AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2024 The Author(s)
ISSN: 0096-3445

2025, Vol. 154, No. 3, 852–863
https://doi.org/10.1037/xge0001690

Adaptive Curiosity About Metacognitive Ability

Samuel Recht[1], Canqi Li[1], Yifan Yang[1], and Kaiki Chiu[2]
[1] Department of Experimental Psychology, University of Oxford
[2] Department of Psychology, Yale University

Metacognition provides control and oversight to the process of acquiring and using knowledge. Efficient metacognition is essential to many aspects of daily life, from health care to finance and education. Across three experiments, we found a specific form of curiosity in humans about the quality of their own metacognition, using a novel approach that dissociates perceptual from metacognitive information searches. Observers displayed a strategic balance in their curiosity, alternating between a focus on perceptual accuracy and metacognitive performance. Depending on the context, this metacognitive curiosity was modulated by an internal evaluation of metacognition, leading to increased feedback requests when metacognition was likely to be inaccurate. Using an ideal observer model, we describe how this curiosity trade-off can arise naturally from a recursive evaluation and transformation of decisions' evidence. These results show that individuals are inherently curious about their metacognitive abilities and can compare perceptual and metacognitive precision to fine-tune performance monitoring. We propose that this form of curiosity may reflect humans' drive to refine their self-model.

Public Significance Statement
Metacognition functions as the brain's built-in "supervisor," monitoring how we acquire and apply knowledge. This study found that humans display a selective curiosity toward metacognition, often preferring to learn about the quality of their metacognitive monitoring rather than reward-related performance. Importantly, this curiosity varied widely between individuals, highlighting diverse preferences for metacognitive ability. Moreover, individuals actively sought additional feedback when they perceive their metacognitive accuracy to be low, suggesting an adaptive monitoring over metacognition. By uncovering this form of introspective curiosity, these findings open up new avenues for our understanding of monitoring and control during perceptual decision making.

Keywords: metacognition, curiosity, perception, self-regulation

Supplemental materials: https://doi.org/10.1037/xge0001690.supp

Metacognition—humans' ability to reflect on their own cognition—has been shown to play an important role in regulating behavior, providing oversight and control to the process of acquiring knowledge and understanding. From infancy to adulthood and across neurotypes, metacognition is often viewed as a critical factor in human's adaptability, and its implication in fostering learning has been extended to the study of artificial systems (Cox, 2005; Lake & Baroni, 2023).

Metacognition is usually defined through two main components: monitoring of first-order thoughts, perceptions, feelings, and memories,

This article was published Online First December 12, 2024.

Rachel Wu served as action editor.

Samuel Recht https://orcid.org/0000-0003-0317-2994

Canqi Li and Yifan Yang contributed equally. This work has not been published elsewhere. A preprint has been made available on PsyArXiv at https://doi.org/10.31234/osf.io/mhnvt. All behavioral data and findings (i.e., except the observer model) in this article were presented in June 2023 at the Perceptual Metacognition Satellite—26th Annual Meeting of the Association for the Scientific Study of Consciousness.

This work was supported by a Fyssen Foundation fellowship and funds from Pembroke College, University of Oxford, to Samuel Recht and a Yinghua Scholarship to Yifan Yang. The authors thank Monica Barbir, Matan Mazor, and Jacques Pesnot Lerousseau for their helpful comments on this project.

Samuel Recht played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, supervision, writing–original draft, and writing–review and editing. Canqi Li played an equal role in conceptualization, data curation, investigation, writing–original draft, and writing–review and editing. Yifan Yang played a supporting role in formal analysis and an equal role in conceptualization, data curation, investigation, writing–original draft, and writing–review and editing. Kaiki Chiu played a supporting role in formal analysis, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Samuel Recht, Department of Experimental Psychology, University of Oxford, Anna Watts Building, Woodstock Road, Oxford OX2 6GG, United Kingdom. Email: samuel.recht@gmail.com

and the active control over these mental operations to improve their efficiency in dealing with the environment (Fleming et al., 2012; Yeung & Summerfield, 2012). A key dimension of metacognition could be found in confidence judgment, which determines our certainty about our decisions and sensory experiences (Mamassian, 2016). Confidence often shapes how we seek information, prioritize tasks, set goals, and communicate (e.g., Aguilar-Lleyda & de Gardelle, 2021; Bahrami et al., 2012; Desender et al., 2018; Pescetelli et al., 2021). Confidence has been shown to involve both the prefrontal cortex as well as a multitude of brain regions active more or less specifically depending on the task (Fleming, 2024). Confidence judgments are also hierarchical across time frames, from local confidence estimations to more global evaluations (e.g., Cavalan et al., 2023; Lee et al., 2021). How well confidence judgments reflect reality is often referred to as "metacognitive ability" and has been shown to vary across individuals and situations (e.g., Smith et al., 2003). Despite being often reliable, metacognitive accuracy is prone to fluctuation, particularly early in development and in certain clinical populations.

Many aspects of daily life can be affected by metacognitive failures, from health care to finance, policymaking, and education. For instance, in bipolar disorder, compromised metacognitive monitoring can delay the recognition of mood episodes, potentially worsening the condition (Van Camp et al., 2019). In financial decision making, overconfidence can lead to increased risk-taking and sensation-seeking behaviors (Grinblatt & Keloharju, 2009; Stotz & von Nitzsch, 2005) and has been tied to the destruction of company value (Ahmed & Duellman, 2013). Overconfidence is also a feature in problem gambling (Friedemann et al., 2024). Finally, ineffective metacognition among students has been described as one of the main obstacles to their academic growth (Bransford et al., 2000). Hence, the negative impacts of metacognitive failures on society are significant enough that efforts should be made to better understand and address them.

However, the precise scope of metacognition's role in guiding subsequent information-seeking remains an open question. Recent research, for instance, suggests that certain forms of uncertainty can drive information search without requiring metacognitive mediation (Edwards-Lowe et al., 2024). Similarly, metacognitive failures associated with certain psychiatric symptoms do not necessarily always lead to failures in information-seeking (Mohr et al., 2024). Therefore, it is important to identify which kinds of information search depend on metacognitive evaluations and which do not.

Improving metacognitive evaluation should also be considered in light of a variety of goals, contexts, and personal experiences, likely requiring nuanced and individualized feedback strategies (e.g., debiasing in finance, Kaustia & Perttula, 2012). For example, guiding patients specifically to identify and address their metacognitive shortcomings has shown promise in psychotherapy (e.g., Fisher & Wells, 2008; Lysaker et al., 2018). Metacognitive interventions have also been found to improve resilience against misinformation (Salovich & Rapp, 2021) and to significantly foster education (Hattie, 2008). Yet, whether humans can effectively probe and improve their own metacognition without explicit guidance or training has not been conclusively answered.

Intuitively, gauging the quality of our own metacognitive judgment seems frequent in daily life. An athlete, for example, might not only track their performance in practice sessions but also assess their confidence in their skills. If they perform well but lack confidence, they might seek additional coaching to better understand the source of the discrepancy between reality and their self-model. A trader, on the other hand, may be aware of the possibility of becoming overconfident and taking greater risk after a series of successful investments and take precautionary measures. In such instances of under- or overconfidence, the decision to seek more information about the quality of their metacognition has a cost. However, refining the self-model an individual possesses could significantly improve their decision making in the future.

Recent work proposes that confidence judgments in perception could indeed involve further levels of self-evaluation, confirming the possibility that individuals may recognize metacognitive failures and remediate them in a self-supervised manner. For example, we recently found evidence of recursive metacognitive evaluations in human vision (e.g., "I know that I know that I saw"; Recht et al., 2022) without the need for training or explicit feedback. Subsequent studies using different research paradigms have further corroborated the existence of recursive metacognitive judgments (Cavalan et al., 2024; Sherman & Seth, 2024; Zheng et al., 2023). These findings for perceptual metacognition are consistent with previous findings from the literature on memory and learning, where higher order inferences have also been identified (Buratti & Allwood, 2015; Dunlosky et al., 2005; Händel & Fritzsche, 2016). Although humans can assess their metacognition, this does not guarantee they will do so willingly without explicit instructions or clear benefits.

A crucial step, therefore, involves examining how people value metacognitive information: Are individuals intrinsically interested in understanding their own thought processes? If so, how do they weigh the significance of insights about their metacognition against more direct perceptual performance information? One challenge in addressing these questions lies in the closely intertwined relationship between perception and metacognition. It is generally understood that metacognitive evidence is at least partially built upon perceptual decision evidence, thereby linking searches for information on one aspect inevitably to the other. A conservative definition of metacognition might require metacognitive evidence not to be reducible to perceptual evidence. Alternatively, it could be defined by its temporal relationship to decision making: Evidence is metacognitive when it is extracted after a first-order (e.g., perceptual) decision and relates to a previous decision. In this work, we use a confidence 2-alternative forced choice (2AFC) as a metacognitive judgment, where the observer selects the more accurate of two preceding perceptual decisions (Mamassian, 2020). This requires comparing the perceptual evidence of each decision to make a confidence judgment. Even without additional noise or signal posterior to the perceptual decisions, this confidence evaluation transforms two independent decision signals into a new one. It is this transformation and its product that we define as metacognitive.

Here, we propose that fluctuations in decision evidence between perception and metacognition drive an adaptive shift in curiosity from perceptual to metacognitive monitoring. We introduce a novel approach that separates perceptual from metacognitive information searches, enabling us to explore how decision evidence influences the preference for seeking information about metacognitive abilities. To test our hypothesis, we used a paradigm where participants made two perceptual judgments (or first-order decisions), followed by a (confidence) metacognitive judgment to select the better of the two initial decisions. A last stage involved giving the option to selectively seek feedback on the accuracy of either perception or metacognition. Crucially, many participants preferred feedback about their metacognitive decision, despite only perception feedback offering direct information about their gain in a given trial. This curiosity for metacognition alternated—over time and between individuals—with a curiosity for perceptual accuracy. Using a computational model, we illustrate how this curious exploration can naturally emerge from the recursive transformation and evaluation of decision evidence, considering that perception fluctuates over time and observers cultivate an appetite for self-knowledge.

Material and Method

Participants

In Experiment 1, 20 naïve adults were recruited from the Prolific platform, averaging 31 ± 7 years (*M* ± *SD*; seven reported their gender as female, 13 as male) and compensated at £8.5/hr. Experiment 2 included 10 students from the University of Oxford, recruited via the University's online platform and compensated with course credits, along with 16 participants recruited via word of mouth, totaling 26 participants (25 ± 11 years, 15 reported their gender as female, 11 as male). In Experiment 3, 62 adults from the Prolific platform (31 ± 8 years, 29 reported their gender as female, 33 as male) were compensated for their time at a rate of £8.5/hr. Participants' age and gender were registered on Prolific and the University's platform and collected with their informed consent. Participants recruited via word of mouth provided free-response information on their age and gender. All participants across the experiments reported normal or corrected vision, no known neurological issues, no psychoactive medication use, and fluency in English. For Experiments 1 and 3, they had an approval rate equal to or above 97% on Prolific. Experiments 1 and 3 took approximately 20 min to complete, with the 50% best performing participants receiving a bonus of £0.6. Experiment 2 took 50 min to complete, and no bonus was paid. All procedures were approved by the University of Oxford's Research Ethics Committee. Experiment 1 was exploratory without formal power analysis. Sample sizes for Experiments 2 and 3 were based on Experiment 1's effect size, using a Bayesian stopping rule (Supplemental Material provides further details). In Experiment 2, one out of the initial 26 participants was excluded due to failed attention checks. Similarly, in Experiment 3, one out of the 62 participants was excluded for the same reason, and another one for not submitting any data (because of a technical issue). In total, data from 105 participants were used for further analysis.

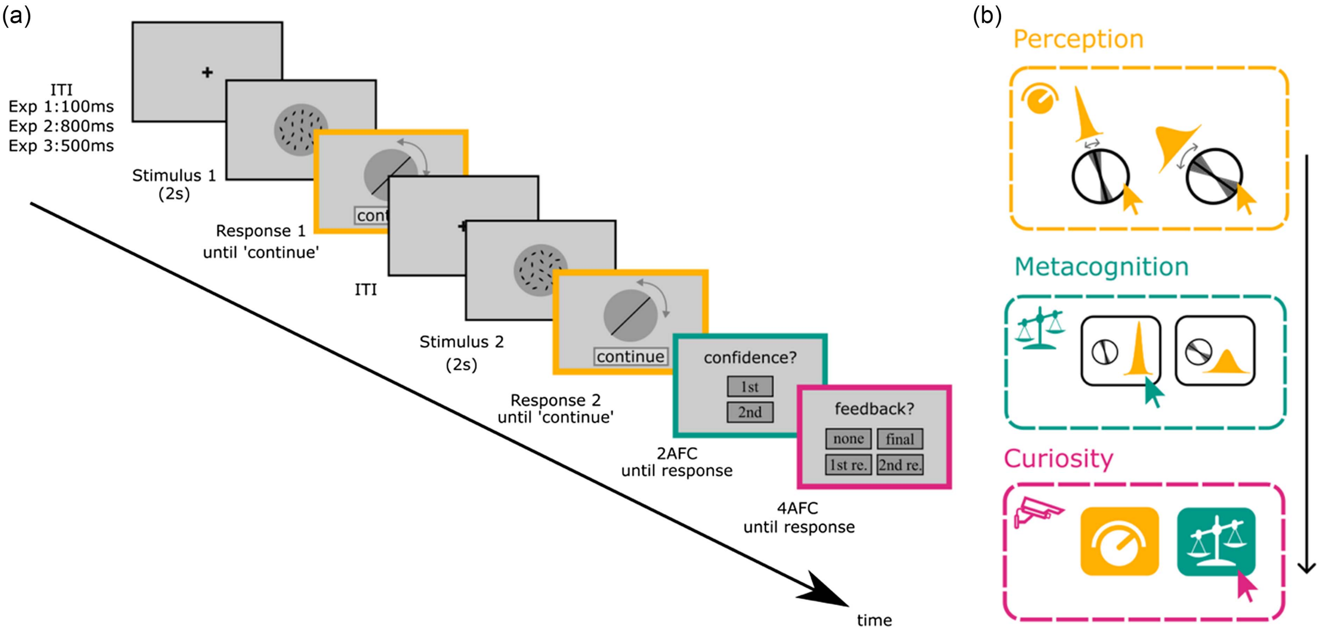
Experimental Protocol

Upon obtaining informed consent, participants were provided with detailed instructions on how to perform the task, where they were encouraged to give unspeeded responses for all trials. To adapt the experiment to their screen size, participants first had to rescale a rectangle using the mouse cursor to match the dimensions of a credit card-sized card held up to the screen. This approach allowed stimuli to be presented in standardized units. Each trial started with a central fixation cross (0.6° of visual angle, assuming a distance of 60 cm from the screen) presented on a light gray background for 100 ms in Experiment 1, 800 ms in Experiment 2, and 500 ms in Experiment 3 (intertrial interval or "ITI"). Following the ITI, a circle of 7.53° and color RGB (187, 187, 187) was displayed at fixation, with 20 equally spaced 0.6° visual gratings composing a grid (spatial frequency = 1.67 cycle/degree, Gaussian window), their individual orientation being sampled from a circular normal distribution (Von Mises). The average orientation of the Von Mises was randomly sampled from a uniform distribution over [0°, 180°], the variance of the Von Mises being specific to the conditions (see the end of the current section for more details). On each trial, participants were asked to first observe the gratings within the circle for 2 s and estimate their average orientation (Figure 1). Then, they were asked to reproduce that average orientation by rotating a line probe with their mouse cursor. Each response was converted into points depending on accuracy (from 0 points, corresponding to a 90° error, up to 100 points, corresponding to a 0° error). Every two trials, participants were asked to choose one of the two previous responses to keep for reward (confidence 2-alternative forced choice, or 2AFC), knowing that only the points from the selected response will be added to their total score. Importantly, following the confidence 2AFC, participants were presented with a 4-alternative forced choice—with the options listed in random order—to decide what type of feedback they would like to receive: no feedback, feedback on the first response, feedback on the second response, or feedback on the final choice (i.e., metacognitive judgment). If they decided to choose the first or second response, they were provided with the accuracy of their response in points. If they opted for the final choice, they were informed whether the response they kept was the best one in the pair (but without its accuracy in points). Finally, the "no feedback" option led to no information. This option was offered to filter out participants not showing any interest for feedback. Following their click, the button turned to the requested information (or remained empty for the "no feedback" option) for 1.2 s, followed by an ITI that varied between experiments.

Every 20 trials, participants were prompted to provide three distinct subjective global estimates of their performance within the block using a rating scale. One more question involved an attentional check to make sure participants were paying attention to the task. The order of these questions was randomized. Except for the attentional check, we did not analyze these ratings in the present study. Following the rating scales, participants were provided with a 15-s break to rest. During this break, they were presented with the number of points earned in the previous 20 trials, along with the total points earned throughout the experiment. All three experiments had the same characteristics, at the exception of the ITI, the number of trials/participants, and the payoff structure. Specifically, Experiment 1 involved 80 trials (40 pairs), Experiment 2 involved 200 trials, and Experiment 3 involved 80 trials. Experiments 1 and 3 involved payment and bonuses, but not Experiment 2. To familiarize participants with the task and the format of the trials, they were presented with four demo trials that only included the perceptual discrimination and visual feedback on the accuracy of their responses. The experiment was displayed in full-screen mode. Exiting full-screen mode during the course of the experiment was automatically reversed in the following trial.

Crucially, we manipulated the relative difficulty level of trials within a pair by altering the variance of the stimulus grid (and therefore the difficulty of reproduction), in four distinct conditions (the Supplemental Material provides further details): easy-easy ("E-E") where the variance for the stimuli's sampling distribution (circular normal) in the two trials was equated and low, easy-hard ("E-H") where the variance was larger for the second trial in the pair, hard-easy ("H-E"), and hard-hard ("H-H"). It is important to note that "condition" in this article refers to

Figure 1
Paradigm



Note. (a) In a perceptual decision task, participants were presented with a grid of oriented gratings for 2s and had to reproduce the average orientation at the end of the trial (i.e., perceptual decision). After two trials, they had to make a confidence judgment (i.e., metacognitive decision) by selecting which of the two previous trials was the best. Finally, they were given the option (i.e., curiosity decision) to seek feedback either on their perceptual or metacognitive accuracy. By design, perceptual accuracy could not be inferred from metacognitive feedback and vice versa. They also had the option to not request any feedback at all. The ITI differed between experiments, as well as the number of participants/trials and the payoff structure. (b) From top to bottom panels: Each panel represents a decision order. In the top "Perception" panel, the observer makes perceptual decisions, with varying degrees of precision (represented by the width of the distribution around the correct angle). The middle "Metacognition" panel represents the selection of the best trial in the pair of perceptual responses, with the aim of selecting the one with greater precision. The bottom "Curiosity" panel represents the choice in feedback requests, where participants have to decide between perceptual or metacognitive feedback (see Figure 2a for the distribution of responses). Exp = experiment; ITI = intertrial interval; AFC = alternative forced choice. See the online article for the color version of this figure.

the difficulty of trial pairs (four levels, "E-E," "E-H," "H-E," "H-H"), instead of individual trials (two levels, i.e., "E" or "H").

Transparency and Openness

The data for all experiments and the scripts to run the model are available on the Open Science Framework repository at https://osf.io/45wvb/. The general approach for statistical (and power) analyses, results for each experiment, and a comprehensive description of the model can be found in the Supplemental Material. The experimental design, main hypotheses, and analyzing pipeline for Experiments 2 (https://aspredicted.org/ii3gb.pdf) and 3 (https://aspredicted.org/y932m.pdf) were preregistered (details provided in the Supplemental Material, including notes on minor or major deviations from preregistration). Experiment 1 was exploratory and not preregistered.

Results

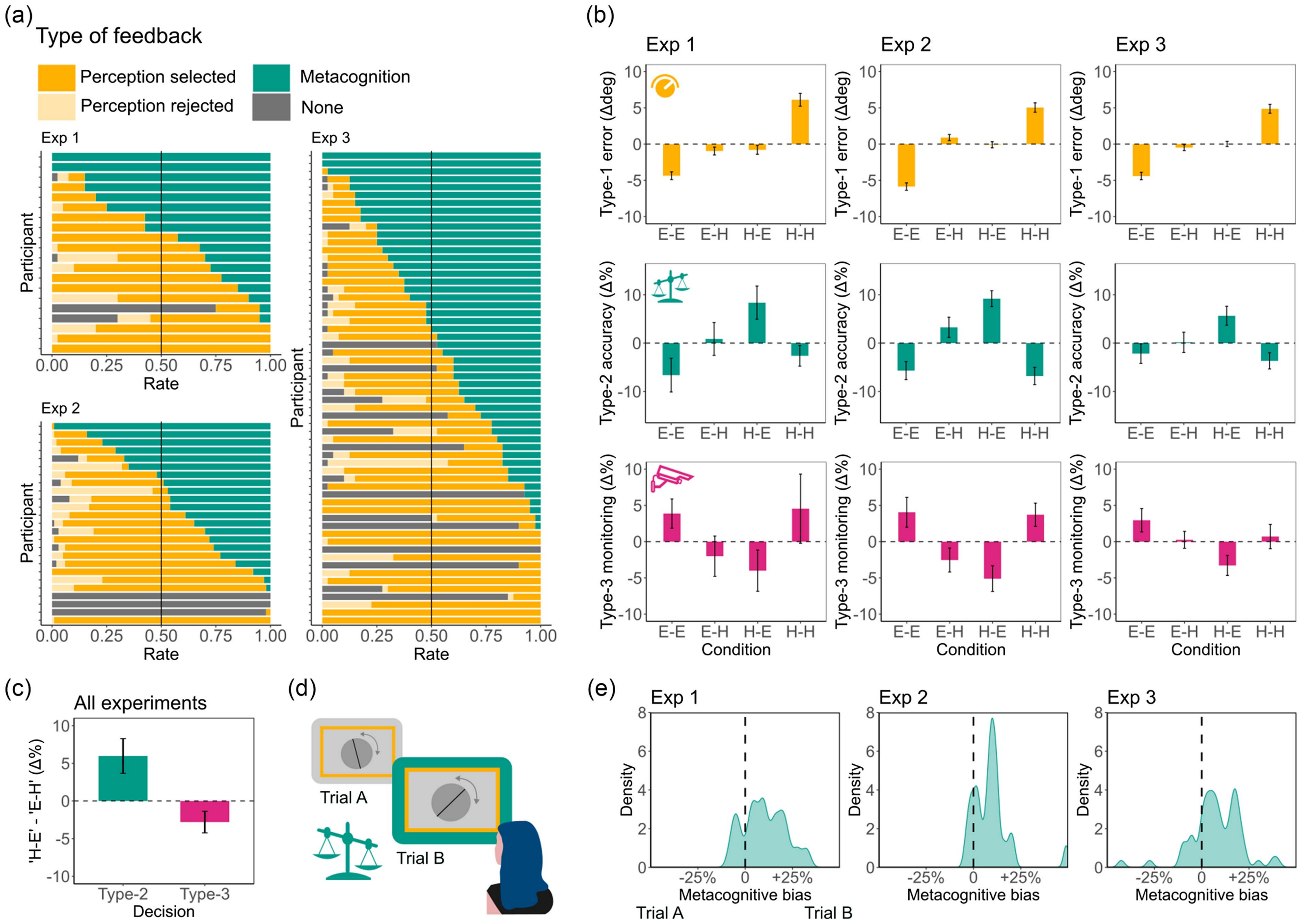
In the following sections, we report all statistical tests we have used for our hypotheses and explicitly state when a test was preregistered. We used linear mixed-effects models (see Supplemental Material for details) and report ΔAIC as the difference between the model without the tested effect and the model with it (a negative value is evidence against the tested effect). Experiment 2 was conducted to replicate Experiment 1, and Experiment 3 was conducted after finding an effect of condition, but no effect of metacognitive accuracy, in Experiment 2 (the preregistrations provide more details).

Perception (Type 1 Decision)

Participants' perceptual accuracy was quantified using the absolute angle difference between the reported orientation and the true orientation (i.e., the average orientation of the distribution). This error was also converted into points (max. of 100 points corresponding to a 0° error). On average, participants did well (in points, $M \pm SD$; Experiment 1: 81.87 ± 8.14, Experiment 2: 81.94 ± 8.78, Experiment 3: 77.55 ± 11.81). We expected lower error on average (across the two trials within a pair) for the "easy" trial pairs ("E-E"), medium average error in the mixed pairs ("E-H" or "H-E"), and greater average error for the "hard" pairs ("H-H"). Figure 2b, top panel, illustrates the relative difference in error between conditions (in degree). We used a linear mixed-effects regression to test the association between condition (four levels, pair-wise) and trial order (i.e., the first vs. second trial in a pair) as predictor(s), and perceptual error per trial as outcome. Across all three experiments, we found that condition was a significant predictor of error (all χ^2s > 47.50, ps < .001,

Figure 2

Behavior



Note. (a) The chart displays the distribution of feedback requests per participant for each experiment. In the chart, the yellow bars represent the proportion of feedback requests related to the quality of the perceptual decision (Type 1) for both the selected and rejected (low α) trials. Notably, the majority of Type 1 feedback requests focused on the selected trials. The green bars represent the proportion of feedback requests related to the quality of the metacognitive decision (Type 2). The gray bars indicate the proportion of instances where participants did not request any specific type of feedback. (b) The figure illustrates the average accuracy for perception (yellow), metacognition (green), and the proportion of times participants requested Type 2 (metacognitive) feedback instead of Type 1 (perceptual) feedback across different conditions (pink, easy-easy, easy-hard, hard-easy, and hard-hard). On the *y*-axis, the values represent the average difference in absolute error in degrees (Δdeg) or the average difference in percent (Δ%), normalized per participant. (c) The difference between hard-easy and easy-hard conditions for metacognitive accuracy (Type 2) and curiosity (Type 3) request rates, pooled across experiments, showing higher metacognitive accuracy for the hard-easy compared to the easy-hard, despite both having the same objective metacognitive difficulty. (d, e) The difference observed in (c) can be understood when considering the overall metacognitive "recency" bias toward the second trial in the pair. Figure (e) plots the distribution of average bias across participants; the majority of participants showed a bias toward the second trial (i.e., were more likely to select the second trial). Error bars represent the standard error of the mean (SEM). E-E = easy-easy; E-H = easy-hard; H-E = hard-easy; H-H = hard-hard; Exp = experiment. See the online article for the color version of this figure.

ΔAICs > 48.50, Supplemental Table S1). The main effect of trial order was not significant (all χ^2s < 1.51, *ps* > .21, ΔAICs < 1.70, Supplemental Table S2), suggesting that the second responses were not always better or worse than the first across conditions, and vice versa. There was, however, an interaction between trial order and condition (all χ^2s > 90.82, *ps* < .001, ΔAICs > 91.43, Supplemental Table S3), which is expected given the existence of mixed pairs ("E-H" vs. "H-E"). Such results confirmed successful manipulation of difficulty level across conditions, which effectively influenced participants' perceptual performance.

Metacognition (Type 2 Decision)

In a second analysis, we quantified participants' metacognitive accuracy. During the metacognitive (or Type 2) judgments, participants had to select the best of their two previous responses to keep for later reward. Here, we define metacognitive accuracy as the probability of correctly selecting the trial with lower error. This definition of metacognitive accuracy is less conservative than the literature's typical definition of the related concept of metacognitive sensitivity (see Mamassian, 2016). In a set of *t* tests, we compared the average metacognitive accuracy to chance level (50%). As expected, participants showed above-chance metacognitive accuracy, being able to pick out the more precise response in all experiments (all rates ≥ 58%, *p*s < .001, BF_{10}s > 388.70, Supplemental Table S4). As it was for perceptual accuracy, we also expected condition to be a significant factor for metacognitive performance (Figure 2b, middle panel). Among the four conditions, we expected that it is more difficult to distinguish performance levels between similar pairs ("E-E" and "H-H"), while it is easier to do so in mixed pairs ("E-H" and "H-E"). Consequently, mixed trial pairs should lead to greater metacognitive accuracy compared to similar ones. For all three experiments, we found a main effect of condition in predicting metacognitive accuracy (all χ^2s > 8.90, *p*s < .035, ΔAICs > 2.90, Supplemental Table S5). Such results suggested that participants were above chance in their metacognitive judgments and that our condition manipulation also affected their metacognitive performance.

An unexpected finding was the tendency for metacognitive accuracy to be greater in the "H-E" compared to the "E-H" condition, despite the two conditions involving similar difficulty levels for metacognition (Figure 2c). An exploratory analysis (over data pooled across experiments) confirmed a significant accuracy difference, $t(104) = 2.60$, $p = .01$, $\text{BF}_{10} = 2.61$, two-sided test. This effect was probably driven by the metacognitive bias toward the second trial in the pair observed in all experiments (Figure 2d and 2e; all *p*s < .001, and BF_{10}s > 52.60, Supplemental Table S6): When the condition conformed to the bias (i.e., in "H-E", the second trial is easier, in line with the bias), metacognitive performance increased.

Curiosity (Type 3 Decision)

Our main aim was to investigate how participants arbitrate feedback-seeking about the quality of their perceptual and metacognitive judgments. A first prediction was that despite only perceptual performance feedback directly informing on the actual points earned, participants would still show an interest in metacognitive feedback, even at the cost of not getting perceptual feedback at all. Considering the proportion of metacognitive compared to other feedback requests, we found above chance-level (> 25%) curiosity about metacognitive ability in all experiments (all rates ≥ 36%, *p*s ≤ .03, BF_{10}s > 2.1, Supplemental Table S7), with strong evidence in Experiment 3 ($\text{BF}_{10} = 26.15$). Figure 2a illustrates the notable interindividual variability for such an appetite: Some participants were mostly interested in their perceptual accuracy (in yellow in the figure), while other participants were mostly interested in their metacognitive accuracy (in green). Many participants also showed alternations over trials. Comparing the curiosity for perceptual accuracy (including selected and rejected trials) to that for metacognitive accuracy using *t* tests (filtering out the "no feedback" requests), we found that the request rate for metacognitive feedback was not significantly below 50% overall (*p*s > .13, BF_{10}s < 0.65, Supplemental Table S8). The similar average curiosity for the two types of feedback suggested that at the population level, participants showed a strong appetite for metacognitive feedback.

Strategic Shifts in Curiosity Between Perception and Metacognition

Similar to metacognitive accuracy, we further hypothesized that the curiosity for metacognitive feedback is also a function of the difficulty posed by the metacognitive task (i.e., how difficult it is to select the better trial in a pair). Participants would request feedback on their metacognitive choice more often when it is challenging (between similar trials), but less so when it is easy (between mixed trials). For all subsequent analyses, we excluded the "no feedback" requests and calculated the rate of metacognitive feedback requests relative to perceptual performance feedback requests (for either trial in each pair). Using a linear mixed-effects logistic regression (preregistered for Experiments 2–3), we found the condition to significantly predict the rate of metacognitive feedback requests in two out of the three experiments. In both Experiments 2 and 3, the models with condition as a predictor were significantly better than the null models, Experiment 2: $\chi^2(3) = 21.05$, $p < .001$, ΔAIC = 15.05; Experiment 3: $\chi^2 = 8.34$, $p = .04$, ΔAIC = 2.34, but not in Experiment 1, $\chi^2(3) = 3.74$, $p = .29$, ΔAIC = −2.26. The fact that Experiment 2 (involving a larger number of trials than Experiment 1) and Experiment 3 (involving a larger number of participants than Experiment 1) showed the main effect of condition suggested that the lack of effect in Experiment 1 was probably due to a lower statistical power. Figure 2b, lower panel, illustrates the relative Type 3 curiosity as a function of condition, once normalized per participant (to account for individual's average curiosity, for readability).

Finally, while the condition effect suggested an effect of metacognitive difficulty on curiosity, our last prediction pertained to the effect of metacognitive accuracy itself. If participants are able to evaluate the quality of their metacognition (a signature of meta-metacognitive ability), they should seek corresponding feedback on metacognition particularly when their metacognitive choice is wrong. In a linear mixed-effects logistic regression (preregistered for Experiments 2–3), we found metacognitive accuracy to predict curiosity in two out of three experiments. In both Experiments 1 and 3, the models including metacognitive accuracy as a predictor were significantly better than the null models, Experiment 1: $\chi^2(1) = 6.89$, $p = .01$, ΔAIC = 4.89; Experiment 3: $\chi^2(1) = 5.42$, $p = .02$, ΔAIC = 3.42. The model for Experiment 2 did not offer significant improvement over the null model, $\chi^2(1) = 0.15$, $p = .69$, ΔAIC = −1.85. We found no significant interaction between metacognitive accuracy and condition in Experiments 1 and 2 (all χ^2s < 8.40, *p*s > .20, ΔAICs < −3.65). In contrast, in Experiment 3, the model with interaction improved the fit significantly, $\chi^2(3) = 11.25$, $p = .01$, ΔAIC = 5.25, Supplemental Table S9. Consistent with the results, one-sample, one-tailed *t* tests (preregistered for Experiments 2–3) also showed a significant increase in metacognitive feedback requests for wrong compared to correct metacognitive choices in Experiments 1 and 3 (only considering participants having data points in both and a nonzero metacognitive feedback requests rate; all *p*s = .01, BF_{10}s > 5.95, Supplemental Table S10), but not in Experiment 2, where we found moderate evidence for the null ($p = .39$, $\text{BF}_{10} = 0.29$). Experiment 2 was the only experiment without monetary incentives or bonuses, and this difference could suggest an effect of reward on metacognitive effort. Following our exploratory analysis on metacognitive bias (Figure 2c–e), we also tested a potential difference in curiosity between "H-E" and "E-H" pairs over data pooled across experiments. Despite a qualitative trend in all experiments (Figure 2b–c), the difference remained at significance threshold, $t(98) = -1.95$, $p = .05$, $\text{BF}_{10} = 0.69$, two-tailed *t* test. Considering our (preregistrered) prediction of an inverse relationship between metacognitive accuracy and curiosity for metacognitive feedback, we also restricted the test to one tail, which only modestly improved the outcome, $t(98) = -1.95$, $p = .03$, $\mathrm{BF}_{10} = 1.33$.

Bayesian Observer Model

An arbitrage in curiosity between perceptual and metacognitive accuracy requires some form of higher order evaluation of metacognitive reliability (i.e., meta-metacognition). However, this does not necessarily mean a supplemental monitoring system; a more parsimonious explanation could be the existence of a unitary system for all higher order decisions, with specific, goal-defined transformation of available evidence as a function of decision order and task requirements (Recht et al., 2022; Zheng et al., 2023). To illustrate how a curiosity trade-off could emerge from a recursive evaluation of decision evidence, we developed a descriptive Bayesian observer model (Figure 3a). In the model, for each pair of trials, the observer first estimates the orientation of the stimulus (perceptual decisions, Figure 3a left panel) and then compares the precision (or perceptual evidence) in each decision, selecting the one with greater precision (metacognitive response, Figure 3a middle panel). In our model, metacognitive evidence perfectly reflected the difference in perceptual evidence. However, postdecisional noise likely corrupts this metacognitive readout, leading to suboptimal metacognition in real-world scenarios. For curiosity, the observer then compares perceptual evidence to metacognitive evidence to request feedback (Figure 3a right panel). When the perceptual evidence of the selected trial is greater than the metacognitive evidence (i.e., than the difference in perceptual evidence between the two trials), the observer requests feedback about metacognition. Conversely, when the perceptual evidence is lower than the metacognitive evidence, the situation is reversed. This strategy would allow the observer to fine-tune performance monitoring depending on the context and curiosity traits, considering that both perceptual and metacognitive evidence are subjectively valuable. The valuation being always in perceptual evidence units, this approach provides a form of common currency across perception, metacognition, and curiosity judgments. For simplicity, we did not consider cases where an observer seeks feedback on rejected (low-confidence) trials, as these are infrequent in our data set. This type of counterfactual curiosity has been discussed elsewhere (Fitzgibbon & Murayama, 2022).

Drawing from earlier research (e.g., van den Berg et al., 2012), we used Fisher's information (J) to measure perceptual precision and quantify the evidence used for metacognitive and curiosity judgments. We hypothesized that perception fluctuates from trial to trial, leading to variable precision in reports. To model these fluctuations, we opted for a γ distribution (two parameters, see Supplemental Material), a model used in both the working memory (Schneegans et al., 2020; van den Berg et al., 2012) and metacognition literature (Geurts et al., 2022; Recht et al., 2021). In line with the literature, we also hypothesized that the observer has access to the perceptual evidence and can use it during metacognitive evaluations. As shown in Figure 3a, the nature of perceptual evidence has a significant impact on predicted metacognitive accuracy and curiosity across precision levels. Our sample size made the fit per participant impractical; we therefore used an aggregated observer model approach.

The average precision range was selected considering the observed precision values in our empirical data. Figure 3b–c shows the observed error, metacognitive accuracy and curiosity rates, and the predicted rate from fitting the model to perceptual responses in the hard and easy trials separately. The observer is deemed ideal for metacognition since no additional noise is added to the perceptual evidence signal. The curiosity decision is more subjective: It involves comparing metacognitive and perceptual evidence (ΔE). The metacognitive evidence, which is the difference in evidence between two perceptual decisions (ΔJ), is invariably lower than the perceptual evidence of the best trial in the pair. This is attributed to the lower limit on precision (J) being 0. Consequently, the absolute difference between the precision of the first trial (J_A) and that of the second trial (J_B) can never exceed the maximum value in the precision pair (J_A and J_B, see the Supplemental Material for more details).

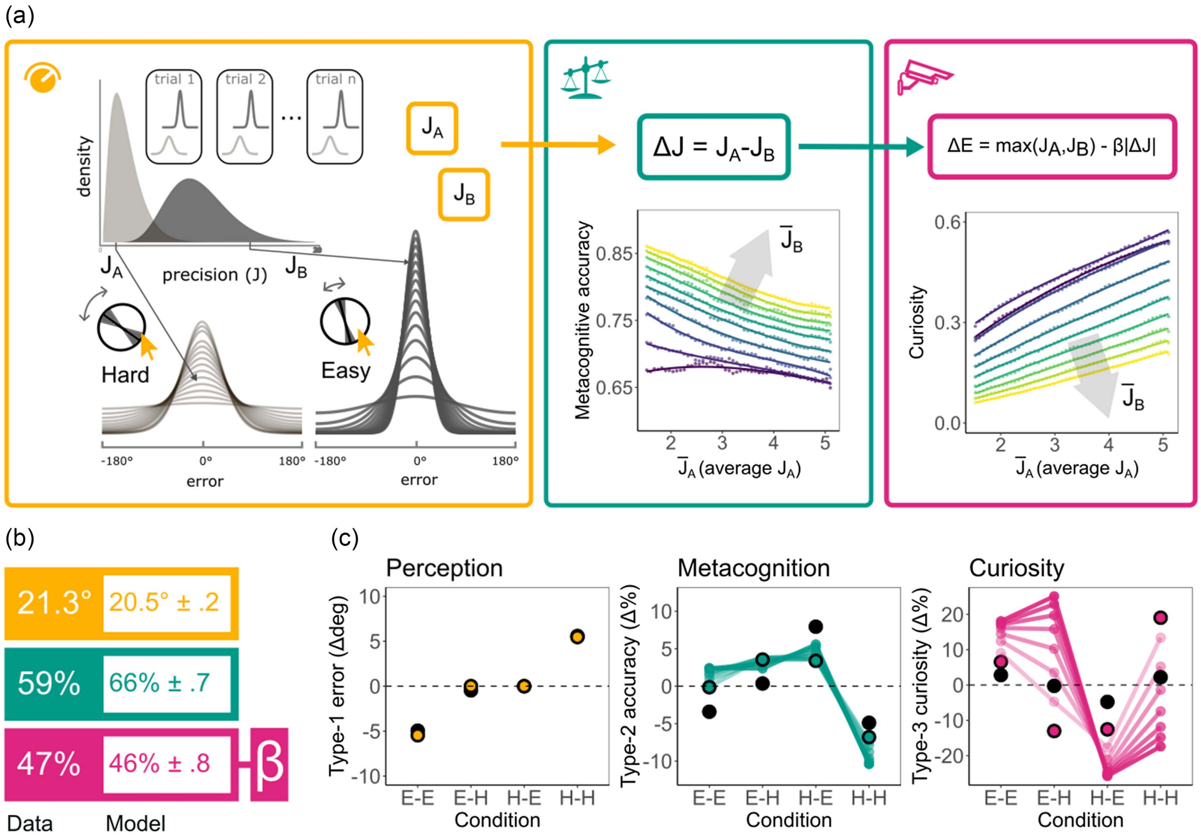
Hence, a scaling factor (β) is required to prevent metacognitive evidence from always being lower than perceptual evidence during the curiosity decision (Figure 3a right panel). This scaling factor is inversely proportional to the curiosity for metacognitive feedback: A small factor leads to lower metacognitive evidence and therefore more curiosity for metacognition. Notably, this factor can be considered as the observer's prior knowledge about the reliability of their metacognition relative to the quality of their perception (as a function of task demand and constraints). To estimate this factor at the group level, we consider the overall probability of selecting metacognitive feedback, which was not significantly different from 50%. We identified the most robust scaling factor across a variety of realistic perceptual evidence distributions ($\beta = 1.42$, see Supplemental Material). Our model therefore has four parameters (the scale and shape of the γ for hard and easy trials) and a fixed scaling factor for curiosity. Our model qualitatively predicts perceptual accuracy across and between conditions and also captures certain trends in metacognitive accuracy (Figure 3c middle panel). Yet, overall metacognitive accuracy was higher for the ideal observer model, suggesting suboptimal metacognitive decisions, a recurring finding in the literature (Mamassian, 2016). An aspect that is not predicted by the model is the difference between "H-E" and "E-H" conditions: While the ideal observer predicts no difference, the empirical data depart from this, showing a marked increase in metacognitive accuracy for the H-E condition, probably also mirrored by curiosity. As shown in Figure 2c–e, this is attributable to a systematic metacognitive recency bias. Regarding curiosity, the model does not capture the "E-H" versus "H-E" imbalance either. A biased observer model, where metacognitive evidence is biased in favor of the second trial in the pair, allows for such variability in both metacognition and curiosity (Figure 3c, colored lines; see Supplemental Material for more details).

Discussion

In his work *On Being a Busybody (De curiositate)*, Plutarch extols the benefits of shifting one's curiosity from things outwards and "turning it inwards" (Plutarch, 1939 ed., p. 477): Only those who actively probe their cognition shall ultimately understand others. During planning, acting, or while exercising restraint, human cognition must systematically adjust to multiple sources of external and internal uncertainty. In a similar vein to how external uncertainty can be reduced by enhancing the predictive power of an agent's world model, internal uncertainty may—to some reasonable extent—be mitigated by improving the model an agent has of themselves. In the current work, we investigated the comparative curiosity for feedback on the accuracy of perception and metacognition. We designed a new paradigm that

Figure 3

Descriptive Observer Model



Note. (a) *Left "perception" panel (yellow)*: For each trial, the response error is drawn from a circular normal distribution (Von Mises), centered on the correct orientation as shown in the bottom figures. The internal precision of the representation varies across trials according to a γ distribution. The top figure illustrates how the γ distribution changes with the scale parameter. Light and dark gray represent hard and easy trials, respectively. *Middle "metacognition" panel (green):* Simulations for an ideal observer. The observer assesses and compares the precision of each trial, preferring the one with greater precision. The panel plots the proportion of correct metacognitive judgments (y-axis) against the difference in average perceptual precision for Trial A (x-axis). The change in precision is achieved through the scale of the γ distribution. The color gradient shows how metacognitive accuracy shifts with an increase in the offset of Trial B's average precision relative to Trial A's following a change in scale. Darker lines represent an increasingly similar average precision between Trials A and B (with Trial B's precision becoming equal to Trial A's at the limit). Note that the "proportion correct" indicates the highest level of performance attainable given the limits of perceptual sensitivity. *Right "Curiosity" panel (pink):* Simulations for an ideal observer. Here, the observer decides to focus on metacognition (or otherwise on perception) by weighing the perceptual evidence of the chosen trial, $\max(J_A, J_B)$, against the metacognitive evidence (ΔJ). To arbitrate between these, metacognitive evidence must be scaled by a fixed proportionality factor, which could be considered as the observer's prior knowledge about metacognitive self-reliability ($\beta = 1.42$; see the main text). This diagram displays the average curiosity for metacognition as a function of average perceptual precision in each trial of a pair (J_A, J_B). Darker lines represent an increasingly similar average precision between Trials A and Trial B. (b) This section details the empirical sensitivity for perception, metacognition, and curiosity, alongside the values derived from the model; β represents the curiosity scaling factor. (c) Displays the relative sensitivity per condition for each decision order as forecasted by the observer model. Black-circled colored dots indicate ideal observer model predictions (with 500-iteration bootstrapped SEM, error bars are all smaller than the dots), based exclusively on perceptual responses from experiments. Black dots show the empirical group averages from experiments. Colored dots and lines represent predictions for an observer model with different metacognitive bias (darker color for greater bias in the [0, 2] range). E-E = easy-easy; E-H = easy-hard; H-E = hard-easy; H-H = hard-hard; SEM = standard error of the mean. See the online article for the color version of this figure.

allowed us to dissociate the interest for each of these dimensions, building on the realistic assumption that confidence is often used to arbitrate between subsequent decisions and direct our actions.

Our main finding is the existence of an inherent curiosity drive about metacognitive ability, which, under adequate circumstances, prevails over a curiosity about perceptual ability and decision outcome. The finding that individuals are ready to sacrifice knowledge about their gain in a specific trial to confirm the accuracy of their metacognitive skills offers compelling support for the idea that humans greatly value metacognitive knowledge. We designed our study to present participants with a choice: to know the outcome of a specific decision they made or to assess the accuracy of their metacognition without knowing the outcome of this decision. This approach intentionally separated perceptual/cognitive and metacognitive information to explore their interactions.

In real life, the choice to explore metacognition could take various forms: It could involve seeking advice, assessing the history of previous confidence judgments, or even practicing self-evaluation specifically to improve decision monitoring. For example, a researcher unsure about the strength of their grant proposal might reflect on their history of application successes and failures. If they recognize a pattern of overconfidence, they might try to understand the main drivers of this miscalibration and seek feedback from colleagues before submission. Interestingly, metacognitive self-reliability may also serve as a benchmark to infer others' reliability: Even when confidence in a decision is high, observers still seek information from an advisor to test the advisor's reliability (Carlebach & Yeung, 2023).

One aspect of our findings is the notable interindividual variability in this curiosity about metacognition (Figure 2a). Previous research has uncovered a variety of motives for human curiosity that transcend the direct value of resolving uncertainty (Kobayashi et al., 2019). Furthermore, general curiosity for the self has been found to vary across the population, with greater curiosity often associated with a heightened sensitivity to other people's expressions, increased levels of distress, and a concern regarding the most effective ways to cope with worry (Litman et al., 2017). Metacognitive representations themselves have been described as at least partially resulting from cultural adaptation, suggesting a rich, varied phenotype of metacognitive traits across individuals and communities (Heyes et al., 2020). Therefore, identifying the specific personality features, cultural underpinnings, and neurotypical factors influencing curiosity about metacognition could be useful in enhancing individualized interventions.

Beyond individual differences, we identified a correlation between curiosity for metacognition and both the difficulty and quality of metacognitive judgment. This link was significant in incentivized experiments (Experiments 1 and 3) but absent in the absence of reward (Experiment 2), hinting at the impact of payoff structures on metacognitive monitoring (Locke et al., 2020). By orthogonally manipulating perceptual and metacognitive evidence, we could discern the influence of each evidence type on curiosity. Participants displayed increased curiosity about their perception in trials with high metacognitive evidence ("E-H" and "H-E" conditions), keen to know their perceptual decision outcome. Conversely, in scenarios with lower metacognitive evidence ("H-H" and "E-E"), curiosity toward metacognition rose. A similar pattern emerged for metacognitive accuracy across different conditions: Incorrect metacognitive decisions led participants to prefer feedback on metacognition rather than perception, suggesting a form of active monitoring. Implementing such a monitoring and error detection at the metacognitive level could be facilitated via a reliability estimate of metacognitive computations (Recht et al., 2022). Our observer model (Figure 3) illustrates how sensory evidence may be gathered, evaluated, and transformed in decisions related to metacognition and curiosity, while also highlighting how metacognitive bias may affect the pursuit of feedback.

Some metacognitive inefficiencies are systematic, while others stem from unpredictable noise (Shekhar & Rahnev, 2021). Attention, for example, is known to systematically affect confidence judgments (e.g., Recht et al., 2019, 2023; Sarı et al., 2024). The literature distinguishes between metacognitive bias and sensitivity: Bias indicates average behavior shifts across evidence levels, whereas sensitivity measures the metacognitive signal's responsiveness to changes in primary evidence (Fleming & Lau, 2014). Metacognitive bias, being systematic, is tied to specific environments, conditions, or agents and remains relatively stable within them. Both bias and sensitivity influence metacognitive evaluations, with our accuracy measure integrating these aspects into a unified metric, both ultimately having real-life consequences. However, our experimental design also allowed us to uncover a metacognitive bias toward the second decision in a pair (Figure 2c–e). This tendency must be considered a bias as it did not reflect any true difference in performance between the two trials but is likely to result from a more general retrieval strategy that underlies a range of memory systems: The most recently primed items may be most readily accessible (e.g., Baddeley, 2007; Baddeley & Hitch, 1977, 1993). In this context, participants may use a fluency heuristic and judge the second trial easier than the first. With regard to our paradigm, the strong evidence for such a bias in all experiments highlights the existence of a postdecisional transformation specific to the metacognitive evaluation (a fluency heuristic being one possible form of postprocessing). Intriguingly, curiosity appeared to respond to this bias qualitatively: with a tendency toward higher curiosity when the environment (i.e., the pair of trials) contradicted the bias and lower when it reinforced it (Figure 2c). While further replication is needed, this suggests that observers can detect subtle imbalances in metacognitive evidence over time, choosing to monitor metacognitive performance more when a given context (i.e., them being worse in the second trial because it is objectively more difficult) contradicts their expectation (i.e., them being better in the second trial). We propose that variations in metacognitive abilities across tasks or contexts may stimulate curiosity and monitoring: When clashing with the environment perceptual evidence, a maladaptive metacognitive bias would lower metacognitive evidence, and the observer may gather metacognitive feedback for in-depth assessment and improvement.

Inefficient decision making is often seen as correctable through effective monitoring and control. Confidence, for example, has been suggested to act as an internal learning signal in the absence of external feedback (Guggenmos et al., 2016; Ptasczynski et al., 2022). However, if supervisory mechanisms themselves are inefficient, they might contaminate the overall decision process and heighten the risk of failure. Consequently, metacognitive processes too may well benefit from monitoring and fine-tuning in certain circumstances. This form of metacontrol can be facilitated by reallocating cognitive focus "on-the-fly" among different decision orders, such as from perceptual to metacognitive monitoring. There is already evidence that such a trade-off occurs between perception and metacognition. For example, Maniscalco et al. (2017) found that reducing metacognitive demand can lessen fatigue's impact on perceptual decision making. Rosenbaum et al. (2022) suggested that first-order decision-making strategies affect subsequent metacognitive evaluations. These findings illustrate an arbitrage between cognition and metacognition, where enhancements in one domain may be offset by impairments in the other. Therefore, it is crucial for the observer to adapt their monitoring strategy in accordance with their goals. The trade-off between perceptual and metacognitive feedback requests observed in the present study supports this notion, highlighting the importance of flexible and goal-dependent monitoring.

To better understand this arbitrage, it is essential to differentiate between the initial source of decision evidence and its subsequent processing and transformation. Although both perceptual and metacognitive decisions may share a common sensory origin, at least initially, this does not preclude the existence of higher level bottlenecks and/or parallel processing streams. For instance, we found the recursive, nested evaluation of a perceptual decision (i.e., "meta-metacognition") to be associated with increasing levels of noise, resulting in lower—albeit still above-chance—accuracy for higher order evaluations (Recht et al., 2022). However, this cost is likely dependent on the nature of the metacognitive evaluation itself (Recht et al., 2022; Zheng et al., 2023). In contrast, when participants are asked to rate their confidence in the same decision twice in close succession, metacognitive accuracy actually increases with reevaluation (Elosegi et al., 2024). This distinction confirms the importance of considering the specific context, task, and goals surrounding metacognitive evaluations and the role of precision at each stage of the decision hierarchy (Yon & Frith, 2021).

In our proposed observer model, metacognitive evidence was estimated as the difference in perceptual evidence between trials in a pair: This approach resulted in metacognitive evidence consistently being lower than perceptual evidence for the best trial in each pair. This difference aligns with the typical decline in available evidence observed during metacognitive computations (Shekhar & Rahnev, 2021). However, if the metacognitive evidence used for curiosity judgment was consistently lower empirically, observers would continuously seek metacognitive feedback, contradicting our findings that participants opted for perceptual feedback half of the time. To address this, we introduced a scaling factor that normalizes metacognitive evidence, making it comparable to perceptual evidence. This adjustment enables a balance where curiosity can shift between perception and metacognition, in a similar vein to our empirical observations (Figure 3c). What could be the theoretical and psychological underpinnings of this normalization? First, it reflects the subjective valuation of metacognitive knowledge relative to cognitive/perceptual knowledge, a valuation likely to depend on a variety of instrumental, hedonic, and cognitive factors. In this sense, our model is in line with a recent framework suggesting that "people will be more likely to want information relating to concepts that are frequently activated and highly interconnected to other concepts in their mental models (for example 'self' or 'human')" (Sharot & Sunstein, 2020, p. 16). Second, the curiosity scaling factor could be considered as a prior or belief an individual has of their own metacognitive reliability: a lower value suggesting lower confidence in their own metacognitive—or "introspective"—ability. As such, it permits a flexible valuation of the metacognitive inferences an agent makes over and above objective perceptual evidence (e.g., Olawole-Scott & Yon, 2023).

Our model assumed that observers had access to the precision of their perception, and used this precision readout to decide at the metacognitive level. A possible alternative view defines confidence as an inference about self-consistency, rather than accuracy or precision. Under this framework, the goal of the observer is to be consistent with themselves across decisions, rather than always being objectively accurate (Caziot & Mamassian, 2021; Koriat, 2024). Although we did not directly compare these approaches, further work may consider which normative framework better accounts for metacognitive curiosity.

Finally, curiosity has been portrayed as an intrinsic motivational force: an appetite for information without an extrinsically rewarding goal (Berlyne, 1954; Gottlieb & Oudeyer, 2018). Other theorists argued for a more liberal definition, considering both extrinsic and intrinsic aspects of curious exploration as irrevocably linked (Kidd & Hayden, 2015). The curiosity drive regarding metacognition found in the present study may well reflect a plurality of distinct motives, encompassing both an urge to know oneself and a potential desire to improve one's decisions in the long run. How these features evolve and interact across contexts, curriculums, and agents is a relevant question for further studies. Is metacognitive curiosity driven by the desire to acquire metacognitive knowledge (an informational goal) or by the motivation to gain self-understanding (a sociocognitive goal)? Moreover, our current paradigm focused on the local evaluation of decisions. The emergence of metacognitive evaluations at different time scales is likely to play an important role in (mal)adaptive decision making, with significant implications for mental health (Seow et al., 2021). Understanding how this form of introspective curiosity interacts with the development of more global and long-term confidence estimates may prove useful in the development of new models and interventions. Overall, our work contributes to a better understanding of self-regulation by highlighting a systematic and adaptive curiosity about one's own metacognitive abilities and suggesting a way to study its function.

Constraints on Generality

Our findings indicate that participants are systematically curious about their metacognitive accuracy in an orientation-reproduction task, with notable interindividual differences. The stimuli consisted of visual gratings in a grid. Subsequent perceptual feedback was provided as points instead of actual orientation, which makes our findings less dependent on stimulus types. We therefore expect these results to generalize to other types of perceptual tasks and potentially to nonperceptual decisions as well. An important aspect of our paradigm is the decoupling of perceptual (cognitive) feedback and metacognitive feedback. Using a task where participants' performance is measured on a continuous scale, rather than an all-or-none basis, is crucial. This approach almost eliminates the possibility for participants to infer their perceptual response from metacognitive feedback alone. Our studies mainly involved students from a U.K. university (Experiment 2) and participants from the Prolific online platform (Experiments 1 and 3), with the latter required to be fluent in English. An overreliance on English-speaking participants is known to reduce the generalizability of findings in cognitive science (Blasi et al., 2022). Given the interindividual and cross-population variance in curiosity, we believe it would be particularly important to investigate whether our results generalize, and to what extent, to other populations. This study was conducted online and was less demanding on the testing environment. Hence, we expect the results would be replicable in laboratory settings, where it is less likely for the participants to be distracted. Experiments 1 and 3 involved monetary compensation, bonus for better performers, and a shorter time frame than Experiment 2. We speculate that these factors had an effect on the results and need to be incorporated in replications. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

References

Aguilar-Lleyda, D., & de Gardelle, V. (2021). Confidence guides priority between forthcoming tasks. *Scientific Reports*, *11*(1), Article 18320. https://doi.org/10.1038/s41598-021-97884-2

Ahmed, A. S., & Duellman, S. (2013). Managerial overconfidence and accounting conservatism. *Journal of Accounting Research*, *51*(1), 1–30. https://doi.org/10.1111/j.1475-679X.2012.00467.x

Baddeley, A. (2007). Recency, retrieval and the constant ratio rule. In A. Baddeley (Ed.), *Working memory, thought, and action* (pp. 103–116). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.003.0006

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, *21*(2), 146–155. https://doi.org/10.3758/BF03202726

Baddeley, A. D., & Hitch, G. J. (1977). Recency reexamined. In S. Dornič (Ed.), *Attention and Performance VI: Proceedings of the Sixth International Symposium on Attention and Performance, Stockholm, Sweden, July 28–August 1, 1975* (1st ed., pp. 647–667). Routledge. https://doi.org/10.4324/9781003309734

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012). What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1350–1365. https://doi.org/10.1098/rstb.2011.0420

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology*, *45*(3), 180–191. https://doi.org/10.1111/j.2044-8295.1954.tb01243.x

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). National Academy Press.

Buratti, S., & Allwood, C. M. (2015). Regulating metacognitive processes—Support for a meta-metacognitive ability. In A. Peña-Ayala (Ed.), *Metacognition: Fundaments, applications, and trends: A profile of the current state-of-the-art* (pp. 17–38). Springer.

Carlebach, N., & Yeung, N. (2023). Flexible use of confidence to guide advice requests. *Cognition*, *230*, Article 105264. https://doi.org/10.1016/j.cognition.2022.105264

Cavalan, Q., Vergnaud, J. C., & de Gardelle, V. (2023). From local to global estimations of confidence in perceptual decisions. *Journal of Experimental Psychology: General*, *152*(9), 2544–2558. https://doi.org/10.1037/xge0001411

Cavalan, Q., Vergnaud, J. C., & de Gardelle, V. (2024). Confidence in metacognition-Type 3 judgments in the context of perceptual decisions. *HAL-SHS Preprint*. https://doi.org/10.13140/RG.2.2.16640.24328

Caziot, B., & Mamassian, P. (2021). Perceptual confidence judgments reflect self-consistency. *Journal of Vision*, *21*(12), Article 8. https://doi.org/10.1167/jov.21.12.8

Cox, M. T. (2005). Metacognition in computation: A selected research review. *Artificial Intelligence*, *169*(2), 104–141. https://doi.org/10.1016/j.artint.2005.10.009

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *The Journal of General Psychology*, *132*(4), 335–346. https://doi.org/10.3200/GENP.132.4.335-346

Edwards-Lowe, G., La Chiusa, E., Olawole-Scott, H., & Yon, D. (2024). *Information seeking without metacognition*. PsyArXiv. https://doi.org/10.31234/osf.io/cf4a7

Elosegi, P., Rahnev, D., & Soto, D. (2024). Think twice: Re-assessing confidence improves visual metacognition. *Attention, Perception, & Psychophysics*, *86*(2), 373–380. https://doi.org/10.3758/s13414-023-02823-0

Fisher, P. L., & Wells, A. (2008). Metacognitive therapy for obsessive–compulsive disorder: A case series. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(2), 117–132. https://doi.org/10.1016/j.jbtep.2006.12.001

Fitzgibbon, L., & Murayama, K. (2022). Counterfactual curiosity: Motivated thinking about what might have been. *Philosophical Transactions of the Royal Society B*, *377*(1866), Article 20210340. https://doi.org/10.1098/rstb.2021.0340

Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology*, *75*(1), 241–268. https://doi.org/10.1146/annurev-psych-022423-032425

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280–1286. https://doi.org/10.1098/rstb.2012.0021

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, Article 443. https://doi.org/10.3389/fnhum.2014.00443

Friedemann, M., Fox, C. A., Hanlon, A. K., Tighe, D., Yeung, N., & Gillan, C. M. (2024). Confidence biases in problem gambling. *Journal of Behavioral Addictions*, *13*(2), 650–664. https://doi.org/10.1556/2006.2024.00030

Geurts, L. S., Cooke, J. R. H., van Bergen, R. S., & Jehee, J. F. M. (2022). Subjective confidence reflects representation of Bayesian probability in cortex. *Nature Human Behaviour*, *6*(2), 294–305. https://doi.org/10.1038/s41562-021-01247-w

Gottlieb, J., & Oudeyer, P. Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), 758–770. https://doi.org/10.1038/s41583-018-0078-0

Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *The Journal of Finance*, *64*(2), 549–578. https://doi.org/10.1111/j.1540-6261.2009.01443.x

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife*, *5*, Article e13388. https://doi.org/10.7554/eLife.13388

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition*, *44*(2), 229–241. https://doi.org/10.3758/s13421-015-0552-0

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://doi.org/10.4324/9780203887332

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in Cognitive Sciences*, *24*(5), 349–362. https://doi.org/10.1016/j.tics.2020.02.007

Kaustia, M., & Perttula, M. (2012). Overconfidence and debiasing in the financial industry. *Review of Behavioral Finance*, *4*(1), 46–62. https://doi.org/10.1108/19405971211261100

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kobayashi, K., Ravaioli, S., Baranès, A., Woodford, M., & Gottlieb, J. (2019). Diverse motives for human curiosity. *Nature Human Behaviour*, *3*(6), 587–595. https://doi.org/10.1038/s41562-019-0589-3

Koriat, A. (2024). Subjective confidence as a monitor of the replicability of the response. *Perspectives on Psychological Science*. Advance online publication. https://doi.org/10.1177/17456916231224388

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*(7985), 115–121. https://doi.org/10.1038/s41586-023-06668-3

Lee, A. L. F., de Gardelle, V., & Mamassian, P. (2021). Global visual confidence. *Psychonomic Bulletin & Review*, *28*(4), 1233–1242. https://doi.org/10.3758/s13423-020-01869-7

Litman, J. A., Robinson, O. C., & Demetre, J. D. (2017). Intrapersonal curiosity: Inquisitiveness about the inner self. *Self and Identity*, *16*(2), 231–250. https://doi.org/10.1080/15298868.2016.1255250

Locke, S. M., Gaffin-Cahn, E., Hosseinizaveh, N., Mamassian, P., & Landy, M. S. (2020). Priors and payoffs in confidence judgments. *Attention, Perception, & Psychophysics*, *82*(6), 3158–3175. https://doi.org/10.3758/s13414-020-02018-x

Lysaker, P. H., Gagen, E., Moritz, S., & Schweitzer, R. D. (2018). Metacognitive approaches to the treatment of psychosis: A comparison of four approaches. *Psychology Research and Behavior Management*, *11*, 341–351. https://doi.org/10.2147/PRBM.S146446

Mamassian, P. (2016). Visual confidence. *Annual Review of Vision Science*, *2*(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Mamassian, P. (2020). Confidence forced-choice and other metaperceptual tasks. *Perception*, *49*(6), 616–635. https://doi.org/10.1177/0301006620928010

Maniscalco, B., McCurdy, L. Y., Odegaard, B., & Lau, H. (2017). Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *The Journal of Neuroscience*, *37*(5), 1213–1224. https://doi.org/10.1523/JNEUROSCI.2271-13.2016

Mohr, G., Ince, R. A., & Benwell, C. S. (2024). Information search under uncertainty across transdiagnostic psychopathology and healthy ageing. *Translational Psychiatry*, *14*(1), Article 353. https://doi.org/10.1038/s41398-024-03065-w

Olawole-Scott, H., & Yon, D. (2023). Expectations about precision bias metacognition and awareness. *Journal of Experimental Psychology: General*, *152*(8), 2177–2189. https://doi.org/10.1037/xge0001371

Pescetelli, N., Hauperich, A. K., & Yeung, N. (2021). Confidence, advice seeking and changes of mind in decision making. *Cognition*, *215*, Article 104810. https://doi.org/10.1016/j.cognition.2021.104810

Plutarch. (1939). *On being a busybody* (Vol. VI), Loeb Classical Library.

Ptasczynski, L. E., Steinecker, I., Sterzer, P., & Guggenmos, M. (2022). The value of confidence: Confidence prediction errors drive value-based learning in the absence of external feedback. *PLOS Computational Biology*, *18*(10), Article e1010580. https://doi.org/10.1371/journal.pcbi.1010580

Recht, S., de Gardelle, V., & Mamassian, P. (2021). Metacognitive blindness in temporal selection during the deployment of spatial attention. *Cognition*, *216*, Article 104864. https://doi.org/10.1016/j.cognition.2021.104864

Recht, S., Jovanovic, L., Mamassian, P., & Balsdon, T. (2022). Confidence at the limits of human nested cognition. *Neuroscience of Consciousness*, *2022*(1), Article niac014. https://doi.org/10.1093/nc/niac014

Recht, S., Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. *Scientific Reports*, *9*(1), Article 11622. https://doi.org/10.1038/s41598-019-48063-x

Recht, S., Mamassian, P., & de Gardelle, V. (2023). Metacognition tracks sensitivity following involuntary shifts of visual attention. *Psychonomic Bulletin & Review*, *30*(3), 1136–1147. https://doi.org/10.3758/s13423-022-02212-y

Rosenbaum, D., Glickman, M., Fleming, S. M., & Usher, M. (2022). The cognition/metacognition trade-off. *Psychological Science*, *33*(4), 613–628. https://doi.org/10.1177/09567976211043428

Salovich, N. A., & Rapp, D. N. (2021). Misinformed and unaware? Metacognition and the influence of inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(4), 608–624. https://doi.org/10.1037/xlm0000977

Sarı, İ. D., Recht, S., & Lunghi, C. (2024). Learning to discriminate the eye-of-origin during continuous flash suppression. *European Journal of Neuroscience*, *60*(1), 3694–3705. https://doi.org/10.1111/ejn.16373

Schneegans, S., Taylor, R., & Bays, P. M. (2020). Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(34), 20959–20968. https://doi.org/10.1073/pnas.2004306117

Seow, T. X. F., Rouault, M., Gillan, C. M., & Fleming, S. M. (2021). How local and global metacognition shape mental health. *Biological Psychiatry*, *90*(7), 436–446. https://doi.org/10.1016/j.biopsych.2021.05.013

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, *4*(1), 14–19. https://doi.org/10.1038/s41562-019-0793-1

Shekhar, M., & Rahnev, D. (2021). Sources of metacognitive inefficiency. *Trends in Cognitive Sciences*, *25*(1), 12–23. https://doi.org/10.1016/j.tics.2020.10.007

Sherman, M. T., & Seth, A. K. (2024). Knowing that you know that you know? An extreme-confidence heuristic can lead to above-chance discrimination of metacognitive performance. *Neuroscience of Consciousness*, *2024*(1), Article niae020. https://doi.org/10.1093/nc/niae020

Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, *26*(3), 317–339. https://doi.org/10.1017/S0140525X03000086

Stotz, O., & von Nitzsch, R. (2005). The perception of control and the level of overconfidence: Evidence from analyst earnings estimates and price targets. *Journal of Behavioral Finance*, *6*(3), 121–128. https://doi.org/10.1207/s15427579jpfm0603_2

Van Camp, L., Sabbe, B. G. C., & Oldenburg, J. F. E. (2019). Metacognitive functioning in bipolar disorder versus controls and its correlations with neurocognitive functioning in a cross-sectional design. *Comprehensive Psychiatry*, *92*, 7–12. https://doi.org/10.1016/j.comppsych.2019.06.001

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

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

Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, *31*(17), R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044

Zheng, Y., Recht, S., & Rahnev, D. (2023). Common computations for metacognition and meta-metacognition. *Neuroscience of Consciousness*, *2023*(1), Article niad023. https://doi.org/10.1093/nc/niad023

Received May 17, 2024
Revision received September 13, 2024
Accepted September 18, 2024 ■