Post-Decision Gaze Reflects Implicit Confidence in Decisions

Yonatan Stern[1], Ophir Netzer[1], Danny Koren[1], Yair Zvilichovsky[1], Uri Hertz[1,2,3], Roy Salomon[1,2,3*]

[1] School of Psychological Sciences, University of Haifa.

[2] The Integrated Brain and Behavior Research Center (IBBRC), University of Haifa.

[3] Department of Cognitive Science, University of Haifa.

* Corresponding Author: Roy Salomon, royesal@gmail.com

Keywords: metacognition, oculomotor behavior, probabilistic learning, decision making, virtual reality

This PDF file includes:

Main Text
Figures 1 to 4
Tables 1

Abstract

Our decisions are naturally accompanied by a feeling of how confident we are in their correctness. While typically measured experimentally via self-reported judgments, decision confidence is typically expressed through implicit sensorimotor behavior. Combining VR-based probabilistic learning paradigms with computational modeling across one exploratory and two pre-registered experiments (total N = 115), we examined whether post-decision gaze dynamics reflects an implicit second-order evaluation of decisions. Gaze direction was consistently aligned with participants' decisions about a target's upcoming location, and fluctuations in this alignment tracked the latent probability of the decision. Moreover, gaze exhibited computational hallmarks of confidence, reflecting a second-order ocular assessment of the decision. Directly comparing gaze-based confidence and explicit, reported confidence revealed a modest yet significant correlation, suggesting that the two are partially overlapping second-order processes. Notably, reported confidence showed higher metacognitive sensitivity and increased serial dependence across trials, while gaze-based confidence offered a more immediate, trial-specific readout of internal confidence. Importantly, we found that under increased uncertainty, gaze and decisions diverged more frequently, reflecting strategic "hedging" behavior. These findings suggest that post-decision gaze reflects an implicit, confidence-like behavior that partially overlaps with confidence judgments. Thus, the embodied interaction between decisions and ocular behavior supports flexible and adaptive behavior when learning and acting in a dynamic environment.

Introduction

Our decisions are usually accompanied by a "sense of knowing," an internal estimate of confidence in their correctness. Confidence is a result of a second-order process that evaluates first-order performance such as on-task decisions. The standard widely used experimental method for directly and explicitly probing this internal sense of confidence is to request a report of one's degree of confidence (e.g., via a self-reported rating; Maniscalco & Lau, 2012; Rahnev et al., 2020). However, this judgment may not fully capture all aspects of one's sense of confidence. For example, confidence judgments have been found to be influenced by interoceptive and sensorimotor processes and to reflect additional sensorimotor information that extends beyond the decision itself (Allen et al., 2016; Kiani et al., 2014). Furthermore, prompted confidence judgments introduce additional evaluation and response processes that do not necessarily occur outside experimental settings and may be unrelated to the decision itself (Shekhar & Rahnev, 2021). In contrast, various sensorimotor processes that naturally occur during the decision process have been shown to immediately and implicitly reflect the second-order process, in a manner akin to confidence judgments. Notably, reaction times are robustly correlated to confidence ratings (Rahnev et al., 2020; Weidemann & Kahana, 2016). Motor adaptation (Dotan et al., 2018; Faivre et al., 2020), speech prosody (Goupil & Aucouturier, 2021), and pupil dilation (Colizoli et al., 2018; Urai et al., 2017) have also been linked to decision certainty and confidence. Moreover, the modulation of sensorimotor behavior by one's sense of confidence may capture additional facets of internal confidence and promote adaptive behavior (Faivre et al., 2020; Goupil et al., 2021).

Confidence is critical for adaptive functioning in a dynamic environment (Fleming, 2024; Nelson, 1984), modulating one's learning rate (Drugowitsch et al., 2019), driving subsequent information-seeking behavior (Desender et al., 2018), and shaping reward maximization strategies (van den Berg et al., 2016). Although the study of confidence and metacognition has historically focused primarily on the domains of perceptual decisions and memory (Rahnev, 2021; Rouault et al., 2018), interest in decision confidence during learning has grown in the past decade (Meyniel et al., 2015). During probabilistic learning, confidence judgments track a decision's underlying likelihood distribution (Meyniel et al., 2015), and reflect the crossing of an internal decision boundary when choosing between competing options (Hertz et al., 2018). In line with confidence's role in assessing the accuracy of one's decisions, confidence ratings robustly distinguish correct and incorrect decisions, and humans

show significant metacognitive sensitivity (Maniscalco & Lau, 2012; Rahnev et al., 2020). However, confidence judgments also diverge from first-order decision process in significant ways. They integrate priors more optimally (Constant et al., 2023) and include parallel or additional post-decision computational processes (Balsdon et al., 2020; Fleming, 2024; Fleming & Daw, 2017). Confidence judgments also exhibit biases such as "leakiness," and are influenced by previous trials even across tasks and domains (Mei et al., 2023; Rahnev et al., 2015, 2020), diverging from a pure assessment of one's current decision. Research focusing on confidence reports has extensively characterized confidence judgments, however it remains an open question whether sensorimotor behavior that reflects a second-order assessment of a decision exhibits similar characteristics as reported confidence.

Notwithstanding extensive research on confidence and sensorimotor behavior, the relation between gaze and confidence in humans remains surprisingly underexplored, despite gaze's close connection to cognitive processes underlying decisions (Spering, 2022). Gaze is typically directed toward features and regions of the environment perceived as important (Niv, 2019; Spering, 2022). For example, viewing naturalistic scenes the learned statistical regularities of scene composition guide attention and visual search, enabling efficient scanning, filtering and processing of complex visual input (for reviews see Nobre & Kastner, 2014; Wolfe, 2020). Thus, gaze is shaped by one's experience, providing a window to an internal model of the environment that one has learned (Bar, 2004; Helmholtz, 1924), and is likely modulated by the strength of one's expectations. Research examining how gaze dynamics is modulated by experience has found that gaze dynamics and decisions rely on similar learning mechanisms (Leong et al., 2017; Vossel et al., 2014). Non-human primate studies have demonstrated that gaze metrics reflect both decision certainty (Seideman et al., 2018) and the reevaluation of one's decision during learning (Ferro et al., 2024). Although decisions and ocular behavior often occur in tandem, previous work in humans has largely examined them separately, leaving the nature of their interaction and contribution to adaptive behavior unclear.

The current study examines whether post-decision gaze reflects an ocular second-order assessment of decisions similar to reported confidence judgments. Across three experiments, participants performed a probabilistic learning task in which they predicted the location of a visual target based on probabilistic cues. In the following sections, we also refer to these predictions as decisions. Leveraging this learning paradigm in which we computationally model the probability underlying the prediction of the target's location

across trials together with the tendency of gaze to be directed towards regions of interest, we examine the relation between decisions and post-decision gaze. Specifically, we ask: (1) Does gaze reflect a second-order, confidence-like evaluation of decisions? (2) How does gaze-based confidence relate to reported confidence judgments and do they exhibit similar characteristics? (3) How do gaze and decisions interact to support adaptive behavior in a dynamic environment?

In two initial experiments (one exploratory, one pre-registered), we found that post-decision gaze direction converged robustly with the predicted target location. Furthermore, the degree of convergence between the decision and gaze tracked the latent probability of the decision and exhibited the statistical hallmarks of confidence. Thus, post-decision gaze reflects a second-order assessment of the decision, which we refer to as gaze-based confidence. In a third experiment, we directly compared self-reported confidence ratings and gaze-based confidence and found a significant but modest correlation, indicating that the two partially overlap. Each exhibited distinct characteristics contributing to adaptive learning: Confidence judgments exhibited superior metacognitive sensitivity, while gaze-based confidence was less influenced by previous trials' confidence and reflects a more momentary readout. Finally, examining the possible synergistic benefits of two partially overlapping measures, we found the decision's predicted location and gaze direction significantly diverge under higher uncertainty, serving to "hedge" one's decision under uncertainty.

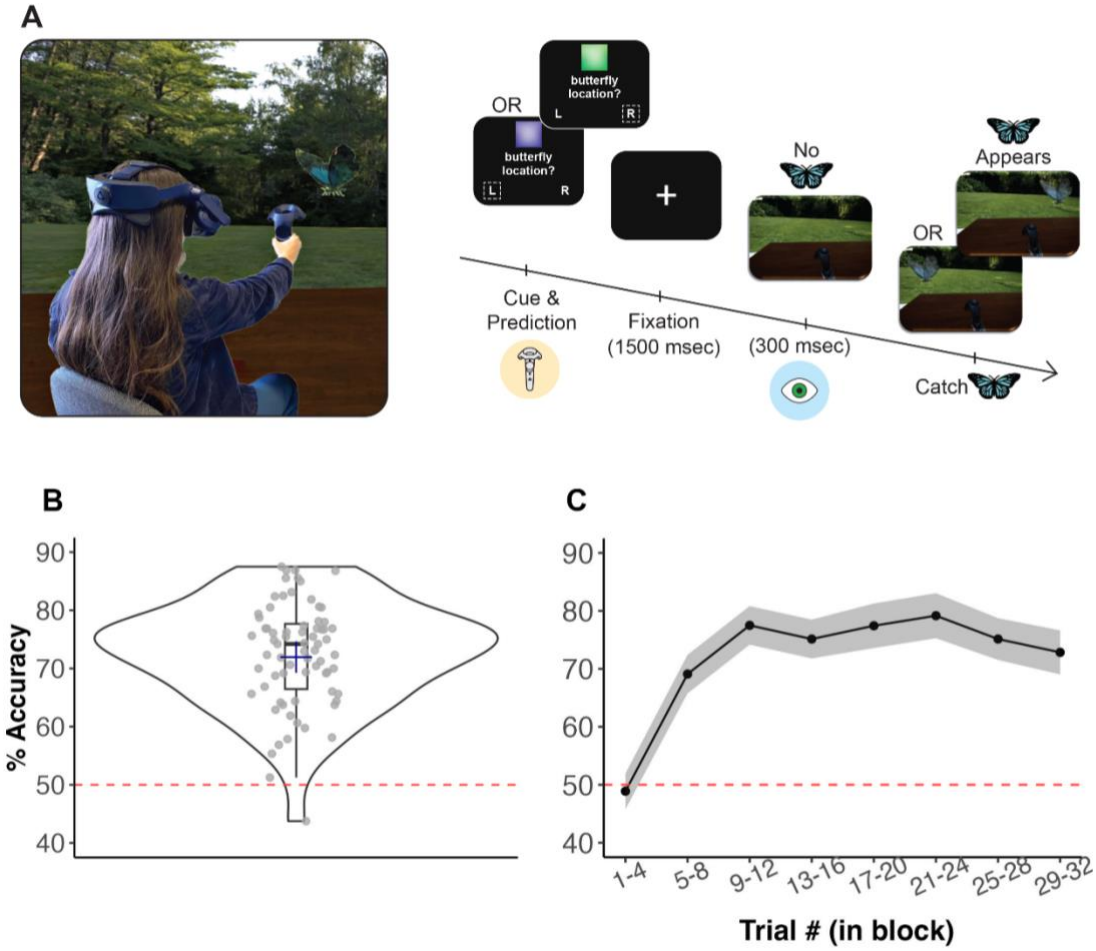

Figure 1. Experimental paradigm. Panel A: (Left) Experimental setup. In VR, participants sat at a table and reached toward the target (i.e., butterfly) when it appeared. (Right) Trial Flow. A colored cue prompted participants to predict the upcoming target's location (Right/Left) via the controller's touchpad, with no display change post-response. Following the response, a fixation cross preceded their entry into the virtual environment, where, after 300 milliseconds, a target appeared. Horizontal gaze direction during this 300 milliseconds interval was used to examine ocular expectations. Following the butterfly's appearance participants were instructed to reach towards it as quickly and accurately as possible. **Panels B–C: Participants' decisions exhibited classical probabilistic learning signatures across experiments 1 and 2.** Following our interest in participants' learning of the underlying rule, we define accuracy as whether the response conformed to the underlying probabilistic rule. **(B) Accuracy exceeded chance**. Gray dots denote individual participants (including those subsequently excluded by pre-registered criteria), the dashed red line indicates chance level (50%), and the blue cross represents the group mean. **(C) Learning curve within blocks.** Averaging trials across blocks, accuracy was low at the start of the block, and quickly increased as participants learn the underlying rule till a plateau was reached. Black line is group average and gray area represents the 95% confidence interval across participants.

Results

Participants learn probabilistic rule

We confirmed that participants' decisions exhibit well-established signatures of probabilistic learning. Across all experiments, accuracy was significantly higher than chance

(M_{accuracy} = 59.6%, 95% CI= [58.6, 60.6], t_{110} = 18.8, p < .001, Cohen's d = 11.1; see Fig. 1B and Fig. S1), and was significantly higher on trials in which a valid, as opposed invalid, cue was given ($M_{\text{difference}}$ = 44.6%, 95% CI= [40.9, 48.4], t_{110} = 2.9, p < .001, Cohen's d =2.3; see Fig. S1). Finally, accuracy within blocks of trials exhibited a classic learning curve shape (see Fig. 1C and Fig. S2). Thus, participants were engaged and learned the underlying probabilistic rule. Given our primary interest in the learning of the underlying rule and following our pre-registration, subsequent analyses define accuracy as whether the prediction conformed to the underlying rule (i.e., "rule accuracy"). Rule accuracy is distinct from whether the prediction matched the target's actual location, which may be partially random due to invalid cues.

Gaze converges with predictions and tracks their underlying probability

Supporting the convergence of predictions and ocular expectations, while awaiting target appearance participants looked in the direction of their prediction for a significantly longer duration than in the opposite direction. Thus, the difference in gaze direction in trials in which the prediction that the target was on the Right as opposed to on the Left was significant and large (*Exp. 1*: $M_{\text{normalized gaze difference}}$ = 0.95, 95% CI= [0.80, 1.10], t_{24} = 660.77, p < .001, Cohen's d =2.56; *Exp. 2*: $M_{\text{normalized gaze difference}}$ = 0.90, 95% CI= [0.79, 1.01], t_{24} = 922.83, p < .001, Cohen's d =3.04; see Fig. 2A). Furthermore, 54 of 56 participants showed a significant difference at the individual level (see Fig. S3). Likewise, converging evidence was obtained by binarizing gaze direction into an ocular response (i.e., Right/Left; see SM for further analyses of binary ocular responses). The ocular response was identical to the prediction in 73% of trials (95% CI = [70%, 75%]; see Fig. 2B and Fig. S4). Thus, as expected, ocular predictions robustly align with the decisions' predictions.

We then examined whether fluctuations in the magnitude of coupling between gaze direction and predictions track latent learning processes underlying the decisions. We calculated the correlation between Gaze Prediction, a continuous measure of the alignment between gaze direction and predicted location, and the underlying probability of the prediction made in each trial (i.e., $p(\text{Choice}_t)$). This probability is derived from the well-established family of Rescorla-Wagner (RW) learning models. While we pre-registered the use of a simple RW model, after comparing the models (see Fig. S5) we opted for a RW model that included a stickiness parameter (Culbreth et al., 2016; Daw et al., 2011; see SM for similar results with simple RW model). The trial-by-trial correlation of $p(\text{Choice}_t)$ and

Gaze Prediction was calculated per participant (see Fig. 2C). In line with our pre-registration, the group distribution of Pearson correlation coefficients was significantly greater than zero in both experiments (*Exp. 1*: Mean $\rho = .13$, 95% CI = [.08, .19], $t_{24} = 4.84$, $p < .001$, Cohen's $d = 0.97$, *Exp. 2*: Mean $\rho = .11$, 95% CI = [.05, .16], $t_{24} = 4.02$, $p = .001$, Cohen's $d = 0.82$; Fig. 2D and Fig. S6). Thus, when the underlying probability of the prediction was greater, gaze was more strongly directed in the direction of the predicted target location. Accordingly, Gaze Prediction's fluctuations mirror the latent probability of the prediction.

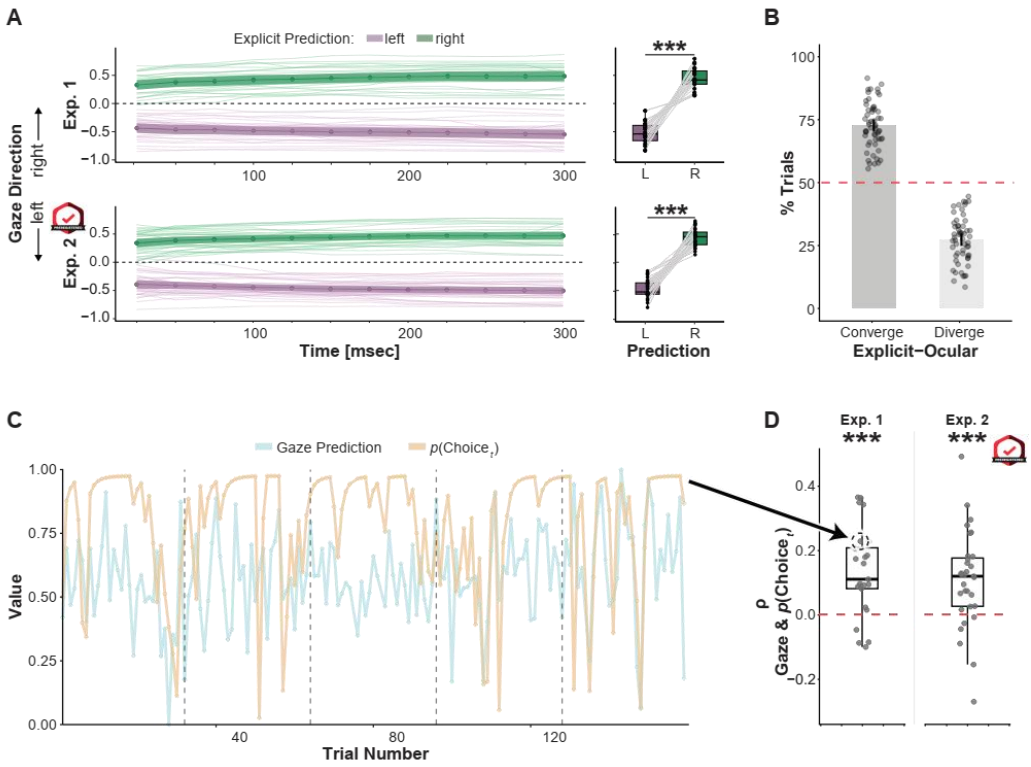

Figure 2. Convergence of predictions and post-decision gaze. Panel A: Gaze Direction is aligned with the prediction's direction. (Left) Time course of normalized post-decision gaze direction in 300 msec intervals prior to target appearance. (Right) Mean normalized gaze direction averaged across intervals and trials is robustly aligned with the predicted location. **Panel B: Convergence and divergence of explicit prediction's and gaze's direction across participants**. Gaze was binarized into a Right/Left direction analogous to explicit predictions. Across both experiments, gaze's direction and the prediction's direction were identical in ~70% of the trials. The black line reflects a 95% confidence interval, and gray dots represent individual participants. **Panel C: Correlation between trial-by-trial Gaze Prediction and latent probability of the decision p(Choice $_t$).** A random participant's trajectory across the experiment of p(Choice $_t$), the probability of making the decision derived from the computational model, and Gaze Prediction, gaze magnitude normalized by the prediction's direction. Note that for the clarity of the visualization, Gaze Prediction values were bound to the limits of p(Choice $_t$) between 0 and 1. **Panel D: Distribution of individual participants' correlation coefficients between p(Choice $_t$) and Gaze Prediction.** In both experiments the coefficients' distributions were significantly positive. Thus, gaze was more strongly directed to the prediction's direction as its underlying probability increased.

Gaze Prediction exhibits computational hallmarks of confidence in decision

Gaze's close relation to the decision's probability (i.e., $p(\text{Choice}_t)$) resonates with the prevalent conceptualization of confidence as the probability of the decision being correct (De Martino et al., 2013; Fleming & Daw, 2017; Hertz et al., 2018; Sanders et al., 2016). Accordingly, we tested whether Gaze Prediction exhibits three computational hallmarks of subjective confidence proposed by Sanders et al. (2016): (1) a monotonic relation with accuracy, (2) a "folded X" pattern in which Gaze Prediction increases with a rise in trial evidence for correct trials and decreases for incorrect trials, and (3) trials with a greater Gaze Prediction magnitude show a stronger association between accuracy and evidence than do trials with low magnitude Gaze Prediction (see Fig. 3 and SM for further details of the analysis). Achieving the first hallmark, we found a positive monotonic relation between accuracy and Gaze Prediction. The group correlation coefficient distribution was significantly greater than zero (*Exp. 1*: Mean $\rho = .33$, 95% CI = [.14, .53], $t_{24} = 3.5$, $p = .002$, Cohen's $d = 0.7$; *Exp. 2*: Mean $\rho = .28$, 95% CI = [.11, .45], $t_{24} = 3.4$, $p = .002$, Cohen's $d = 0.61$; Fig. 3A and Fig. S7). To test the second hallmark and following our pre-registration, we used trial-by-trial expectation strength (v) derived from the learning model as a proxy for trial evidence. In correct trials, Gaze Prediction increased with v, whereas for incorrect trials gaze decreased, exhibiting a "folded X" pattern (see Fig. 3B). The interaction effect of accuracy and v was significant (*Exp. 1*: $\chi^2_{(1)} = 9.32$, $p = .002$; Exp. *2*: $\chi^2_{(1)} = 20.67$, $p < .001$; see Eq. 1 in Table 1 and Table S5 and see Table S6 for pairwise comparisons). Finally, fulfilling the third hallmark, high Gaze Prediction trials exhibited a stronger association between v and accuracy than did low gaze trials (see Fig. 2C), and the interaction was significant (*Exp. 1*: mean (SEM) $\beta = 0.15$ (0.05), $t_{24} = 3.16$, $p = .002$; *Exp. 2*: mean (SEM) $\beta = 0.16$ (0.04), $t_{30} = 3.60$, $p < .001$; see Eq. 2 in Table 1). Thus, Gaze Prediction meets the three computational hallmarks of subjective confidence, and exhibits confidence-like characteristics.

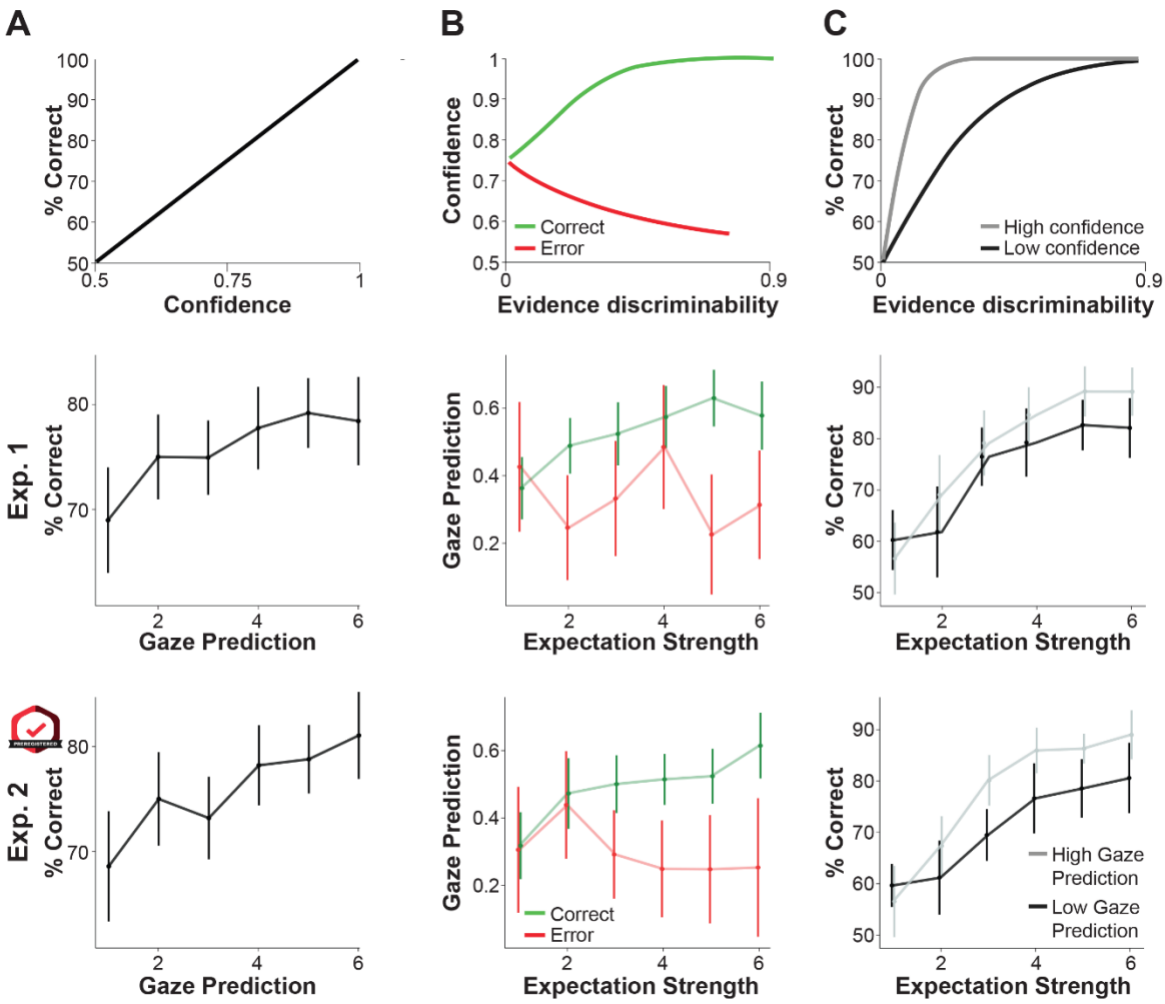

Figure 3. Gaze Prediction exhibits computational hallmarks of confidence. Top panel: Subjective confidence exhibits three computational hallmarks of statistical confidence: (a) A monotonic relation between confidence and accuracy, (b) a "folded X" pattern, such that as evidence increases, confidence in correct trials (green) increases, whereas for incorrect trials (red) it decreases, and (c) Steeper slope between evidence strength and accuracy for high confidence trials (gray) as opposed to low confidence trials (black). Figure adapted from Sanders et al., (2016). **Middle and bottom panels:** Gaze Prediction, the degree of alignment between gaze's direction and magnitude and the explicit prediction's direction, exhibits computational hallmarks of confidence in an exploratory (Experiment 1) and pre-registered replication (Experiment 2) experiment. In these analyses, Gaze Prediction was used in lieu of subjective confidence, and expectation strength (v) served as a proxy for the trial's evidence.

Table 1. Regression Models

Equation #	Hypothesis	Model Equation
1	Gaze Prediction exhibits "folded X" pattern (hallmark 2).	Gaze Prediction ~ accuracy*expectation strength + (1|participant)

2	Stronger relation between accuracy and expectation for high Gaze Prediction trials (hallmark 3).	logit(accuracy) ~ Gaze Prediction*expectation strength + (1|participant)
3	Confidence judgments are linked to Gaze Prediction beyond other factors.	Gaze Prediction ~ confidence rating +accuracy +p(Choice $_t$) +δ_{t-1} +(1| participant)

Gaze-based confidence is significantly but modestly correlated to confidence judgments

In Experiment 3, after predicting the target's location, participants rated their confidence in their decision (see SM for complete results of the experiment's pre-registered analyses). To examine the relation between confidence and latent learning processes, for each participant we correlated the confidence rating with the underlying probability of each trial's decision (i.e., p(Choice $_t$)), and found that the group distribution of the correlation coefficients was significant (Mean ρ = .38, 95% CI = [.33, .44], t_{31} = 14.47, p < .001, Cohen's d = 2.56; see Figure S17 for correlations between confidence ratings and other latent learning variables). In addition, 30 of the 32 participants exhibited a significant correlation at the individual level. Thus, in line with previous work (Hertz et al., 2018; Meyniel et al., 2015), trial-by-trial fluctuations in confidence ratings closely tracked the underlying probability of the decision made.

Turning to examine the relation between confidence ratings and gaze-based confidence, we calculated the trial-by-trial correlation between confidence ratings and Gaze Prediction for each participant. We found that the group-level distribution of correlation coefficients was significantly greater than zero (see Figure 4A, Mean group ρ = .076, 95% CI = [.038, .11], t_{31} = 4.1, p < .001, Cohen's d =0.72), which indicates a significant degree of convergence between the two. To confirm that the correlation between gaze-based confidence and confidence ratings is not driven by confounding factors (e.g., p(Choice $_t$), response accuracy), a linear mixed model incorporating these factors (see Table 1 Eq. 3) revealed a significant relationship between gaze and confidence ($\chi^2_{(1)}$ = 9.5, p = .002), even after controlling for these factors. Nonetheless, despite the significant association between confidence ratings and gaze-based confidence, the correlation magnitude was modest and below our pre-registered criterion (r >. 1) for defining the two processes as similar.

Confidence ratings exhibits superior metacognitive sensitivity relative to gaze-based confidence

After identifying the partial overlap of these two types of confidence, we examined their distinct characteristics. To confirm that the two measures are directly comparable, we converted Gaze Prediction into a 5-point rating that is analogous to the confidence ratings (see SM for further details of the conversion). Comparing the metacognitive sensitivity of reported and gaze-based confidence via the difference in ratings in correct and incorrect trials (Δ confidence; see SM for full justification of the use of this measure), both measures exhibited significant metacognitive sensitivity, with higher confidence in correct trials compared to incorrect trials (i.e., Δ confidence > 0; see Figure 4B, *Reported Δ Confidence*: Mean = 0.58, 95% CI = [.47, .70], t_{31} = 10.44, p < .001, Cohen's d =1.85, *Gaze-based Δ Confidence*: Mean = 0.10, 95% CI = [.03, .17], t_{31} = 3.01, p = .004, Cohen's d =0.54). However, reported confidence ratings were significantly more sensitive than gaze-based confidence (t_{31} = 7.7, p < .001, Cohen's d =1.36). Interestingly, reported and gaze-based metacognition were not significantly correlated across participants, and Bayesian statistics offered anecdotal evidence for the lack of correlation (r_{31} = .09, p = .63, BF_{10} = .43; see Fig. S8). This lack of correlation should be interpreted with caution given the small sample size, yet it suggests that distinct mechanisms underlie reported and gaze-based metacognition.

To further explore the mechanisms behind each type of confidence, we examined whether they are influenced by the previous trial's confidence and exhibit serial dependence (Mei et al., 2023; Rahnev et al., 2015, 2020). For each participant, we calculated the autocorrelation between adjacent trials' confidence (i.e., lag 1) for reported and gaze-based confidence separately. Reported confidence ratings were significantly autocorrelated across trials, and the group distribution of the coefficient was significantly greater than zero ($M_{autocorrelation}$ = 0.35, 95% CI = [.27, .43], t_{31} = 8.77, p < .001, Cohen's d =1.54). In contrast, the group-level autocorrelation of gaze-based confidence was greater than zero across experiments, with results varying by significance (see SM for further details). Importantly, directly comparing the autocorrelation coefficient for each participant revealed that confidence ratings are significantly more strongly autocorrelated compared to gaze-based confidence (see Figure 4C, t_{31} = 7.85, p < .001, Cohen's d = 1.39; see SM for similar results obtained while controlling for additional factors such as previous trial accuracy). These results suggest that reported confidence relies on a mechanism that is temporally extended and influenced by previous trials, while gaze-based confidence offers a more short-term, trial-specific readout of decision certainty.

To summarize**,** reported and gaze-based confidence reflect two partially overlapping second-order assessments of decisions. They are significantly correlated, and both exhibit metacognitive sensitivity, suggesting shared underlying processes. However, reported confidence has significantly greater metacognitive sensitivity and exhibits robust serial dependence, pointing to a temporally extended assessment of one's internal confidence. In contrast, gaze-based confidence reflects a more immediate readout of the current state (see SM results for converging results obtained by examining binary ocular responses and their respective learning mechanisms). These findings highlight that while both reported and gaze-based confidence reflect second-order processes evaluating a decision, they have distinct characteristics, and their integration may be key to adaptive behavior.

Gaze and prediction directions diverge under increased uncertainty

To examine how decisions and gaze contribute to adaptive behavior, we analyzed the convergence and divergence of the predicted target location and gaze direction. As previously noted, their direction was identical in ~70% of trials. In the remaining trials, we found that divergence was more likely under increased uncertainty. Exploratory trials with low certainty (i.e., $p(\text{Choice}_t) < .5$) exhibited significantly more divergence than exploitation trials with high certainty (see Figure 4D, $M_{\text{divergence rate in exploratory trials}}$: 35%, 95% CI = [31%, 40%]; $M_{\text{divergence in exploitation trials}}$: 26%, 95% CI = [23%, 28%]; $t_{55} = 4.13$, $p = .001$, Cohen's $d = 0.56$; see SM and Fig. S13). This finding suggests that the divergence of directions enables one to "hedge" one's decisions under increased uncertainty. In contrast, trials in which prediction and gaze directions converged exhibited significantly higher prediction accuracy ($M_{\text{accuracy in converging trials}}$: 78%, 95% CI = [76%, 79%]; $M_{\text{accuracy in diverging trials}}$: 71%, 95% CI = [68%, 74%]; $t_{55} = 4.75$, $p < .001$, $d = 0.63$; see SM Fig. S12). Thus, fluctuations in the alignment of gaze and decision enable adaptive behavior based on one's decision certainty.

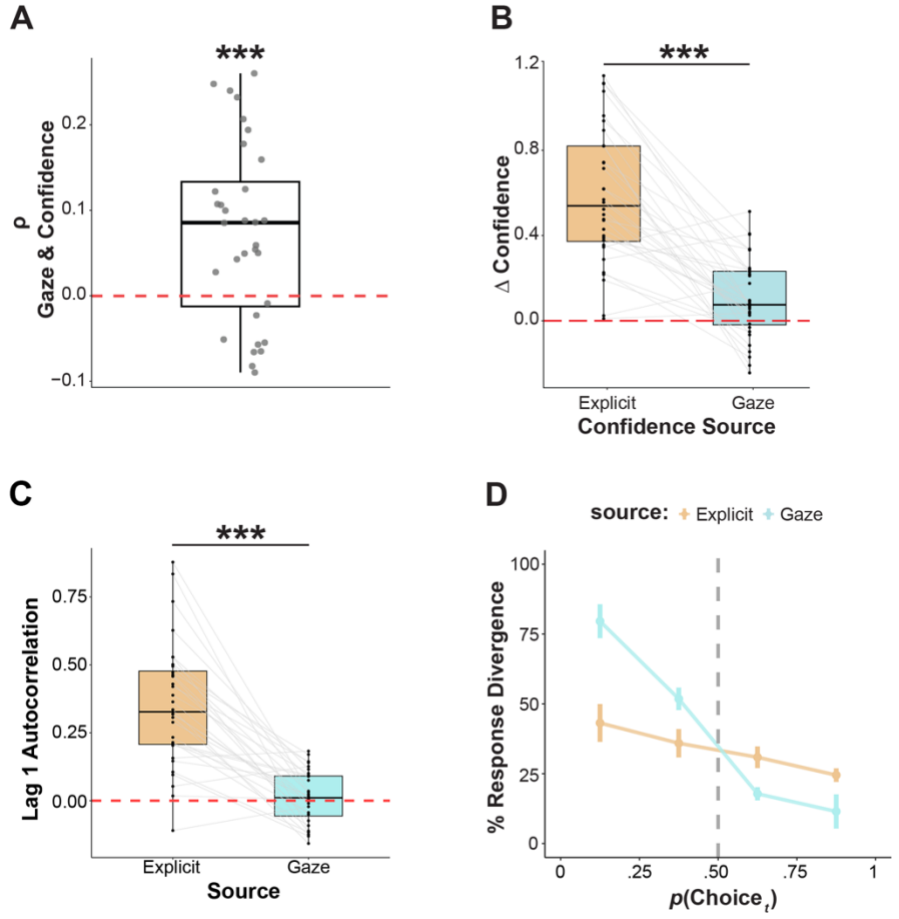

Figure 4. Comparison of reported and gaze-based confidence. Panel A: Distribution of individual participants' correlation coefficients between reported explicit confidence ratings and gaze-based confidence in Experiment 3. Mean correlation was 0.07 and was significantly greater than zero ($p < .001$), yet its magnitude was modest and smaller than our pre-registered criterion for defining the two processes as similar ($r >. 1$). **Panel B: Comparison of metacognitive sensitivity (i.e., Δ confidence) of reported and gaze-based confidence.** Reported confidence ratings were significantly more sensitive than were gaze ($p < .001$). Importantly, both sources of confidence exhibited significant sensitivity and their Δ confidence was greater than chance (dashed red line; both *p*s < .001). **Panel C: Comparison of serial dependence in reported and gaze-based confidence.** Serial dependence was assessed by calculating the lag-1 autocorrelation of confidence ratings. Reported confidence exhibited significantly greater serial dependence than gaze-derived confidence ($p < .001$), indicating that ratings were more temporally extended and influenced by confidence in the preceding trial. **Panel D: Increased divergence of decision's and gaze's direction under increased uncertainty.** Decision and gaze "response" direction were more likely to differ in exploratory trials in which the response made had a lower probability (i.e., $p(\text{Choice}_t)$) that was calculated based on decision (brown) or ocular (turquoise) "response" trajectories. Both showed a similar effect: In exploratory trials (i.e., $p(\text{Choice}_t) < .5$) the two responses were more likely to differ in contrast to exploitation trials (i.e., $p(\text{Choice}_t) > .5$).

Discussion

This study investigated how post-decision gaze dynamics relate to confidence in one's decisions during probabilistic learning. As expected, across three experiments, gaze direction consistently aligned with decisions. Furthermore, fluctuations in gaze-decision alignment tracked the underlying probability of the decision and exhibited the computational hallmarks of confidence. Thus, post-decision gaze reflects a second-order ocular assessment of one's decisions during probabilistic learning. A direct comparison of gaze-based and reported confidence revealed a significant yet modest correlation, suggesting that these two second-order processes only partially overlap. Reported confidence was more metacognitively sensitive and exhibited robust serial dependence effects reflecting a more temporally extended assessment of internal confidence. In contrast, gaze-based confidence reflected a more momentary readout of internal confidence. Finally, examining how gaze and decision jointly contribute to adaptive behavior, we found that they tended to diverge under uncertainty, reflecting a strategic "hedging" of responses.

A central finding of the current study is that post-decision gaze dynamics reflects a second-order evaluation of one's decision. This relation may serve a functional role. Post-decision gaze is part of an anticipatory motor preparation process that entails directing attention towards the expected location of an upcoming target in preparation for reaching towards it. Importantly, anticipatory gaze is sensitive to the decision's certainty and the strength of one's expectation. Gaze's sensitivity to decision certainty resonates with pre-decision gaze patterns that are sensitive to the subjective value of options during value-based decisions (Krajbich et al., 2010; Thomas et al., 2019). Given its critical role in guiding attention, gaze is closely linked to internal evaluation processes that are essential in shaping behavior. An alternative but complementary explanation for the relation is representational. Gaze may reflect the reinstatement of internal models of the environment. Activation of stimuli in working memory results in an involuntarily gaze shift to a stimulus' location even when location is task irrelevant (Van Ede et al., 2019; van Ede et al., 2020). Similarly, during the use of abstract concepts such as a "mental number line," gaze patterns reflect the concept's spatial organization (Viganò et al., 2024). Thus, the strength of the internal representation of the upcoming target's location may automatically drive gaze towards the expected location regardless of task demands. In the current study, we cannot distinguish between these complementary explanations since gaze always followed the participant's response and preceded a reaching movement. It therefore remains unclear to what degree contextual task demands drive the coupling of decisions and gaze. Future studies may address

this question by examining gaze dynamics during passive viewing, in which case their coupling could be explained by a representational connection, or instances in which the speed of one's motor response is of critical importance (e.g., playing tennis) and according to the functional explanation it should result in a stronger coupling.

In the current study, post-decision gaze exhibited the computational hallmarks of confidence (Sanders et al., 2016). Whether these hallmarks are sufficient to equate gaze with sensorimotor confidence resonates with similar debates regarding other measures such as neural activation and animal behavior that exhibit these hallmarks (Bang & Fleming, 2018; Rausch & Zehetleitner, 2019; Schmack et al., 2021). Importantly, both explicit confidence and Gaze Prediction tracked the probability of the choice made (i.e., $p(\text{Choice}_t)$), and this may be a core feature of second-order processes during learning (Hertz et al., 2018; Meyniel et al., 2015). Following the significant, albeit modest, correlations between explicit confidence and gaze-based confidence in Experiment 3, we argue that gaze exhibits confidence-like characteristics: It reflects a partially overlapping second-order sensorimotor assessment that resembles reported confidence yet should not be equated with it.

Examining the unique characteristics of reported confidence and gaze-based confidence, we found that the two differed in their serial dependence and metacognitive sensitivity. Although providing a full mechanistic explanation for these two differences is beyond the scope of the current paper, these differences point to key differences in these second-order processes. In line with previous work (Mei et al., 2023; Rahnev et al., 2015, 2020), confidence judgments exhibited significant serial dependence effects. This presumably reflects a "continuity field" in the metacognitive domain that creates a temporally extended assessment of one's performance, inducing a shift in one's internal confidence criteria that persists across trials and even tasks (Rahnev et al., 2015). Serial dependence is considered a source of metacognitive inefficiency in tasks whose trials are independent of each other (Rahnev et al., 2015; Shekhar & Rahnev, 2021), yet during probabilistic learning in which trials are highly dependent, serial dependence may enable a more coherent sense of overall performance that is consistent across trials. Thus, reported confidence may reflect a higher-order process that aspires to self-consistency across trials (Koriat, 2012), in contrast to gaze-based confidence that reflects a more momentary readout of one's confidence. Turning to metacognitive sensitivity, although gaze exhibited significant metacognition characteristics, it remains unclear what biases caused its reduced sensitivity. One possibility is that gaze-based

confidence calculations were based on a more volatile perception of the environment that affects confidence (Behrens et al., 2007). At the start of the experiment, participants were informed that there is a fixed ratio (i.e., 3:1) of valid and invalid trials. Reported confidence ratings may, unlike gaze-based confidence, be able to use this "meta-knowledge" of the environment, which in turn enhances their sensitivity. Additional possibilities may be that gaze-based confidence is more susceptible to noise or may rely on different parts of the evidence accumulation process, compared to self-reported confidence (Balsdon et al., 2020; Kiani & Shadlen, 2009). Future work may attempt to systematically characterize the different sources of metacognitive inefficiency for these different processes.

Directly comparing reported and gaze-based confidence is also related to ongoing debates on whether metacognition is a unitary domain-general or domain-specific capacity (Arbuzova et al., 2023; Rouault et al., 2018). The lack of correlation between two measures metacognitive sensitivity, despite both being made on the same decision, is more easily interpreted within a domain-specific framework in which there is no assumed relation between different types of metacognitions. The lack of correlation should be interpreted with caution given the small sample size and that the addition of confidence ratings in Experiment 3 slightly altered gaze dynamics and weakened the relation between gaze and confidence. Nevertheless, our results raise the intriguing possibility that even uncorrelated second-order processes may complement one another. In line with this possibility, we found that the divergence of the decision and gaze directions in an uncertain environment enabled a "hedging" of responses. This finding underscores the importance of understanding different cognitive processes within the broader context of parallel processes whose integration enhances adaptive behavior. Furthermore, understanding how systems interact is relevant in understanding psychopathology in which this integration may be disrupted, and is especially relevant for schizophrenia spectrum disorders in which an aberrant interaction between explicit and implicit sensorimotor systems is postulated to be a core deficit (Krugwasser et al., 2021; Nelson et al., 2014; Salomon et al., 2021; Stern et al., 2020).

The current study highlights how cognitive capacities are a multimodal process that spans both decisions and the implicit sensorimotor behavior that evaluates it. Examining the interplay between these distinct processes revealed instances of convergence, such as the case of gaze reflecting a confidence-like assessment of decisions, as well as divergence, such as the unique characteristics of this gaze-based confidence. The integration of these distinct

processes enables adaptive and flexible behavior, and understanding it is critical in developing a more ecological and embodied understanding of human behavior.

Materials and Methods

We conducted three experiments: an initial exploratory Experiment 1, and two pre-registered experiments (Experiment 2: https://osf.io/p4hyq, 09.05.2023; Experiment 3: https://osf.io/ngsx6. 17.07.2023). The pre-registered analysis plan was followed, unless otherwise stated. For consistency, we retrospectively executed the pre-registered analysis plan for Experiment 1. The primary inclusion criteria were: (1) response accuracy (i.e., the prediction matches the target's location) above 55%. This value was selected to ensure that parameter fitting was reliable and that participants understood the task; (2) The VR and eye-tracking recording was valid in at least 70% of trials.

Participants

For Experiment 1 we collected data from 35 participants; 25 participants (19 female, M_{age} = 23.9, SD = 4 years) were included in analysis. For Experiment 2, we pre-registered 40 participants, anticipating that ~30 would provide sufficient statistical power to replicate Experiment 1's effects (see https://osf.io/p4hyq for power analysis). The final sample included data from 31 participants (21 female, M_{age} = 23.7, SD = 3.4 years). For Experiment 3, we pre-registered 40 participants, and the final sample size included data from 32 participants (23 female, M_{age} = 25.3, SD = 5.3 years).

All participants were right-handed, with normal or corrected-to-normal vision and self-reported no-history of neurological impairments, and were recruited via social media. All participants gave written informed consent and were compensated (50 NIS per hour or course credits) for their participation. Experiment 1 and 2 were performed at Bar-Ilan University and Experiment 3 was performed at Haifa University. Each university's Institutional Review Board approved this study: Approval ISU202107003 for Experiments 1 and 2 and Approval HU448/19 for Experiment 3.

Hardware and Setup

Participants wore an HTC Vive Pro Eye head-mounted display connected to a high-performance desktop, which projected a virtual environment created in Unity (2018.3.2) at a 90 Hz refresh rate and tracked their hand and eye movements. Responses were recorded

through the controller's trackpad, and eye movements were monitored using the built-in Tobii trackers (~90Hz; see Applebaum et al. [2024] and SM for additional technical details).

Experimental Procedure

We adapted a well-established previously modeled probabilistic learning task (Frässle et al., 2021; Harrison et al., 2021; Iglesias et al., 2013) to VR. In 160 trials, participants were presented with one of two cues (i.e., a square that varied in its color), and they predicted the upcoming target's location (Right/Left) by pressing a button on the controller. Following their prediction, a fixation cross appeared (duration 1.5 seconds), and then participants entered a virtual environment and the target (a butterfly) appeared after a brief 300 msec delay. Participants were instructed to reach towards the target once they identified it, and a virtual hand tracked their actual hand movement (see Fig. 1A for trial flow). Participants' goal was to learn the underlying rule governing the cue-location association and adjust their predictions accordingly. Participants were informed that: (1) In 75% of the trials, the cue is valid (it follows an underlying rule) and in the remaining trials the cue is invalid; (2) The underlying rule is contingent between the two cues (e.g., if cue A indicates that the target will appear on the right, cue B necessarily indicates that it will appear on the left); (3) Throughout the experiment, the underlying rule may change multiple times, requiring participants to detect these changes. The rule was changed five times during the experiment, with each block lasting between 29 and 35 trials (see SM for further details). In Experiment 3 participants performed an identical task and additionally they rated their confidence in their prediction on a five-point vertical slide. After completing the probabilistic learning task, participants also performed a sensorimotor conflict detection task to assess their sense of agency (Applebaum et al., 2024), and completed several self-report questionnaires pertaining to psychotic-like and anomalous self-experiences (Cicero et al., 2017; Foa et al., 2002; Loewy et al., 2011), whose results are beyond the scope of the current work.

Measures

In each trial, we recorded participants' prediction ("The target will appear in Right or Left location") and their post-decision ocular expectation of the target's location. Predictions were assessed for *actual accuracy* (whether the target appeared as predicted) and *rule accuracy* (the prediction conformed to the underlying rule). Our primary focus was on rule accuracy due to our interest in the rule-learning process, except when noted otherwise.

Ocular expectations were derived from a pre-registered 300 msec interval in the VR environment, after participants made their decision but before the target appeared. We normalized and averaged horizontal gaze directions across this interval for each trial, yielding a single continuous Gaze Direction value per trial. From Gaze Direction, we derived: (1) *Gaze Response* (Right or Left), a binary measure that is analogous to the response made via the controller; and (2) *Gaze Prediction*, a continuous measure obtained by normalizing Gaze Direction by the decision's direction. This measure indicates the alignment between the prediction and post-decision gaze direction, with positive values denoting gaze in the direction of the predicted location and negative values indicating gaze in the opposite direction (see SM for transformation details).

Modeling

To analyze learning dynamics throughout the experiment, we employed a model-based approach. The model space consisted of: (1) the well-established Rescorla-Wagner (RW) model (Rescorla, 1972); (2) a Rescorla-Wagner model with an additional stickiness parameter (sRW) (Culbreth et al., 2016; Daw et al., 2011); and (3) a noisy Win-Stay Lose-Switch (WSLS) model (Wilson & Collins, 2019).

The RW model consists of a perceptual and decision model. In the perceptual model each option's value (v) is updated by the prediction error (δ)—the discrepancy between the previous trial's predicted and actual outcomes (r)—modulated by the learning rate (α), a free parameter that controls the extent of updating by prediction errors (Eq. 4A). Lower α values suggest slower learning over more trials. The value (v) from the perceptual model informs the decision model, which employs a SoftMax function to calculate the choice probability (i.e., $p(\text{Choice}_t)$). The function's steepness and randomness are dictated by the inverse temperature (β), another free parameter, reflecting decision consistency (Eq.4B).

(Eq. 4A) $$v_{t+1} = v_t + \alpha * \delta_t$$

Where $$\delta_t = r_t - v_t$$

(Eq. 4B) $$\boldsymbol{P}._{Choice} = \frac{e^{\beta * v}}{e^{\beta * v} + e^{\beta * (1-v)}}$$

The "sticky" sRW model extends the traditional RW model by adding a perseveration parameter (ρ) to quantify the tendency to use the rule used in the previous decision (Culbreth et al., 2016; Daw et al., 2011). This parameter is included in the decision model (Eq. 5), multiplied by *rep (a)*, an indicator function that equals 1 for decisions that apply the rule used in the previous decision and 0 otherwise. Importantly, decisions are coded in a cue-contingent manner (cue A and a "right" decision are coded identically to cue B and a "left" response), which implies that perseveration reflects that a decision adheres to the same rule across trials, rather than mere motor repetition. Notably, the RW model is nested within the sRW model when the parameter ρ is set to zero.

Eq. 5 $$P._{Choice} = \frac{e^{\beta * \boldsymbol{v} + \rho * rep(\mathrm{a})}}{e^{\beta * \boldsymbol{v} + \rho * rep(\mathrm{a})} + e^{\beta * (1-\boldsymbol{v}) + \rho * rep(\mathrm{a})}}$$

Unlike the RW model, which uses a free parameter to determine the integration window, in the WSLS model (Eq. 6) decisions are based solely on the previous trial's outcome, and a free parameter ($\boldsymbol{\varepsilon}$) incorporates probabilistic noise in the decision-making.

Eq. 6 $$P._{Choice} \begin{cases} 1 - \varepsilon \; if \; C_{t-1} \; correct \\ \varepsilon \; if \; C_{t-1} \; incorrect; \end{cases}$$

For all models, stimuli and responses were encoded within a unified contingency space to minimize free parameter estimates, in line with our pre-registration and previous work (Harrison et al., 2021; Iglesias et al., 2013). Model parameters were estimated via the Bayesian adaptive direct search (BADS) toolbox (Acerbi & Ji, 2017), optimizing each model's log likelihood. Parameter recovery across models was reliable (see Fig. S19). Model comparison was performed via the Bayesian Information Criteria (BIC; Schwarz, 1978). All models showed good model recovery (see Fig. S20).

Linear mixed models were fitted using the lme4 package (Bates et al., 2014), and likelihood ratio tests were used to compare the full model against a version without the targeted effect to obtain p-values.

Pre-registrations, analysis code, and notebooks are available at:

Data are available at:

https://osf.io/6nj8u/?view_only=0852199aacb94825a3ef81c760b972a6

Bibliography

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Advances in neural information processing systems*, *30*.

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *Elife*, *5*, e18103.

Applebaum, A., Netzer, O., Stern, Y., Zvilichovsky, Y., Mashiah, O., & Salomon, R. (2024). The Body Knows Better: Sensorimotor signals reveal" Suboptimal" inference of the Sense of Agency in the human mind. *bioRxiv*, 2024–02.

Arbuzova, P., Maurer, L. K., & Filevich, E. (2023). Metacognitive domains are not aligned along a dimension of internal-external information source. *Psychonomic Bulletin & Review*, *30*(3), 1125–1135.

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1), 1753.

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *115*(23), 6082–6087.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Cicero, D. C., Neis, A. M., Klaunig, M. J., & Trask, C. L. (2017). The Inventory of Psychotic-Like Anomalous Self-Experiences (IPASE): Development and validation. *Psychological Assessment*, *29*(1), 13–25. https://doi.org/10.1037/pas0000304

Colizoli, O., de Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, *8*(1), 13702.

Constant, M., Pereira, M., Faivre, N., & Filevich, E. (2023). Prior information differentially affects discrimination decisions and subjective confidence reports. *Nature Communications*, *14*(1), 5473.

Culbreth, A. J., Westbrook, A., Daw, N. D., Botvinick, M., & Barch, D. M. (2016). Reduced model-based decision-making in schizophrenia. *Journal of Abnormal Psychology*, *125*(6), 777.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778.

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

Drugowitsch, J., Mendonça, A. G., Mainen, Z. F., & Pouget, A. (2019). Learning optimal decisions with confidence. *Proceedings of the National Academy of Sciences*, *116*(49), 24872–24880.

Faivre, N., Vuillaume, L., Bernasconi, F., Salomon, R., Blanke, O., & Cleeremans, A. (2020). Sensorimotor conflicts alter metacognitive and action monitoring. *Cortex*, *124*, 224–234.

Ferro, D., Cash-Padgett, T., Wang, M. Z., Hayden, B. Y., & Moreno-Bote, R. (2024). Gaze-centered gating, reactivation, and reevaluation of economic value in orbitofrontal cortex. *Nature Communications*, *15*(1), 6163.

Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology*, *75*, 241–268.

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

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485–496. https://doi.org/10.1037/1040-3590.14.4.485

Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., Harrison, S. J., Heinzle, J., Iglesias, S., & Kasper, L. (2021). TAPAS: an open-source software package for translational neuromodeling and computational psychiatry. *Frontiers in Psychiatry*, *12*, 680811.

Goupil, L., & Aucouturier, J.-J. (2021). Distinct signatures of subjective confidence and objective accuracy in speech prosody. *Cognition*, *212*, 104661.

Goupil, L., Ponsot, E., Richardson, D., Reyes, G., & Aucouturier, J.-J. (2021). Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature. *Nature Communications*, *12*(1), 861.

Harrison, O. K., Köchli, L., Marino, S., Luechinger, R., Hennel, F., Brand, K., Hess, A. J., Frässle, S., Iglesias, S., & Vinckier, F. (2021). Interoception of breathing and its relationship with anxiety. *Neuron*, *109*(24), 4080–4093.

Helmholtz, H. (1924). Treatise on physiological optics (Vol. 3). *Menasha, WI: Optical Society of America*.

Hertz, U., Bahrami, B., & Keramati, M. (2018). Stochastic satisficing account of confidence in uncertain value-based decisions. *PLoS One*, *13*(4), e0195399.

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E. M., & Stephan, K. E. (2013). Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*, *80*(2), 519–530. https://doi.org/10.1016/j.neuron.2013.09.009

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krugwasser, A. R., Stern, Y., Faivre, N., Harel, E. V., & Salomon, R. (2021). *Impaired Sense of Agency and Associated Confidence in Psychosis*.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463.

Loewy, R. L., Pearson, R., Vinogradov, S., Bearden, C. E., & Cannon, T. D. (2011). Psychosis risk screening with the Prodromal Questionnaire—Brief Version (PQ-B). *Schizophrenia Research*, *129*(1), 42–46. https://doi.org/10.1016/j.schres.2011.03.029

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Mei, N., Rahnev, D., & Soto, D. (2023). Using serial dependence to predict confidence across observers and cognitive domains. *Psychonomic Bulletin & Review*, *30*(4), 1596–1608.

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS Computational Biology*, *11*(6), e1004305.

Nelson. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109.

Nelson, B., Whitford, T. J., Lavoie, S., & Sass, L. A. (2014). What are the neurocognitive correlates of basic self-disturbance in schizophrenia?: Integrating phenomenology and neurocognition. Part 1 (Source monitoring deficits). *Schizophrenia Research*, *152*(1), 12–19.

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553.

Nobre, A. C. (Kia), & Kastner, S. (Eds.). (2014). *The Oxford Handbook of Attention*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199675111.001.0001

Rahnev, D. (2021). Visual metacognition: Measures, models, and neural correlates. *American Psychologist*, *76*(9), 1445.

Rahnev, D., Desender, K., Lee, A. L., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., & Bang, J. W. (2020). The confidence database. *Nature Human Behaviour*, *4*(3), 317–325.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664–1680.

Rausch, M., & Zehetleitner, M. (2019). The folded X-pattern is not necessarily a statistical signature of decision confidence. *PLoS Computational Biology*, *15*(10), e1007456.

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning, Current Research and Theory*, *2*, 64–69.

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: Insights from individual differences and neuroimaging. *Personality Neuroscience*, *1*, e17.

Salomon, R., Kannape, O. A., Debarba, H. G., Kaliuzhna, M., Schneider, M., Faivre, N., Eliez, S., & Blanke, O. (2021). Agency Deficits in a Human Genetic Model of Schizophrenia: Insights From 22q11DS Patients. *Schizophrenia Bulletin*, sbab143. https://doi.org/10.1093/schbul/sbab143

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron*, *90*(3), 499–506.

Schmack, K., Bosc, M., Ott, T., Sturgill, J. F., & Kepecs, A. (2021). Striatal dopamine mediates hallucination-like perception in mice. *Science*, *372*(6537), eabf4740. https://doi.org/10.1126/science.abf4740

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

Seideman, J. A., Stanford, T. R., & Salinas, E. (2018). Saccade metrics reflect decision-making dynamics during urgent choices. *Nature Communications*, *9*(1), 2907.

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

Spering, M. (2022). Eye movements as a window into decision-making. *Annual Review of Vision Science*, *8*(1), 427–448.

Stern, Y., Koren, D., Moebus, R., Panishev, G., & Salomon, R. (2020). Assessing the Relationship between Sense of Agency, the Bodily-Self and Stress: Four Virtual-Reality Experiments in Healthy Individuals. *Journal of Clinical Medicine*, *9*(9), 2931.

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635.

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637.

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*, *26*(23), 3157–3168.

van Ede, F., Board, A. G., & Nobre, A. C. (2020). Goal-directed and stimulus-driven selection of internal representations. *Proceedings of the National Academy of Sciences*, *117*(39), 24590–24598.

Van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience*, *22*(3), 477–483.

Viganò, S., Bayramova, R., Doeller, C. F., & Bottini, R. (2024). Spontaneous eye movements reflect the representational geometries of conceptual spaces. *Proceedings of the National Academy of Sciences*, *121*(17), e2403858121.

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial attention, precision, and Bayesian inference: A study of saccadic response speed. *Cerebral Cortex*, *24*(6), 1436–1450.

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*, *3*(4), 150670.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for? *Annual Review of Vision Science*, *6*, 539–562.