an effect of confidence ($F(1,17) = 39.93$, $MSE = 0$, $p < 0.001$) but no effect of condition ($F(1,17) = 0.95$, $MSE=0$, $p = 0.34$) and no interaction ($F(1,17) = 1.41$, $MSE = 0.0$, $p=0.25$). Participants were therefore able to discriminate between fast and slow response times, and, as expected, condition did not significantly affect this ability. For all subsequent analyses, we therefore combined both conditions together.

In a similar vein to the first task, we evaluated how second-order comparison judgments (here, RT comparisons) could be predicted from the difference in first-order performance between the two trials (here, the difference in RTs), using a logistic regression. We then compared the simple model (ΔRT , log-likelihood: -1458.66) to the scaling model ($\frac{\Delta RT}{\Sigma RT}$, log-likelihood: -1454.09) using the likelihood ratio. Contrary to the first task, we found evidence in favour of the scaling model over the simple model (LLR = 9.14). We did not use the centred model for this analysis, given the nature of the task. Indeed, participants had to select the trial in which they were the fastest, not the trial in which the RT was closest to the mean RT in the condition. This result suggests a potential difference in the nature of the metacognitive evidence used in each task. We therefore used the scaling model in all subsequent analyses (group-

level average $R^2 = 0.10 \pm 0.02$ SEM). Metacognitive ability (the β in the model) was significantly positive at the group level ($t(17) = 6.83$, $p < 0.001$; $M = 5.47$, 95% CI [3.78 7.16]; $BF_{10} = 6.79 \times 10^3$), and we found no significant evidence for a metacognitive bias (the intercept in the model, $t(17) = -1.72$, $p = 0.10$; $M = -0.38$, 95% CI [-0.84 0.09]; $BF_{10} = 0.83$).

We then tested whether our results for metacognition in the main attention task could be a general ability that extends to the timing of another cognitive and sensorimotor process (i.e., metacognition in the RT task), by evaluating across participants the correlation between metacognitive abilities in the attention and sensorimotor domain. A lack of correlation here would suggest that metacognitive ability in our main task does not necessarily reduce to sensorimotor uncertainty, and might thus be related to another process, probably linked to stimulus encoding. We found no strong evidence for a correlation between the beta coefficient in the RT task and the coefficient in the clock task, for any of the conditions (correlation with the pre-cue condition: $t(16) = 2.32$, $p = 0.034$; $r = 0.50$, 95% CI [0.05 0.79]; $BF_{10} = 2.78$; exogenous: $t(16) = 1.67$, $p = 0.11$; $r = 0.39$, 95% CI [-0.10 0.72]; $BF_{10} = 1.30$; endogenous: ($t(16) = 0.21$, $p = 0.83$; $r = 0.05$, 95% CI [-0.42 0.51]; $BF_{10} = 0.50$) using Bonferroni correction ($\alpha=0.05/3$).

In sum, given the lack of evidence for a strong correlation between sensorimotor and attentional metacognition, and given the different models for these two domains (with the scaling model best fitting the RT task but not the main task data), our data points to partially distinct sources of evidence for the two tasks. In other words, metacognition of errors in the first task and metacognition of RT in the second task are relying on – at least partially – distinct sources of second-order evidence.

4. DISCUSSION

Our results shed light on three important aspects of attention orientation that we will discuss in turn. The first is the time taken by the orienting process, which appeared to be different across conditions but relatively constant within each condition. The second element is the seemingly oversight of delay:

metacognition appeared to be strongly bound to attention, to the point of making confidence blind to the temporal cost of attentional deployment. Yet, such a confidence-attention bound still allowed confidence to discriminate between different levels of response precision. Finally, confidence judgments appeared to be degraded during the deployment of voluntary attention, revealing a novel constraint on the metacognitive system.

4.1. TIMING VOLUNTARY AND INVOLUNTARY ATTENTION

Our data fit well with the results from both the attention and time perception literature. First, our results replicated previous studies that have found exogenous and endogenous attention to affect the perceived phase of moving clocks (Carlson et al., 2006; Chakravarthi & VanRullen, 2011; Hogendoorn et al., 2010). These results are also consistent with the observation that the reported time of visual events is directly affected by their relative distance from the attentional locus (Jovanovic & Mamassian, 2019, 2020).

A canonical experimental design in the spatial attention literature involves a location-specific pre-cue followed by a target, with a cue-to-target delay known to maximise attentional effects and target discrimination (e.g., 300ms for endogenous attention). These paradigms however can overlook the variability of the orienting process from trial to trial: sometimes attention is allocated earlier, sometimes later (Hogendoorn et al., 2010; Zivony & Eimer, 2020). In our paradigm, the orientation of attention is expected to occur either at the beginning of the trial (pre-cue condition) or at the very moment the observer needs to register the phase of the clock (exogenous and endogenous condition).

The current experimental design allowed us to measure the time cost and variability of attentional deployment for the exogenous and endogenous condition compared to the pre-cue condition: we found that on average the reported phase was delayed, in accordance with the known average latency for each attention type. Importantly, even if our paradigm led to reasonable delay estimations for exogenous (~101ms) and endogenous (~262ms) attention, the absolute value of these mean errors in milliseconds is not necessary directly interpretable, as the use of temporally autocorrelated stimuli is known to increase perceived lag compared to decorrelated ones (Callahan-Flintoft, Holcombe, & Wyble, 2020; Sheth, Nijhawan, & Shimojo, 2000). In other words,

temporally autocorrelated features widen the latency of information sampling, leading to less moments being sampled in a given time range. It is mainly for this reason that they should be interpreted relatively to the pre-cue condition, where attention is pre-allocated at the correct location.

On the other hand, we found no strong evidence for a difference between conditions regarding the variability of this delay. Notably, precision in the pre-cue and endogenous condition was matched, suggesting that the deployment of attention takes a relatively fixed time in this task, and that the precision we measured mostly reflects the quality of encoding. This allowed for a systematic analysis of confidence with roughly equated performance across attention conditions. Such a stable performance was also robust to changes made to the underlying descriptive model (von Mises distributions). Specifically, adding a guess parameter or allowing precision to fluctuate from trial to trial did not alter the original pattern (see Supplementary Material).

4.2. METACOGNITION IGNORES ATTENTIONAL LATENCY BUT TRACKS PRECISION

Participants appeared fully oblivious to the delay of both exogenous and endogenous attention (fig. 3B). This inability to monitor the delay of spatial attention mirrors what has been recently found for temporal attention (Recht et al., 2019), as well as previous results on executive control (Corallo, Sackur, Dehaene, & Sigman, 2008; Marti, Sackur, Sigman, & Dehaene, 2010). These studies, using dual-task paradigms, showed that metacognition ignores the delay in temporal attention induced by the Attentional Blink (Recht et al., 2019) or the delay in response times induced by the Psychological Refractory Period (Corallo et al., 2008; Marti et al., 2010). The present study suggests that the principle of a ‘metacognitive blind spot’ regarding the temporal cost of certain cognitive processes is even generalisable to single task paradigms, thought to involve lower cognitive load.

However, the inability to monitor the average timing of cognitive processes needs not preclude a fine-grained introspection of other aspects of processing, like the deviation from average latency (or relative error magnitude). Participants in our study were able to discriminate between error magnitudes, giving higher confidence judgments to more precise trials for all

conditions (fig. 3C). They were also able to discriminate between larger and shorter reaction times in our RT task. We note that in a similar vein, observers have been shown to be metacognitively aware of some of the processing stages during visual search and implicit spatial shifts of attention (Reyes & Sackur, 2014, 2017).

4.3. ALLOCATING VOLUNTARY ATTENTION IN SPACE WEAKENS METACOGNITION

Voluntary attention takes time to be allocated, and requires cognitive control to be maintained (Carrasco, 2011). Our results demonstrate a direct metacognitive cost of endogenous attention orientation in space: during orienting, the sensitivity of confidence to precision is altered compared to a condition where attention is already pre-allocated to the correct location. This metacognitive cost is found despite overall perceptual precision remaining unaffected by attentional orienting. Therefore, it is likely that a bifurcation occurs at some stage between the stream of evidence used for perceptual report and the evidence used for metacognitive judgment. The relationship between first-order (here, phase reproduction) and second-order (here, confidence judgment) decisions is the subject of ongoing debates: the account of confidence using only the first-order decision evidence (e.g. Kiani & Shadlen, 2009) is challenged by empirical dissociations between subjective and objective performance (e.g. Rahnev et al., 2011; Recht et al., 2019) and the existence of changes of mind (e.g. Resulaj, Kiani, Wolpert, & Shadlen, 2009). At first glance, the ability to assess the magnitude of self-made errors might seem paradoxical: if the participant knows about the size of the estimation error, why not correct for this error in the first place? Such comment has often been made in the context of error detection, but also in confidence studies (Yeung & Summerfield, 2012): an influential account solves this contradiction by distinguishing between the evidence used for perceptual decision and the evidence used by metacognition (see Fleming & Daw, 2017 for a review). Our results support such distinction, and suggest that the process of deploying voluntary attention has a targeted effect on metacognitive - but not perceptual - evidence. Interestingly, participants' ability to introspect on their reaction times did not particularly correlate with their metacognition of attention, suggesting that each type of metacognitive ability was driven by (partially) distinct sources of second-order evidence.

The current finding that the process of allocating endogenous attention elicited a metacognitive impairment might suggest that the top-down, frontal mechanisms needed for both voluntary orienting of attention and metacognition could share certain central resources. For example, the neuroanatomical and functional bases of visual attention have been located within a large fronto-parietal network involving, amongst other areas, the frontal-eye-field (Buschman & Kastner, 2015), while the neural bases of visual metacognition are proposed to be mostly residing within the dorsolateral and anterior parts of the prefrontal cortex (Fleming & Dolan, 2012; Fleming, Ryu, Golfinos, & Blackmon, 2014; Fleming, Van Der Putten, & Daw, 2018; Shekhar & Rahnev, 2018). All of these regions have a strong implication in top-down cognitive control, biasing incoming signals from early visual cortices and monitoring perceptual selection and decision-making (Gilbert & Li, 2013; Rahnev, 2017). Further work is needed to address how attention and metacognition interact at the functional level, to better understand the neural underpinnings of the metacognitive cost observed in the present study.

5. CONCLUSION

Metacognition allows individuals to reflect on the quality of their perceptual decisions. Yet, our results demonstrate that metacognition can be oblivious to the latency of spatial attention, an important modulator of perceptual accuracy. Furthermore, this experiment taps into the computational limitations of metacognition: the very process of voluntarily deploying attention in space was found to weaken metacognitive ability. Together, our results provide invaluable information to our understanding of metacognition and its relationship with spatial attention.

DATA AVAILABILITY

The data for the experiments is freely available via Open Science Framework:
<https://osf.io/ynm3k>

ACKNOWLEDGEMENTS

This research was supported by the Agence Nationale de la Recherche (ANR-16-CE28-0002 and ANR-16-ASTR-0014 to VdG, ANR-17-EURE-0017 and ANR-18-CE28-0015 to PM) and by PSL University (doctoral scholarship to SR).

CONTRIBUTIONS

SR, VdG and PM designed the experiment. SR conducted the experiment. SR, VdG and PM analyzed the data and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Aguilar-Lleyda, D., Lemarchand, M., & de Gardelle, V. (2020). Confidence as a Priority Signal. *Psychological Science*, *31*(9). <https://doi.org/10.1177/0956797620925039>
- Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1), 1753. <https://doi.org/10.1038/s41467-020-15561-w>
- Barthelmé, S., & Mamassian, P. (2009). Evaluation of Objective Uncertainty in the Visual System. *PLoS Computational Biology*, *5*(9), e1000504. <https://doi.org/10.1371/journal.pcbi.1000504>
- Barthelmé, S., & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences*, *107*(48), 20834–20839. <https://doi.org/10.1073/pnas.1007704107>
- Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144. <https://doi.org/10.1016/j.neuron.2015.09.017>
- Callahan-Flintoft, C., Holcombe, A. O., & Wyble, B. (2020). A delay in sampling information

- from temporally autocorrelated visual stimuli. *Nature Communications*, 11(1), 1852.
<https://doi.org/10.1038/s41467-020-15675-1>
- Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? *Journal of Vision*, 6, 1406–1411. <https://doi.org/10.1167/6.12.6>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*.
<https://doi.org/10.1016/j.visres.2011.04.012>
- Chakravarthi, R., & VanRullen, R. (2011). Bullet trains and steam engines: Exogenous attention zips but endogenous attention chugs along. *Journal of Vision*.
<https://doi.org/10.1167/11.4.1>
- Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on Introspection. *Psychological Science*. <https://doi.org/10.1111/j.1467-9280.2008.02211.x>
- de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a Common Currency between Vision and Audition. *PloS One*, 11(1), e0147901.
<https://doi.org/10.1371/journal.pone.0147901>
- de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychological Science*, 25(6), 1286–1288.
<https://doi.org/10.1177/0956797614528956>
- de Gardelle, V., & Mamassian, P. (2015). Weighting Mean and Variability during Confidence Judgments. *Plos One*, 10(3), e0120870. <https://doi.org/10.1371/journal.pone.0120870>
- Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences*, 201717720. <https://doi.org/10.1073/pnas.1717720115>
- Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science*.
<https://doi.org/10.1177/0956797617744771>
- Fechner, G. T. (1964). *Elemente der Psychophysik*. 2 (1964). Bonset.
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision performance: A general Bayesian framework for metacognitive computation. *Psychological Review*, 124(1), 1–59. <https://doi.org/10.1037/rev0000045>
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical*

- Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1594), 1338–1349. <https://doi.org/10.1098/rstb.2011.0417>
- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*. <https://doi.org/10.1093/brain/awu221>
- Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*. <https://doi.org/10.1038/s41593-018-0104-6>
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*. <https://doi.org/10.1038/nrn3476>
- Gorea, A., Mamassian, P., & Cardoso-Leite, P. (2010). Introspective duration estimation of reactive and proactive motor responses. *Acta Psychologica*, 134(2), 142–153. <https://doi.org/10.1016/j.actpsy.2010.01.005>
- Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *ELife*. <https://doi.org/10.7554/eLife.13388>
- Hainguerlot, M., Vergnaud, J. C., & De Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports*. <https://doi.org/10.1038/s41598-018-23936-9>
- Hogendoorn, H., Carlson, T. A., VanRullen, R., & Verstraten, F. A. J. (2010). Timing divided attention. *Attention, Perception, and Psychophysics*, 72(8), 2059–2068. <https://doi.org/10.3758/APP.72.8.2059>
- Jovanovic, L., & Mamassian, P. (2019). When an Event Is Perceived Depends on Where We Attend. *I-Perception*, 10(3), 1–10. <https://doi.org/10.1177/2041669519858096>
- Jovanovic, L., & Mamassian, P. (2020). Temporal context affects the perceived time of visual events. *Psychonomic Bulletin and Review*, 27, 56–61. <https://doi.org/10.3758/s13423-019-01682-x>
- Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science (New York, N.Y.)*, 324(5928), 759–764. <https://doi.org/10.1126/science.1169405>
- Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The Influence of

- Endogenous and Exogenous Spatial Attention on Decision Confidence. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-06715-w>
- Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, 2(1), 459–481. <https://doi.org/10.1146/annurev-vision-111815-114630>
- Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*. <https://doi.org/10.1016/j.cognition.2010.01.003>
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. Retrieved from <https://cran.r-project.org/package=BayesFactor>
- Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1080/00335558008248231>
- R Core Team. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <http://www.r-project.org/>
- Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, 14(12), 1513–1515. <https://doi.org/10.1038/nn.2948>
- Rahnev, Dobromir. (2017). Top-Down Control of Perceptual Decision Making by the Prefrontal Cortex. *Current Directions in Psychological Science*. <https://doi.org/10.1177/0963721417709807>
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 849–860. <https://doi.org/10.1037/0096-1523.18.3.849>
- Recht, S., Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-48063-x>
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461(7261), 263–266. <https://doi.org/10.1038/nature08275>

- Reyes, G., & Sackur, J. (2014). Introspection during visual search. *Consciousness and Cognition*.
<https://doi.org/10.1016/j.concog.2014.08.009>
- Reyes, G., & Sackur, J. (2017). Introspective access to implicit shifts of attention. *Consciousness and Cognition*, 48, 11–20. <https://doi.org/10.1016/j.concog.2016.10.003>
- Shekhar, M., & Rahnev, D. (2018). Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *Journal of Neuroscience*.
<https://doi.org/10.1523/JNEUROSCI.3484-17.2018>
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*. <https://doi.org/10.1126/science.3629243>
- Sheth, B. R., Nijhawan, R., & Shimojo, S. (2000). Changing objects lead briefly flashed ones. *Nature Neuroscience*. <https://doi.org/10.1038/74865>
- Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*.
<https://doi.org/10.1073/pnas.1117465109>
- van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. *Current Biology*.
<https://doi.org/10.1016/j.cub.2016.10.021>
- Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision*, 8(5), 7. <https://doi.org/10.1167/8.5.7>
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1594), 1310–1321. <https://doi.org/10.1098/rstb.2011.0416>
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*. <https://doi.org/10.1038/nature06860>
- Zivony, A., & Eimer, M. (2020). Distractor Intrusions Are the Result of Delayed Attentional Engagement: A New Temporal Variability Account of Attentional Selectivity in Dynamic Visual Tasks. *Journal of Experimental Psychology: General*.
<https://doi.org/10.1037/xge0000789>
- Zizlsperger, L., Sauvigny, T., & Haarmeier, T. (2012). Selective attention increases choice

certainty in human decision making. *PLoS ONE*.

<https://doi.org/10.1371/journal.pone.0041136>

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. *Nature Communications*.

<https://doi.org/10.1038/ncomms4940>

Deploying voluntary spatial attention weakens perceptual metacognition

SUPPLEMENTARY MATERIAL

TIMING SPATIAL ATTENTION

THE VON MISES MODEL

We assume that the distribution of errors in a given condition follows a von Mises distribution, an approximation of the wrapped normal distribution:

$$f(x | \mu, k) = \frac{e^{k \cos(x-\mu)}}{2\pi I_0(k)}$$

where I_0 is the modified Bessel function of the first kind of order 0, μ is the location parameter (equivalent to the mean in a normal distribution), k is the concentration parameter ($1/k$ is analogous to the variance in a normal distribution) and x the angular error in a given trial. All our analyses were carried out using the R programming language.

To ensure that the von Mises distribution was also preferable to alternative (e.g. asymmetrical) distributions, we performed a Watson test, a circular adaptation of the Cramér-von Mises test (Watson, 1961; Lockhart & Stephens, 1985), using a bootstrapping procedure (Sun, 2009). A von Mises was fitted to the original N samples using Maximum Likelihood Estimation (grouped by participant and condition), and the estimated parameters were used to generate N random samples from a von Mises (the process was repeated 2,000 times per original fit). We then compared the Watson's U^2 statistic for the original dataset to the bootstrapped U^2 distribution, counting the number of times our statistic calculated on the surrogates was greater than the statistic of the real data. We selected two approaches to check significance level: we computed a group-level p-value per condition, to check the overall evidence against a von Mises model in a given condition. We find no evidence for rejecting the von Mises model, neither for the pre-cue ($p_{\text{bootstrap}} = 0.26$), exogenous ($p_{\text{bootstrap}} = 0.34$) nor endogenous ($p_{\text{bootstrap}} = 0.44$) conditions. We also looked at the individual level, considering the individual rank for each participant/condition, but correcting for multiple tests (Bonferroni-corrected for 60 tests, $\alpha = 0.05/60 = 0.0008$). We only found 3 situations out of 60 with significant evidence against the von Mises, all in the pre-cue condition. Relaxing slightly the corrected alpha to 0.001 did not increase the number of significant tests. We checked that the three participants concerned were not driving our conclusion by doing a reanalysis of the dataset without them. Excluding these three participants did not change

the observed patterns regarding latency blindness or weakened metacognition in the endogenous condition.

ALTERNATIVES TO THE SIMPLE VON MISES MODEL

It has been suggested that the distribution of errors following the encoding of a stimulus into working memory can be modelled as a mixture of a von Mises and a uniform distribution, the later accounting for guesses, which represent no encoding at all (Zhang & Luck, 2008). We therefore checked that a model involving a mixture of the two was better at explaining our data. The mixture model was defined with one additional parameter g for the guess rate, as follows:

$$f(x | \mu, k, g) = g \frac{1}{2\pi} + (1 - g) \frac{e^{k \cos(x-\mu)}}{2\pi I_0(k)}$$

The guess rate however could be shared across attentional conditions or not, we therefore had two variants of the von Mises + guess model: one with a shared, fixed guess rate across conditions ('VM+FG', 7 parameters) and one with a specific guess rate for each condition ('VM+G', 9 parameters).

A second line of thought in the working memory literature proposes that encoding from trial to trial is of variable precision rather than constant (Fougnie, Suchow, & Alvarez, 2012; Van Den Berg, Shin, Chou, George, & Ma, 2012). In such a case, errors are coming from a mixture of von Mises distributions with their concentration following a higher order distribution (often a Gamma distribution). We therefore tested a third, variable-precision model (adapted from Van Den Berg et al., 2012). Contrary to Van Den Berg and colleagues, we did not use the Fisher information (J) as the measure of precision, but we directly used the concentration parameter (k) instead. The Fisher information being monotonically related to k , we kept the latter to make it comparable to our main model ('pure VM').

$$f(x | \mu, \bar{k}, \tau) = \int_0^{\infty} \frac{e^{k \cos(x-\mu)}}{2\pi I_0(k)} \text{Gamma}(k; \frac{\bar{k}}{\tau}, \tau) dk$$

where I_0 is the modified Bessel function of the first kind of order 0, μ is the location parameter of the von Mises distributions, $\frac{\bar{k}}{\tau}$ is the shape parameter (with \bar{k} as the mean concentration) and τ the scale parameter of the gamma distribution, x the angular error in a given trial. The variable-precision model ('VP') is fitted separately for each condition, and therefore has 9 parameters. We also tested three other variants: one with a fixed shape, but variable scale parameter across conditions ('VP-FSh', 7 parameters), one with fixed scale but variable shape parameter ('VP-FSc', 7 parameters) and finally, one with both shape and scale parameters fixed across all attentional conditions ('VP-F', 5 parameters).

All of the tested models involved fitting a specific location parameter (μ) for each condition, in light of the strong and systematic difference in average latency observed between attentional conditions (Carlson, Hogendoorn, & Verstraten, 2006; Chakravarthi & VanRullen, 2011; Hogendoorn, Carlson, VanRullen, & Verstraten, 2010). Note that our model comparison approach was not meant to be fully exhaustive, but rather to check that our results hold when considering possible alternatives to our definition of precision.

Models were fitted using MLE. All analyses were carried out using R programming language. BIC and AIC were estimated for each model, and the difference between the pure VM model and the other models for each estimator is denoted ΔBIC and ΔAIC (fig. S1). A negative value suggests a better fit for the pure VM model. BIC is known to penalize more heavily the number of parameters than AIC.

When considering the von Mises + guess family models, a first important observation is that the VM+G model, which supposes a variable guess rate between conditions, was significantly worse than the pure von Mises, according to ΔBIC ($T(19) = 27$, $p=0.002$) and not significantly different according to ΔAIC ($T(19) = 146$, $p=0.133$). Importantly, it also performed significantly worse than the model with shared guess rate across conditions relative to BIC ($T(19) = 201$, $p < 0.001$), the difference in AIC between these two models was not significant ($T(19) = 145$, $p = 0.143$). It is therefore highly unlikely that a change in guess rate

between attentional conditions would explain the difference in metacognition observed in our data. The benefit of adding a stable guess rate across condition (VM+FG) was unclear, with only the AIC favouring this model ($T(19) = 166$, $p=0.021$), but not the BIC ($T(19) = 74$, $p = 0.261$), and only when not correcting for multiple comparisons.

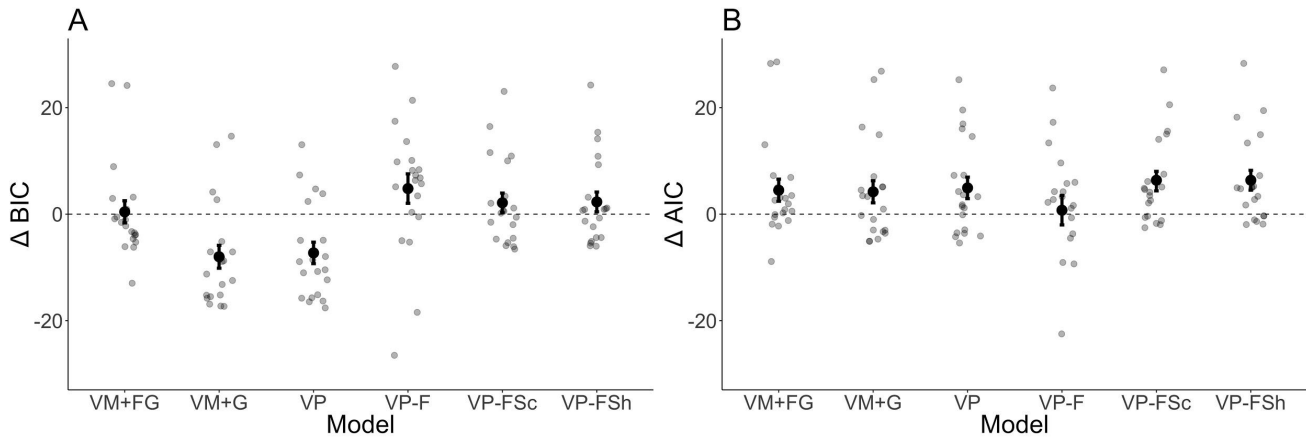


Figure S1: BIC and AIC comparison. (A) The difference in the Bayesian Information Criterion (BIC) between the pure von Mises and each of the other models. A negative value suggests evidence in favour of the pure von Mises model. (B) Same measure but using the Akaike Information Criterion (AIC). Low alpha dots correspond to individual participants in the given model. Black dots represent the mean across participants. Error bars represent across participants SEM.

When considering the variable-precision models, the worst model was the full VP, which fitted a specific set of shape and scale parameters to each condition. This model's BIC was significantly worse than the pure VM ($T(19)=26$, $p=0.002$) and there was no significant difference in AIC ($T(19) = 154$, $p=0.069$). The VP-F, which fixes the parameters across conditions, was not significantly better than the pure VM for BIC ($t(19)=1.74$, $p=0.098$) nor AIC ($t(19) = 0.27$, $p=0.793$). When fixing one parameter of the VP, we found no significant difference in BIC (for VP-FSh: $T(19)=118$, $p=0.647$; for VP-FSc: $T(19) = 119$, $p=0.622$), but a lower AIC for both models (for VP-FSh: $T(19) = 187$, $p=0.001$; for VP-FSc: $T(19) = 188$, $p=0.001$). The average ΔAIC was 6.37 for the model with fixed shape and 6.19 for the fixed scale model. Therefore, both of these two models were accounting

equally well for the data, but the evidence favouring these models over the pure von Mises was fairly low, particularly when using BIC.

ROBUSTNESS OF PRECISION ESTIMATES

Despite the low evidence in favour of a guess rate, we nevertheless checked that concentration remained stable across conditions when accounting for guesses. We analysed the effect of condition on the precision parameters of the von Mises + Fixed Guess model which was the only plausible candidate given our model comparison. The estimated values are shown in fig. S2 A and B. The latency parameter was not expected to vary at all from one model to the other, but the concentration could change - in theory uniformly - because of the added guess rate parameter. A repeated-measures ANOVA with latency as dependent variable and condition as independent variable revealed a main effect of condition on latency ($F(1.52,28.79) = 200.34$, $MSE=1656.13$, $p<0.001$). Bonferroni corrected paired t-tests confirmed that latency was lower in the pre-cue condition than for the exogenous ($t(19) = -6.47$, $p<0.001$) and endogenous ($t(19) = -15.57$, $p<0.001$) conditions, and that the latency in the exogenous condition was lower than in the endogenous condition ($t(19) = -15.07$, $p<0.001$).

A second ANOVA with concentration as dependent variable showed a main effect of condition ($F(1.95,37.06) = 3.95$, $MSE=0.94$, $p=0.03$). Bonferroni-corrected ($\alpha=0.05$, corrected for 3 tests), paired t-tests showed no significant difference between the pre-cue and exogenous conditions ($t(19) = -1.03$, $p=0.315$) nor between the pre-cue and endogenous conditions ($t(19) = 1.69$, $p=0.108$), but a significant difference between the exogenous and endogenous conditions ($t(19)=3.00$, $p=0.007$). The von Mises + Fixed Guess model was therefore leading to the exact same conclusions as the pure von Mises regarding both latency and concentration.

The absence of difference in guess rate between conditions, and the stability of the concentration parameter pattern between the pure von Mises and the von Mises + Fixed Guess rate models rule out a strong effect of guess rate during attentional orienting. Furthermore, it has been suggested that models involving guess rates (like the Zhang & Luck model used in the present analysis) should be interpreted with caution given the risk of inflated guess rate estimates. In

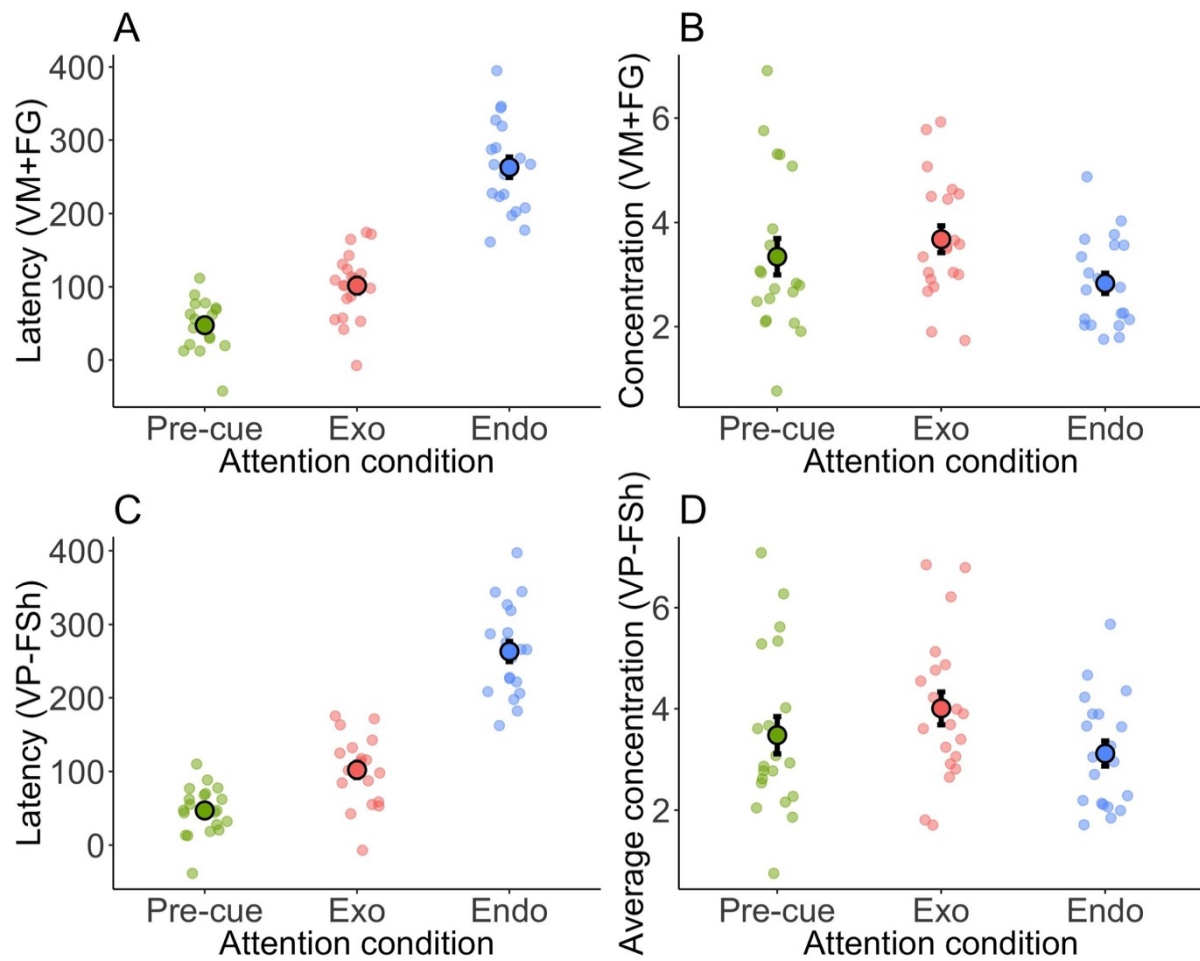


Figure S2: Estimated attention parameters for the two best alternative models. (A) The latency for each attentional orienting condition in the von Mises + Fixed Guess model. (B) The precision for each condition in the von Mises + Fixed Guess model. There is no significant difference between the pre-cue and exogenous/endogenous conditions. (C) The latency for each attentional orienting condition in the Variable-precision with fixed shape parameter model. (D) The average concentration for each attentional orienting condition in the Variable-precision with fixed shape parameter model. There is no significant difference between the pre-cue and exogenous/endogenous conditions. Coloured dots correspond to individual participants in the given condition. Black-outlined dots represent the mean across participants. Error bars represent across participants SEM.

particular, this risk has been shown to exist when the true generative process is a variable precision model involving zero guess rate (Ma, 2018), or when the

error space is non-linearly related to the stimulus space (Schurgin, Wixted, & Brady, 2020).

In a second control analysis, we considered the possibility of a variable precision across trials. We selected the Variable-precision with fixed shape model (fig. S2, C and D). A repeated-measure ANOVA was first applied on latency. It confirmed the effect of condition on latency ($F(1.51, 28.74) = 203.46$, $MSE=1642.46$, $p<0.001$). The difference between the pre-cue and exogenous conditions ($t(19) = -6.62$, $p<0.001$), the pre-cue and endogenous conditions ($t(19) = -15.7$, $p<0.001$) and between the exogenous and endogenous conditions ($t(19) = -15.12$, $p<0.001$) were all significant after Bonferroni-correction ($\alpha=0.05/3$). A second ANOVA was applied with average concentration as a dependent variable, and condition as an independent variable.

The effect of condition on the average concentration was significant ($F(1.98, 37.69) = 4.06$, $MSE=1.00$, $p=0.03$), but this effect was driven by a higher average precision in the exogenous compared to endogenous condition ($t(19) = 2.78$, $p=0.012$). The difference between the pre-cue and endogenous/exogenous conditions was not significant (all $p>0.117$, Bonferroni-corrected with $\alpha=0.05/3$). These results are all fully consistent with what was observed using the pure VM model and confirm the robustness of this model in analysing our experimental results.

Together these results, in tandem with the low evidence for a strictly better performing model over the pure von Mises, suggest that our attentional manipulation strongly affected average latency (μ) but not precision (k) of the response distributions, particularly when considering endogenous (i.e. voluntary) attention. Importantly, this was true regardless of the metric used (fixed or variable precision). Moreover, adding a guess rate parameter was only very weakly beneficial when the guess rate was fixed across conditions. Adding this stable guess rate did not alter the original pattern, a fixed guess rate being equivalent to adding a single probability constant to all conditions (figure S2).

CLOCKS' ECCENTRICITY

Our experimental paradigm involved two distinct eccentricities: 4 of the clocks were located at 4° eccentricity and 2 clocks at 6° eccentricity. The

eccentricity here was also depending on the position relative to the horizontal meridian (the 6° eccentricity landing on the meridian). We nonetheless checked if eccentricity has an effect on our results. We added the eccentricity factor to the ANOVAs reported in section 3.1.2. of the main manuscript, where latency (or concentration) was predicted by confidence (higher vs. lower) and attention condition (pre-cue vs. exogenous vs. endogenous).

We found no evidence for an effect of eccentricity on latency ($F(1,19) = 0.08$, $MSE = 721.78$, $p = 0.78$), no eccentricity x confidence interaction ($F(1,19) = 0.51$, $MSE = 224.13$, $p = 0.48$) and no eccentricity x condition interaction ($F(1.74, 33.09) = 0.40$, $MSE = 609.91$, $p = 0.64$). The triple interaction confidence x condition x eccentricity was also not significant ($F(1.71, 32.57) = 0.43$, $MSE = 528.87$, $p = 0.62$).

Regarding concentration, we found a main effect of eccentricity ($F(1, 19) = 13.56$, $MSE = 1.86$, $p = 0.002$), but no confidence x eccentricity interaction ($F(1, 19) = 2.98$, $MSE = 0.72$, $p = 0.10$) or condition x eccentricity interaction ($F(1.99, 37.72) = 0.73$, $MSE = 1.27$, $p = 0.49$). The triple interaction confidence x condition x eccentricity was also not significant ($F(1.46, 27.78) = 0.72$, $MSE = 1.97$, $p = 0.45$).

Therefore, we can conclude that eccentricity (or the clock's position relative to the horizontal meridian) affected perceptual performance via encoding precision, as reflected by the main effect on the concentration parameter. However, it did not affect metacognition, since the effect of eccentricity on precision did not interact with confidence level. The effect of eccentricity on performance also did not seem to depend on attention. However, the eccentricities considered here are small and similar to each other. It would be interesting to consider peripheral presentations in future studies.

REFERENCES

- Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? *Journal of Vision*, 6, 1406–1411. <https://doi.org/10.1167/6.12.6>
- Chakravarthi, R., & VanRullen, R. (2011). Bullet trains and steam engines: Exogenous attention zips but endogenous attention chugs along. *Journal of Vision*.
<https://doi.org/10.1167/11.4.1>
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*. <https://doi.org/10.1038/ncomms2237>
- Hogendoorn, H., Carlson, T. A., VanRullen, R., & Verstraten, F. A. J. (2010). Timing divided attention. *Attention, Perception, and Psychophysics*, 72(8), 2059–2068.
<https://doi.org/10.3758/APP.72.8.2059>
- Lockhart, R. A., & Stephens, M. A. (1985). Tests of fit for the von Mises distribution. *Biometrika*, 72(3), 647–652. <https://doi.org/10.1093/biomet/72.3.647>
- Ma, W. J. (2018). Problematic usage of the Zhang and Luck mixture model. *BioRxiv*.
<https://doi.org/10.1101/268961>
- Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, 1-17.
<https://doi.org/10.1038/s41562-020-00938-0>
- Sun, Z. (2009). Comparing measures of fit for circular distributions (Doctoral dissertation). Retrieved from
https://dspace.library.uvic.ca/bitstream/handle/1828/2698/zhengsun_master_thesis.pdf
- Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*.
<https://doi.org/10.1073/pnas.1117465109>
- Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, 48(1/2), 109–114.
<https://doi.org/10.1093/biomet/48.1-2.109>
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*. <https://doi.org/10.1038/nature06860>