Journal of Vision (2024) 24(8):12, 1–11

Sensorimotor confidence for tracking eye movements

Alexander Goettker* Abteilung Allgemeine Psychologie,
Justus-Liebig University Giessen, Giessen, Germany

Shannon M. Locke* Laboratoire des Systèmes Perceptifs,
Département d'Études Cognitives,
École Normale Supérieure,
PSL University, Paris, France

Karl R. Gegenfurtner Abteilung Allgemeine Psychologie,
Justus-Liebig University Giessen, Giessen, Germany

Pascal Mamassian Laboratoire des Systèmes Perceptifs,
Département d'Études Cognitives,
École Normale Supérieure,
PSL University, Paris, France

For successful interactions with the world, we often have to evaluate our own performance. Although eye movements are one of the most frequent actions we perform, we are typically unaware of them. Here, we investigated whether there is any evidence for metacognitive sensitivity for the accuracy of eye movements. Participants tracked a dot cloud as it followed an unpredictable sinusoidal trajectory and then reported if they thought their performance was better or worse than their average tracking performance. Our results show above-chance identification of better tracking behavior across all trials and also for repeated attempts of the same target trajectories. Sensitivity in discriminating performance between better and worse trials was stable across sessions, but judgements within a trial relied more on performance in the final seconds. This behavior matched previous reports when judging the quality of hand movements, although overall metacognitive sensitivity for eye movements was significantly lower.

Introduction

When was the last time you moved your eyes? The answer is probably only a few hundred milliseconds ago, when your eyes jumped from word to word while reading this text. Despite the importance of eye movements for visual perception (Findlay & Gilchrist, 2003; Gegenfurtner, 2016; Schütz, Braun, & Gegenfurtner, 2011), we usually only have limited awareness of when and how our eyes move (Goettker, Braun, Schütz, & Gegenfurtner, 2018; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001; van Zoest & Donk, 2010). Eye movements are important because they allow us to place the high-acuity foveal region on visual elements of interest. Eye movements are conducted with high precision and low latency (Carpenter, 1988; Leigh & Zee, 2015; Liston & Stone, 2014). However, eye movements are not perfect and will incur some amount of error because of different sources of noise (van Beers, 2007). Errors in eye movements can be registered by the brain as shown by oculomotor adaptation, where these error signals drive motor learning (for a review, see Pélisson, Alahyane, Panouillères, & Tilikete, 2010). But can error signals be used by the observer to judge the accuracy of their own eye movements?

For visually-guided hand movements, recent research indicates that humans can estimate their error to report sensorimotor confidence (Locke, Mamassian, & Landy, 2020; Mole, Jersakova, Kountouriotis, Moulin, & Wilkie, 2018). Sensorimotor confidence is a subjective report of performance in a task that takes into account (1) the quality of the perceptual information, (2) the quality of the motor execution, and (3) the sensorimotor goal (Locke et al., 2020). In these studies, participants used their hands to control an input device in a computerized game. Subsequently, they made a confidence judgement about their performance in the task (i.e., a sensorimotor confidence report about their accuracy). This confidence judgment agreed with

Citation: Goettker, A., Locke, S. M., Gegenfurtner, K. R., & Mamassian, P. (2024). Sensorimotor confidence for tracking eye movements. *Journal of Vision*, *24*(8):12, 1–11, https://doi.org/10.1167/jov.24.8.12.

https://doi.org/10.1167/jov.24.8.12 Received April 8, 2024; published August 23, 2024 ISSN 1534-7362 Copyright 2024 The Authors

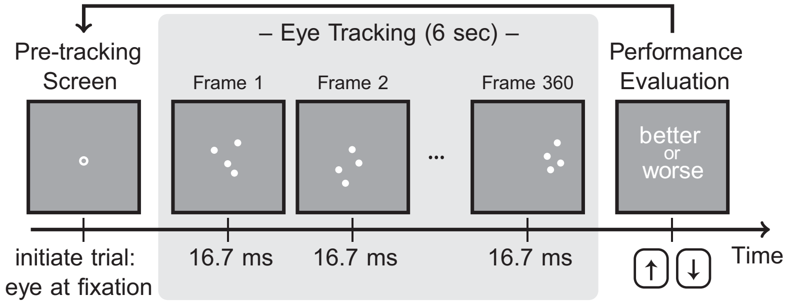

Figure 1. Trial Structure. Participants initiated the trial with a fixation check, then tracked the center of a moving dot cloud for six seconds. Four dots were drawn each frame from a 2D Gaussian distribution whose horizontal position followed an unpredictable sum-of-sinusoids trajectory. Sensorimotor confidence was reported as a binary better/worse performance evaluation, where participants reported their belief about their tracking accuracy relative to their personal average level of tracking accuracy.

the participants' true objective performance, albeit with some variability across participants. But can we generalize these studies to eye movements? There are reasons to doubt that sensorimotor confidence for eye movements will be similar because their function (Land & Hayhoe, 2001) and the neuroanatomy of oculomotor circuits (Leigh & Zee, 2015) substantially differ from that of voluntary hand movements.

Although we are usually not aware of how our eyes move, we can bring them under cognitive control and voluntarily direct our gaze at a certain visual stimulus, for example a fixation cross (Gegenfurtner, 2016; Thaler, Schütz, Goodale, & Gegenfurtner, 2013) or look back at a remembered location (Pierrot-Deseilligny, Rivaud, Gaymard, & Agid, 1991). The few studies that have investigated the subjective experience of eye movements have focused on motor awareness not sensorimotor confidence. Although sensorimotor confidence specifically considers the quality of perceptual information and the goals of the movement, motor awareness is only the knowledge about whether one's eyes have moved. Recently, Vencato and Madelain (2020) could even show that after training, observers were able to estimate their own saccade latency with an accuracy of about 40 ms. Eye movement latencies can also be altered by reinforcement learning (Vullings & Madelain, 2018) and are under discriminative control depending on the visual consequences (Vullings & Madelain, 2019). While these results clearly show that observers can sometimes control and monitor some aspects of their eye movements, other results indicate that eye movements sometimes escape voluntary control. In particular, there are circumstances where the eyes can react to things that are imperceivable (see Spering & Carrasco, 2015). For example, displacements of a target during a saccade are not visible, but can still lead to saccade adaptation (Bridgeman, Van der Heijden, & Velichkovsky, 1994; Deubel, Schneider, & Bridgeman, 1996; Klingenhoefer & Bremmer, 2011). Such results speak for limited motor awareness of eye movements (Morvan & Maloney, 2012; Tavassoli & Ringach, 2010) and therefore possible limited sensorimotor confidence. However, it is important to remember that corrections without awareness of changes in the target position are also possible for hand movements (Goodale, Pelisson, & Prablanc, 1986; Prablanc & Martin, 1992). Still, for hand movements, decisions seem to be based on knowledge about their own variability (Trommershäuser, Maloney, & Landy, 2008), and as reviewed above, reliable sensorimotor confidence has been reported (Locke et al., 2020; Mole et al., 2018).

Here we wanted to directly address the question whether an observer can judge the accuracy of their own eye movements by measuring sensorimotor confidence for tracking movements. Participants had to track the trajectory of an unpredictably moving target, whose position was only indicated by randomly sampled dots (see Figure 1). Subsequently, participants had to judge whether they tracked the target more accurately than their average performance. We expected that observers would show sensorimotor confidence for eye movements as reflected in an above-chance performance to sort better from worse trials. In addition, we expected to find this confidence ability not only across all trajectories in the experiment, but also within repeats of the same trajectories. Furthermore, we were interested in whether sensorimotor confidence is stable within trials and across different sessions across days.

Methods

Open practices statement

This study was preregistered on OSF February 18, 2022, with the full details available at https://osf.io/6s5gx/?view_only=b3bd3f1b5d8647f1b3181d898731105d. The code for running the experiment and performing the data analysis is available on Github at https://anonymous.4open.science/r/RepSinusoidTask-3484/, with repository release version 2 used for the data analysis reported within this article. All raw and summary data for this experiment and its pilot version are available on the OSF repository at https://osf.io/j9asn/?view_only=843596106542427fa2acdc8fc8f12387. Also included in this repository is the information about the target trajectories, additional results figures, pilot study report, and the data and code to generate the figures presented in this article.

Participants

Thirty participants (mean age of 24.56 years old, 21 female) with normal or corrected-to-normal vision took part in the study. All participants but one author were naive to the design of the experiment. Testing was conducted in accordance with the ethics requirements of the ethics committee of the Justus-Liebig University Giessen. Participants received details of the experimental procedures and gave written informed consent before the experiment and were paid 8 Euros per hour for taking part in the experiment.

Setup

Stimuli were displayed on a Philips Brilliance 288P Ultra Clear monitor (60 × 32 cm, 3840 × 2160 pixel, 60 Hz; Philips, Amsterdam, the Netherlands). Participants sat 70 cm from the monitor with their head stabilized by a chin rest. Gaze was recorded from one eye with a desk-mounted eye tracker (EyeLink 1000 Plus; SR Research, Kanata, ON, Canada) at a sampling frequency of 1000 Hz. Before each block a nine-point calibration was used, and an additional drift check was performed at the start of each trial.

All confidence judgements were entered on a standard computer keyboard. The experiment was conducted using custom-written code in MATLAB version R2021b (The MathWorks, Natick, MA), using Psychtoolbox version 3.0.16 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

Stimuli

We presented a horizontally-moving dot cloud to be gaze-tracked for 6 seconds (see Figure 1). On each frame, four static white dots (diameter: 0.25 degree of visual angle (deg)) were sampled from a 2D circularly-symmetric Gaussian distribution (standard deviation: 2 deg; centered vertically) and presented on a mid-gray background. Dots disappeared and were replaced every frame, giving the stimulus a twinkling quality. The tracking target was the mean of this invisible dot-generating distribution, which followed one of ten precomputed horizontal trajectories. The same ten trajectories were used for all participants and presented multiple times but with new dots sampled for each repeat. The trajectory was vertically mirrored (the same trajectory could start moving to the left or to the right) on half of the repeats to minimize recognition or memorization.

Trajectories were randomly generated sum-of-sinusoid patterns generated from six individual sinusoidal components. Component frequencies always included the 0.05 Hz fundamental frequency, and then five randomly-sampled harmonic frequencies up to 0.4 Hz. The respective component amplitudes were sampled from the uniform distribution $U(-1, 1)$, and assigned in the order such that the largest absolute magnitude amplitude went to the lowest-frequency component of the six different frequency components and so forth until the highest-frequency component was assigned the smallest-magnitude amplitude. We used this procedure to ensure that the target overall had slow movements from side-to-side with added perturbations on that path because of the higher frequencies. Trajectories were then scaled to a maximum horizontal deviation randomly sampled from the Normal distribution $N(12, 1^2)$ in degrees of visual angle. The variation in the maximum horizontal deviation was to introduce spatiotemporal uncertainty in the reversal points in the target trajectory. All components had a phase of 0 to ensure the trajectory started at the screen center. There were two additional criteria a trajectory had to satisfy: (1) the maximum speed had to be less than 30 deg/s and (2) more than 75% of the time, the target had to be moving faster than 2 deg/s. These constraints were imposed to ensure the trajectory was likely to elicit smooth pursuit eye movements. If a sample trajectory did not meet these two criteria, it was discarded and resampled. This procedure resulted in fast, unpredictable, and varied trajectories that remained within the confines of the computer screen (see Figure 2 for an example trajectory).

Task

On each trial, participants performed two tasks. First, they were asked to track with their gaze the mean of four dots that were sampled from a Gaussian distribution centered on one of the 10 trajectories. The trajectory lasted six seconds, and the dots were regenerated every frame at 60 Hz. Then, participants were asked to judge whether their gaze tracking was better or worse than the average of all trials they performed so far. Tracking performance was defined to participants to reflect the spatial distance between the mean of the dots and their gaze at each time point of the trajectory.

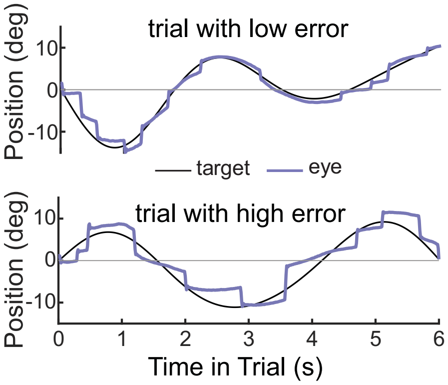

Figure 2. Example tracking behavior. The target trajectory was a sum-of-sinusoids (black smooth curve) and was shown to participants as a dot-cloud whose distribution was centered on the trajectory. The participant's gaze (blue jagged curve) closely followed the target trajectory. Eye movements were a combination of periods of smooth tracking interspersed with saccadic jumps in position. Top shows an example trial with low tracking error, Bottom shows an example trial with higher tracking error chosen from one participant.

The experiment was performed in two one-hour sessions held on different days. Each session began with a calibration procedure for the eye tracker, followed by a brief training task, and then the main task. The purpose of the training was to familiarize the participant with the stimuli and allow them to develop a sense of how well they can track the target. Participants performed five practice trials with no need to report their sensorimotor confidence afterward. Participants were offered to repeat the training phase as many times as they wished before moving on to the main experiment. The trajectory set used during training was not the same as in the main task. In the main task, participants tracked the stimulus and then reported their confidence. In each session, participants performed 20 repeats of the 10 unique trajectories, half of which were horizontally mirrored, in a pseudorandom order. Thus there were 40 repeats per trajectory total and 400 test trials in the whole experiment.

The trial structure was the same for the training and main tasks (see Figure 1). It began with a drift-correction, where the participant had to look at the screen center and press the space key before the trial would start. A blank screen was presented for 255 ms, after which the stimulus would appear at the screen center and move horizontally for six seconds. Then a blank screen was displayed, and participants reported their confidence in their tracking performance as either "better" or "worse" than their average by pressing the right or left arrow key, respectively. On the first trial, participants were given an onscreen reminder of the two confidence-response options. Small breaks were given every 10 trials in the main task for the participant to rest their eyes.

Hypotheses and data analyses

We preregistered four hypotheses and the matching analyses plan:

H1 (main): Participants have above chance metacognitive sensitivity for sorting objectively better eye-tracking performance from objectively worse performance across the entire set of repeated trajectories (i.e., the area under the ROC curve (AUROC) is significantly greater than 0.5; see below and Locke et al., 2020).

H2 (main): Averaged across multiple repetitions of each trajectory, participants report "better" performance significantly more often for the 50% of repeats with lower tracking error than the 50% with higher tracking error for an individual repeated trajectory (i.e., a median split on objective error).

H3 (secondary): The group-averaged temporal metacognitive sensitivity curve shows a recency effect (i.e., greater metacognitive sensitivity for later time bins).

H4 (secondary): Metacognitive sensitivity for eye tracking is stable across days. Specifically, metacognitive sensitivity does not significantly differ between Session 1 and Session 2.

Pilot experiments revealed that participants performed the tracking task with a mixture of smooth pursuit and saccadic eye movements (see Figure 2 for an example tracking trace). We quantified tracking accuracy as the two-dimensional root-mean-squared-error (RMSE) between the gaze position and the tracking target. Both horizontal and vertical error contributed to tracking accuracy. To quantify the relationship between behavior and metacognition, we followed the approach developed by Locke et al. (2020). This involved collecting the distribution of tracking-error separately for trials labeled by the participant with a "better" confidence response and the trials labeled "worse" and measuring the separation between these two distributions. A quantile-quantile comparison gives a receiver operating characteristic (ROC)–like curve, and the degree of separation in the confidence-conditioned distributions is reflected by the AUROC. An area of 0.5 indicates no metacognitive sensitivity, an area of 1 indicates perfect metacognitive sensitivity, and intermediate values indicate intermediate levels of metacognitive sensitivity. We used this approach to evaluate Hypotheses 1, 3, and 4. For Hypothesis 2, we used a novel approach to quantify metacognitive sensitivity that leverages the repeated-trajectories design of the current experiment. For each unique trajectory, we made a median-split of the repeats according to objective tracking accuracy: the half with better accuracy versus the half with worse accuracy according to the RMSE in the trial. Note for this analysis, we included all trials including those with RMSE scores that differed greatly from the participant's mean accuracy. It is then possible to test if significantly higher confidence is given to the better half of repeats, for each trajectory.

Exclusion criteria

Following our preregistered protocol, we had to apply some exclusion criteria. We recruited 30 participants to reach our intended sample size of 27 participants. Three additional participants were recruited to replace two participants who only ever reported "better-than-average" tracking and a third participant who had a technical error occur in eye-tracking during recording, which we were able to correct later in post-processing. We detected an extreme confidence bias in 17 of the total 30 participants (56.7%), where the criterion was more than 75% of confidence responses favoring one the two choice alternatives. Thus, we report the results for the full sample ($n = 30$) as well as the no-extreme bias sample ($n = 13$) for all of the tested hypotheses. We also removed trials with problematic eye recordings using the criteria that the trial RMSE was ±3 SD from that individual participant's mean. This occurred rarely, with an average of 98.6% of trials kept for analysis per participant.

Results

Participants tracked an unpredictable moving trajectory with a mixture of smooth pursuit and saccadic eye movements (see Figure 2 for an example tracking trace). Along our four preregistered hypotheses, we first look at the relationship between the reported performance and actual performance across all trials for an estimate of overall sensorimotor confidence. Then we will look at whether metacognitive sensitivity is influenced by the trajectory characteristics, by computing confidence within repetitions of the same trajectories. Finally, we will look how metacognitive sensitivity changed within a trial or across sessions.

Overall metacognitive sensitivity (Hypothesis 1)

We first examined if participants had above-chance levels of metacognitive sensitivity for their eye-tracking behavior. We compared the distribution tracking accuracy for trials with reported better or worse performance and used that to compute an ROC-like curve (see Methods for more details). An area of 0.5 under this curve indicates no metacognitive sensitivity, an area of 1 indicates perfect metacognitive sensitivity, and intermediate values indicate intermediate levels of metacognitive sensitivity. In the full sample, we could compute the ROC-like curves for 28 of the 30 participants (see Figure 3), with the average area under the ROC-curve (AUROC) being 0.61 ± 0.02 SEM. The two remaining participants only ever responded "better-than-average" and so there was no "worse" error distribution available for the calculation. The level of metacognitive sensitivity in the full sample minus these two participants was significantly above chance ($t(27) = 6.69$, $p < 0.001$, Cohen's $d = 1.27$). This result remains unchanged if we exclude all participants with an extreme confidence bias (0.61 ± 0.02, $t(12) = 4.34$, $p < 0.001$, Cohen's $d = 1.20$; see Methods for more details).

To compare the present results with the metacognitive sensitivity found for hand movements found in a previous study (Locke et al., 2020), we show the metacognition of the average participant in the present study using an equivalent d′ calculation (red curve in Figure 3) and contrast it against the previous results in visuomotor tracking (yellow curve). More precisely, the equivalent d′ is computed as follows: for a given empirical AUROC computed from two conditional error-distributions, that do not follow any particular standard probability distribution, the equivalent d′ is the separation in the means of two standard normal distributions that would result in the exact same AUROC value. Mathematically, the equivalent d′ is computed as follows: $d'_{\text{equiv}} = \sqrt{(2)}\ \Phi^{-1}(\text{AUROC})$. The equivalent d′ for the present eye-tracking study is 0.39. In comparison to the earlier study by Locke et al. (2020) using a similar paradigm with hand-tracking (Figure 3), the d′ for eye movements is lower than the equivalent d′ of 0.66 found for hand movements (although we do acknowledge there are some potential differences in the available cues, see Discussion for more details).

Because there was substantial variance in the level of metacognitive sensitivity across our observers, we performed three additional exploratory analyses. First, we tested whether people with overall better tracking (lower RMSE) were the ones who showed better metacognitive sensitivity. A Spearman correlation between the RMSE and the AUROC did not reveal a significant relationship ($r(26) = 0.23$, $p = 0.24$), indicating no relationship between objective task performance and metacognitive sensitivity in our sample. Second, we looked at whether we could improve metacognitive sensitivity when using a RMSE measurement that takes the response latency into

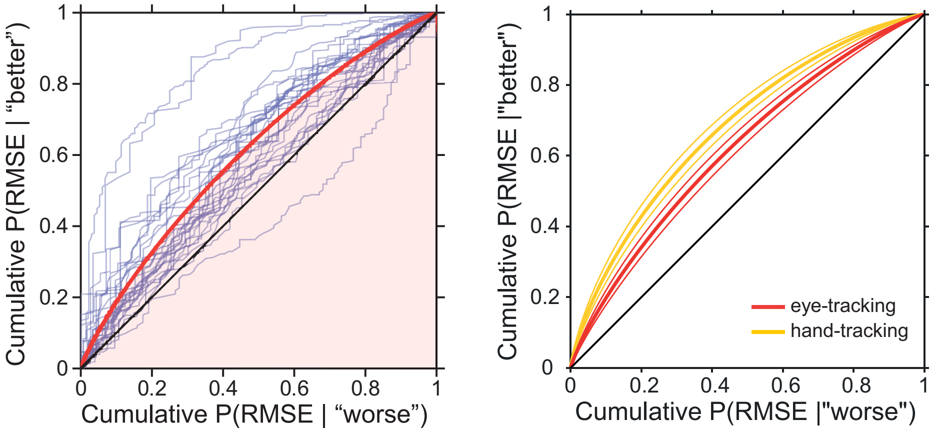

Figure 3. Metacognitive sensitivity results. (Left) ROC-like analysis was applied to the confidence-conditioned tracking-error distributions for each participant (blue), with metacognitive sensitivity reflected by the area under the curve. Two participants had too few "worse" responses for this analysis and therefore are not shown. Tracking error was calculated as the RMSE between gaze and target across the entire tracking period. Solid diagonal line: theoretical curve for a no-sensitivity observer. Red shading shows average metacognitive sensitivity across participants. (Right) Red curve demonstrating the equivalent d′ for the current eye tracking experiment. The superior metacognitive sensitivity of the average observer in the hand-tracking study of Locke et al. (2020) is shown in yellow for comparison. Thin lines represent standard error the mean.

account. For that purpose, we computed a RMSE error on each trial that was shifted by the response lag given in a certain trial estimated via cross-correlation. On average observers lagged the target movements by 71 ± 6 ms. Using this tracking latency-corrected RMSE, we obtained an AUROC of 0.58 ± 0.01 SEM, a value again comparable to the level we obtained for the overall RMSE measurement. Third, we tested whether metacognitive sensitivity improved when we computed the AUROC not from the average RMSE error across all trials, but from the relative error computed on the recent n-back trials to determine whether performance was better or worse than the average (see also Locke et al., 2020). For that we allowed the n-back trials that lead to the highest AUROC to vary per participant and observed that performance overall stayed quite comparable to using the average RMSE. Best performance was achieved when integrating 21 ± 5 trials, but the gain in AUROC performance compared to that obtained with the average across all trials was minimal (0.64 ± 0.02 SEM compared to 0.61 ± 0.02 SEM). This suggests that observers overall had a stable representation of their average tracking performance.

Repeated-trials analysis (Hypothesis 2)

A difficulty with the metacognitive-sensitivity analysis presented above is that it is unclear whether there are some difficulty cues observable in each stimulus that participants can use to infer their tracking performance. For example, stimuli that move faster are often harder to track. This cue-based method of inferring tracking performance is in contrast to a more monitoring-based approach where the participant is genuinely keeping track of the errors they make as the trial progresses. One way we propose to discard the strategy that participants simply estimate how difficult a stimulus is, is to present the same stimulus multiple times. Tracking error will vary between these repeats, and we can measure if the confidence responses of the participant can detect these performance variations. In other words, did observers also show more sensorimotor confidence in better tracking for the same trajectory? For this purpose, the errors on all the repetitions of a trajectory were sorted, and then we split trials into halves of lower and higher errors (see Methods for more details). On average, participants were 8.21% ± 1.79% more likely to rate a more-accurate tracking trial as "better-than-average" in their confidence report. This pattern of higher confidence for more accurately-tracked trials was observable for all 10 unique trajectories (see Figure 4). This difference in confidence reports is significant ($t(29) = 4.57$, $p < 0.001$, Cohen's $d = 0.83$), and the results are unchanged if the extremely biased participants are excluded from the analysis (13.27% ± 3.30%, $t(12) = 4.03$, $p = 0.002$, Cohen's $d = 1.12$). This supports a monitoring-based explanation of sensorimotor confidence for eye tracking.

Temporal dynamics of metacognition (Hypothesis 3)

We also performed a temporal analysis of the metacognitive sensitivity of observers to test if there

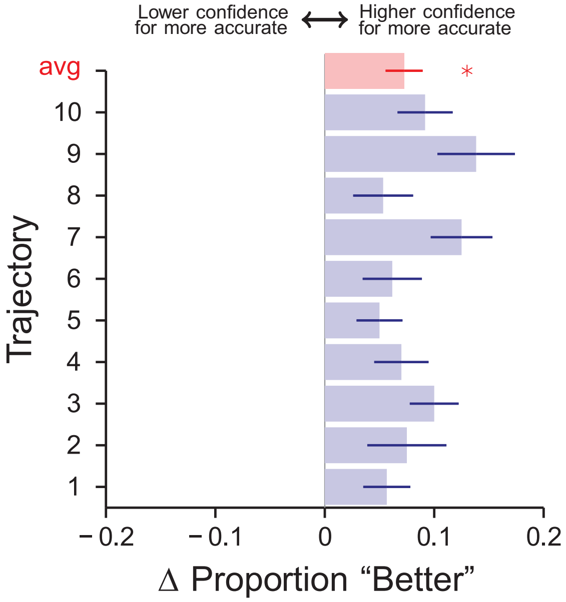

Figure 4. Results of the repeated-trials analysis. For all repeated trajectories, participants on average responded "better" more often for the half of trajectory repeats with the better tracking (i.e., lower RMSE). Individual trajectory results are shown in blue, and the average of all trajectories in red. Error bars: SEM.

was a recency bias in metacognition. Previously, it has been shown that confidence in hand-tracked targets disproportionally relies on the tracking error at the end of the trial (Locke et al., 2020). We followed the procedure of this previous study to compute the temporal metacognitive sensitivity curves by separately computing the AUROC using the error in each of the six one-second time bins (see Figure 5A). We then tested for a recency effect by measuring if the average AUROC in the 5- and 6-second bins was significantly greater than the average AUROC of the three- and four-second bins. The first two seconds of tracking were ignored in this analysis as tracking error can take a couple of seconds to stabilize and we did not observe a systematic difference in tracking accuracy across the remaining trial. The AUROC was significantly higher by 0.05 ± 0.01 for the final two seconds as compared to the middle two seconds ($t(27) = 4.00$, $p < 0.001$, Cohen's $d = 0.76$) in the full population minus the two participants for which we could not compute their AUROC. However, this analysis did not reach significance in the sample excluding extremely-biased participants (0.03 ± 0.02, $t(12) = 1.60$, $p = 0.13$, Cohen's $d = 0.44$).

Metacognitive sensitivity across sessions (Hypothesis 4)

Our final analysis considered whether any metacognitive learning occurred between the first and second sessions of the experiment that were conducted on different days. To test this hypothesis, we computed the metacognitive sensitivity separately for each session. For both the larger sample and the extreme-bias-excluded sample, we did not find any significant difference in metacognitive sensitivity between the sessions. The differences were -0.004 ± 0.02 ($t(27) = -0.27$, $p = 0.79$, Cohen's $d = -0.05$) and -0.01 ± 0.01 ($t(12) = -0.76$, $p = 0.46$, Cohen's $d = -0.21$), respectively. This indicates that metacognitive sensitivity was stable across the experiment. As an additional exploratory measurement, instead of focusing on the average, we also looked at the individual variation in metacognitive sensitivity (see Figure 5B). We observed that sensitivity stayed comparable ($r(26) = 0.58$, $p = 0.001$), suggesting temporally stable differences in sensitivity across participants.

Discussion

The goal of the present study was to test whether humans are capable of actively monitoring the performance of their own eye movements. Across all trials we found a low but significant agreement between the reported and actual tracking performance of observers, demonstrating sensorimotor confidence for eye movements (Figure 3). When calculating the metacognitive sensitivity separately for multiple repeats of each of the different target trajectories (Figure 4), sensitivity was present for each of them suggesting that the effect was not related to stimulus characteristics. When looking into the temporal evolution of the effect, we observed that the reports of performance were mostly related to tracking performance late in the trial (Figure 5A), but that overall sensitivity was stable across sessions measured on separate days (Figure 5B).

Heuristics or performance monitoring

While our results suggest a limited, but successful performance monitoring, the information was used inefficiently. One factor that needs to be carefully considered is how other potential and even irrelevant cues could have affected the subjective reports. It is known that confidence reports can be affected by heuristics and assumptions about the task (Mole et al., 2018; Navajas et al., 2017; Spence, Dux, & Arnold, 2016). For example, despite similar performance, participants are biased to report higher confidence if the task involves their own active movement in comparison to a passive movement or a visual task (Charles, Chardin, & Haggard, 2020). Confidence can even be manipulated independently of performance, for example when motion characteristics (de Gardelle

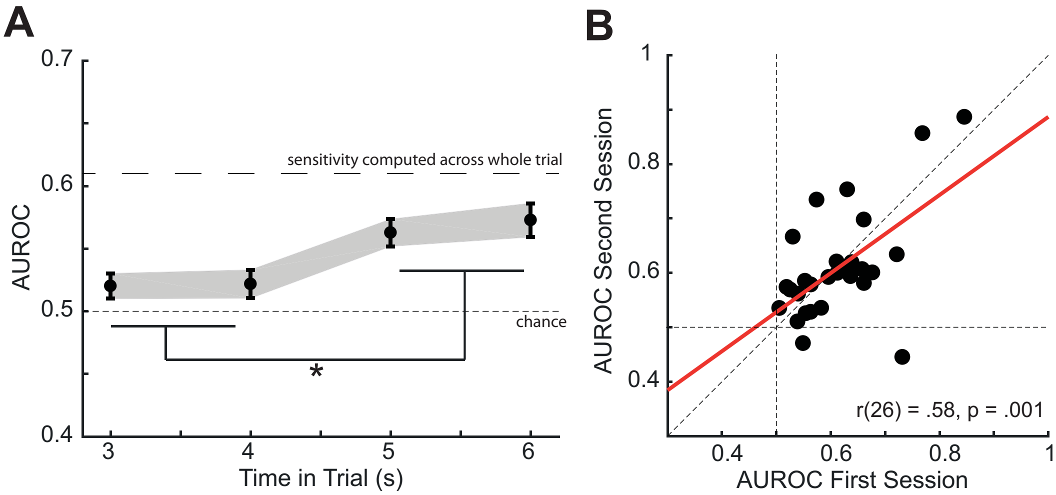

Figure 5. Temporal dynamics and stability of metacognitive sensitivity. (**A**) The average AUROC for all observers across time within in a trial. The AUROC is computed separately for each second of the trial while ignoring the first two seconds. The shaded area shows the SEM. (**B**) The AUROC for each participant, represented by separate dots, for the first and second sessions, which were separated by multiple days. Red line shows a linear regression fitted to the data.

& Mamassian, 2015) or stimulus visibility (Maniscalco, Peters, & Lau, 2016; Rausch, Müller, & Zehetleitner, 2015) are changed across trials. For our current results, we believe it is unlikely that heuristics are able to explain our sensorimotor confidence results due to three points: First, we observed the effect consistently across all trajectories which rules out different judgements depending on their characteristics. Second, our AUROC is not affected by potential biases (similarly to d′ in signal detection theory). Third, with simple heuristics, there should not have been a stronger influence of tracking error towards the end of the trial. This recency effect demonstrates that the subjective report was related to performance monitoring and that some moments were treated differently than others (Locke et al., 2020). Locke and colleagues could show that such a recency effect also occurs for unpredictable trial durations and interpreted it as a signature of the temporal accumulation of a confidence signal with “leaky accumulation” and memory-limitations (see their article for a detailed discussion). Together, these points suggest, that although heuristics can play an important role for confidence judgements, in our task metacognitive sensitivity seems to be mainly driven by limited and inaccurate but successful performance monitoring.

Confidence across sensory and sensorimotor processing

So far, previous studies mainly have considered the contribution of movements in perceptual confidence (Fleming & Daw, 2017; Kiani, Corthell, & Shadlen, 2014; Yeung & Summerfield, 2012), or used motor behavior as index of perceptual confidence (Dotan, Meyniel, & Dehaene, 2018; Patel, Fleming, & Kilner, 2012; Resulaj, Kiani, Wolpert, & Shadlen, 2009). However, these studies mostly focused on simple behavior and did not include an assessment of the potential role of additional motor variability. Extending the work from Locke and colleagues (2020) on hand movements, we show that people also can make successful confidence judgements about eye movements and their variability. This supports the idea that confidence could serve as a common currency to evaluate performance across tasks (Ais, Zylberberg, Barttfeld, & Sigman, 2016; de Gardelle, le Corre, & Mamassian, 2016; Faivre, Filevich, Solovey, Kühn, & Blanke, 2018). This is reflected in results showing that confidence judgements can be correlated across different perceptual tasks (Charles et al., 2020). Similar to measures of perceptual confidence (Ais et al., 2016; de Gardelle & Mamassian, 2015; Navajas et al., 2017), we also observed stable individual differences in the sensitivity of observers that stayed consistent across multiple days. Thus it would be interesting to explicitly test whether variability in sensorimotor confidence is also linked to other instances of confidence measurements.

Differences in sensitivity between eye and hand movements

When comparing metacognitive sensitivity measured for eye movements with the previous work using a similar paradigm for hand movements (Locke et al., 2020), sensitivity for eye movements was lower (Figure 3). Although the paradigms were similar, there was a potentially important difference: In the hand movement task, there was an additional visual representation of the hand position on the screen, which could have allowed for a more direct assessment of tracking performance. A similar cue based on eye position was not possible because it would have added a salient and dynamic distractor that would have interfered with tracking performance. While an accurate eye position signal should provide the same information, having an additional cue to monitor performance might have made the evaluation of hand movement performance easier, which could explain the higher sensitivity.

Next to the difference in performance feedback, an additional explanation could be a difference in the quality of evaluating and monitoring eye and hand movements. While eye (Spering & Carrasco, 2015; Tavassoli & Ringach, 2010) and hand movements (Fourneret & Jeannerod, 1998; Goodale et al., 1986; Prablanc & Martin, 1992) can be controlled by information we are perceptually unaware of, there seems to be a particular unawareness about how we move our eyes (Goettker et al., 2018; Nieuwenhuis et al., 2001; van Zoest & Donk, 2010). Such differences could be based on differences in control: Although eye movements are controlled by only a few muscles and do not need much energy, successful hand movements require much more effort. In addition, although eye movements only increase our understanding of the environment, hand movements can modify it and sometimes produce injuries from contact with the environment. Therefore there might have been greater evolutionary pressure to be better at monitoring one's hand than eye movements.

Additionally, although we observe metacognitive sensitivity for the eye, it might not be based on a judgement of actual motor performance, but a judgment of a more general state. Earlier work showed that humans are able to estimate their own saccade latency reasonably well (Vencato & Madelain, 2020), however such a performance could have been achieved without actually knowing about when the eye movement began. Instead, participants could be judging the state of attention and preparedness at the beginning of the trial (Tomassini, Spinelli, Jacono, Sandini, & Morrone, 2015; VanRullen, 2016). Similarly, knowing the current level of attention could serve as a proxy for the quality of the movement, without actually having access to the movement signal. The recency effect can also be explained by simply focusing on the level of concentration people were paying toward the end of the trial. Therefore, whereas for the eye and hand movements an overall sensitivity for current state of the observer could be available, for hand movements additional monitoring and evaluation processes might be accessible.

Conclusions

Our goal was to understand whether observers can judge the accuracy of their own eye movements. We found across all trials and even within repetitions of the same trajectory, that participants did show a sensitivity for distinguishing good from bad trials, thus displaying metacognitive sensitivity. The observed sensitivity was mostly related to performance towards the end of the trial, which matched previous reports when judging the quality of hand movements. However, overall sensitivity for eye movements was significantly lower than that found for hand movements. These results provide an additional piece of evidence for sensorimotor confidence, and open outstanding questions about why it differs across movements and how it could be related to purely perceptual and other instances of confidence.

Keywords: eye movements, sensorimotor confidence, confidence

Acknowledgments

The authors thank Nina Küpke for her help in data collection. S.L. & P.M. were supported by an Anneliese Maier Award from the Alexander von Humboldt Foundation awarded to PM. A.G., K.G. & P.M. were supported by the Deutsche Forschungsgemeinschaft (DFG; project number 222641018-SFB/TRR 135 Project A1).

Commercial relationships: none.

Corresponding author: Alexander Goettker.
Email: alexander.goettker@psy.jlug.de.
Address: Abteilung Allgemeine Psychologie,
Justus-Liebig University Giessen, Giessen 35394,
Germany.

[*] AG and SML contributed equally as first authors.

References

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition, 146*, 377–386, https://doi.org/10.1016/j.cognition.2015.10.006.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Bridgeman, B., Van der Heijden, A. H. C., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements.

Behavioral and Brain Sciences, 17(2), 247–258, https://doi.org/10.1017/S0140525X00034361.

Carpenter, R. H. S. (1988). *Movements of the eyes* (2nd rev.). London: Pion Limited.

Charles, L., Chardin, C., & Haggard, P. (2020). Evidence for metacognitive bias in perception of voluntary action. *Cognition, 194*, 104041, https://doi.org/10.1016/j.cognition.2019.104041.

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. (2015). Weighting mean and variability during confidence judgments. *PLOS ONE, 10*(3), e0120870, https://doi.org/10.1371/journal.pone.0120870.

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Research, 36*(7), 985–996, https://doi.org/10.1016/0042-6989(95)00203-0.

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition, 171*, 112–121, https://doi.org/10.1016/j.cognition.2017.11.001.

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *The Journal of Neuroscience, 38*(2), 263–277, https://doi.org/10.1523/JNEUROSCI.0322-17.2017.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing*. Oxford: Oxford University Press.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review, 124*(1), 91–114, https://doi.org/10.1037/rev0000045.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia, 36*(11), 1133–1140, https://doi.org/10.1016/S0028-3932(98)00006-2.

Gegenfurtner, K. R. (2016). The interaction between vision and eye movements. *Perception, 45*(12), 1333–1357, https://doi.org/10.1177/0301006616657097.

Goettker, A., Braun, D. I., Schütz, A. C., & Gegenfurtner, K. R. (2018). Execution of saccadic eye movements affects speed perception. *Proceedings of the National Academy of Sciences of the United States of America, 115*(9), 2240–2245, https://doi.org/10.1073/pnas.1704799115.

Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature, 320*(6064), 748–750, https://doi.org/10.1038/320748a0.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84*(6), 1329–1342, https://doi.org/10.1016/j.neuron.2014.12.015.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*(14), 1–16.

Klingenhoefer, S., & Bremmer, F. (2011). Saccadic suppression of displacement in face of saccade adaptation. *Vision Research, 51*(8), 881–889, https://doi.org/10.1016/j.visres.2010.12.006.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25–26), 3559–3565.

Leigh, R. J., & Zee, D. S. (2015). *The Neurology of Eye Movements*. Oxford: Oxford University Press, https://doi.org/10.1093/med/9780199969289.001.0001.

Liston, D. B., & Stone, L. S. (2014). Oculometric assessment of dynamic visual processing. *Journal of Vision, 14*(14), 12.

Locke, S. M., Mamassian, P., & Landy, M. S. (2020). Performance monitoring for sensorimotor confidence: A visuomotor tracking study. *Cognition, 205*, 104396, https://doi.org/10.1016/j.cognition.2020.104396.

Maniscalco, B., Peters, M. A. K., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, & Psychophysics, 78*(3), 923–937, https://doi.org/10.3758/s13414-016-1059-x.

Mole, C. D., Jersakova, R., Kountouriotis, G. K., Moulin, C. J., & Wilkie, R. M. (2018). Metacognitive judgements of perceptual-motor steering performance. *Quarterly Journal of Experimental Psychology, 71*(10), 2223–2234, https://doi.org/10.1177/1747021817737496.

Morvan, C., & Maloney, L. T. (2012). Human visual search does not maximize the post-saccadic probability of identifying targets. *PLoS Computational Biology, 8*(2), e1002342, https://doi.org/10.1371/journal.pcbi.1002342.

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour, 1*(11), 810–818, https://doi.org/10.1038/s41562-017-0215-1.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38*(5), 752–760, https://doi.org/10.1111/1469-8986.3850752.

Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proceedings of the Royal Society B: Biological Sciences, 279*(1748), 4853–4860, https://doi.org/10.1098/rspb.2012.1847.

Pélisson, D., Alahyane, N., Panouillères, M., & Tilikete, C. (2010). Sensorimotor adaptation of saccadic eye movements. *Neuroscience & Biobehavioral Reviews, 34*(8), 1103–1120, https://doi.org/10.1016/j.neubiorev.2009.12.010.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pierrot-Deseilligny, C., Rivaud, S., Gaymard, B., & Agid, Y. (1991). Cortical control of memory-guided saccades in man. *Experimental Brain Research, 83*(3), 607–617, https://doi.org/10.1007/BF00229839.

Prablanc, C., & Martin, O. (1992). Automatic control during hand reaching at undetected two-dimensional target displacements. *Journal of Neurophysiology, 67*(2), 455–469, https://doi.org/10.1152/jn.1992.67.2.455.

Rausch, M., Müller, H. J., & Zehetleitner, M. (2015). Metacognitive sensitivity of subjective reports of decisional confidence and visual experience. *Consciousness and Cognition, 35*, 192–205, https://doi.org/10.1016/j.concog.2015.02.011.

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.

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision, 11*(5), 9.

Spence, M. L., Dux, P. E., & Arnold, D. H. (2016). Computations underlying confidence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance, 42*(5), 671–682, https://doi.org/10.1037/xhp0000179.

Spering, M., & Carrasco, M. (2015). Acting without seeing: Eye movements reveal visual processing without awareness. *Trends in Neurosciences, 38*(4), 247–258, https://doi.org/10.1016/j.tins.2015.02.002.

Tavassoli, A., & Ringach, D. L. (2010). When your eyes see more than you do. *Current Biology, 20*(3), R93–R94.

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research, 76*, 31–42, https://doi.org/10.1016/j.visres.2012.10.012.

Tomassini, A., Spinelli, D., Jacono, M., Sandini, G., & Morrone, M. C. (2015). Rhythmic oscillations of visual contrast sensitivity synchronized with action. *Journal of Neuroscience, 35*(18), 7019–7029, https://doi.org/10.1523/JNEUROSCI.4568-14.2015.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*(8), 291–297, https://doi.org/10.1016/j.tics.2008.04.010.

van Beers, R. J. (2007). The sources of variability in saccadic eye movements. *Journal of Neuroscience, 27*(33), 8757–8770, https://doi.org/10.1523/JNEUROSCI.2311-07.2007.

van Zoest, W., & Donk, M. (2010). Awareness of the saccade goal in oculomotor selection: Your eyes go before you know. *Consciousness and Cognition, 19*(4), 861–871, https://doi.org/10.1016/j.concog.2010.04.001.

VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences, 20*(10), 723–735, https://doi.org/10.1016/j.tics.2016.07.006.

Vencato, V., & Madelain, L. (2020). Perception of saccadic reaction time. *Scientific Reports, 10*(1), 17192, https://doi.org/10.1038/s41598-020-72659-3.

Vullings, C., & Madelain, L. (2018). Control of saccadic latency in a dynamic environment: Allocation of saccades in time follows the matching law. *Journal of Neurophysiology, 119*(2), 413–421, https://doi.org/10.1152/jn.00634.2017.

Vullings, C., & Madelain, L. (2019). Discriminative control of saccade latencies. *Journal of Vision, 19*(3), 16, https://doi.org/10.1167/19.3.16.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1310–1321, https://doi.org/10.1098/rstb.2011.0416.