Adaptive curiosity about metacognitive ability

Supplementary Material

Samuel Recht[1], Canqi Li[2*], Yifan Yang[1*] & Kaiki Chiu[3]

[1] Department of Experimental Psychology, University of Oxford, Oxford, UK

[2] Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, USA

[3] Department of Psychology, Yale University, New Haven, USA

* These authors contributed equally

Corresponding author: SR, samuel.recht@gmail.com

The experiments were programmed using Javascript (JsPsych library). Statistical analyses were conducted using the *R* programming language. Observer models were programmed in Python.

General approach for statistical analyses

We employed linear (or generalised linear) mixed-effects models using the lme4 package in *R* (version 1.1-35.1). This approach provides the flexibility to incorporate both fixed effects, representing our experimental predictors, and random effects, capturing the variability among participants. For model comparison and validation, we used likelihood ratio tests and Akaike Information Criterion (AIC) values. This enabled us to systematically assess the explanatory power of the models and ensure the selection of the most parsimonious representation of the underlying relationships. To test main effects and interactions, we employed a hierarchical modelling strategy. Initially, we gauged the impact of independent variables individually by contrasting each one to a null model that only considered participant variability (random intercept). After identifying significant predictors, a comprehensive model incorporating all main effects and their interaction was developed and compared to the previously selected best model.

Bayes factors were calculated using the “ttestBF” functions for *t*-tests, from the BayesFactor *R* package (version 0.9.12-4.7, Morey & Rouder, 2018). For all analyses, we used the default prior distribution provided with the package. In cases where a one-tailed t-test was conducted, we adjusted the null interval accordingly (e.g., setting the nullInterval parameter to [0,Inf) when testing for a positive difference). This approach was also applied to rate differences, which theoretically should be constrained to the interval [0,1]. However, we maintained the [0,Inf) interval for consistency with the classical parametric t.test() function, which does not allow for restricting the alternative hypothesis to a specific range. This approach effectively makes the BF_{10} estimate more conservatory in these cases, as the wider interval reduces the likelihood of favouring the alternative hypothesis. We always report the Bayes factor in favour of the alternative hypothesis (BF_{10}), with values above 3 providing evidence in favour of the alternative hypothesis and values below 0.33 evidence in favour of the null.

Tables

T-tests are all two-tailed unless stated (following a pre-registered directional effect).

Table S1

Mixed-effects model comparison for effect of condition on perceptual error

Experiments	df	χ^2	ΔAIC	*p*
Exp 1	3	47.5	48.51	<.001
Exp 2	3	69.2	69.12	<.001
Exp 3	3	86.8	85.29	<.001

Table S2

Mixed-effects model comparison for effect of trial order on perceptual error

Experiments	df	χ^2	ΔAIC	*p*
Exp 1	1	1.50	1.70	0.22
Exp 2	1	0.02	0.08	0.88
Exp 3	1	<.001	0.75	0.99

Table S3

Mixed-effects model comparison for effect of condition x trial order interaction on perceptual error

Experiments	df	χ^2	ΔAIC	*p*
Exp 1	4	90.83	91.44	<.001
Exp 2	4	145.52	143.37	<.001
Exp 3	4	142.58	141.95	<.001

Table S4

Results of t-test for metacognitive accuracy relative to chance level

Experiments	df	Above Chance (>0.5) Mean [95% CI]	*t*-value	Cohen's *d*	Bayes Factor	*p*
Exp 1	19	0.10 [.06, .14]	5.06	1.13	388.71	< .001
Exp 2	24	0.09 [.05, .14]	7.78	1.55	2.82×10^5	< .001
Exp 3	59	0.08 [.06, .10]	8.72	1.13	2.68×10^9	< .001

Table S5

Mixed-effects model comparison for effect of condition on metacognitive accuracy

Experiments	df	χ^2	ΔAIC	*p*
Exp 1	3	8.92	2.92	0.03
Exp 2	3	32.56	26.56	<.001
Exp 3	3	10.73	4.73	0.01

Table S6

Results of t-test for metacognitive bias in favour of the second trial

Experiments	df	Mean [95% CI]	*t*-value	Cohen's *d*	Bayes Factor	*p*
Exp 1	19	0.10 [.05, .15]	4.07	0.91	52.66	< .001
Exp 2	24	0.09 [.05, .14]	4.36	0.87	138.32	< .001
Exp 3	59	0.08 [.04, .11]	4.43	0.57	490.79	< .001

Table S7

Results of t-test for average curiosity for metacognitive feedback relative to chance in the 4AFC task (> 25%, including all options in rate calculation, one tailed)

Experiments	df	Mean [95% CI]	*t*-value	Cohen's *d*	Bayes Factor	*p*
Exp 1	19	0.17 [.03 Inf]	2.07	0.46	2.6	0.02
Exp 2	24	0.11 [.01 Inf]	1.96	0.39	2.1	0.03
Exp 3	59	0.13 [.06 Inf]	3.20	0.41	26.15	0.001

Table S8

Results of t-test for average curiosity relative to 50% (excluding 'no feedback' requests)

Experiments	df	Above Chance (>0.5) Mean [95% CI]	*t*-value	Cohen's *d*	Bayes Factor	*p*
Exp 1	19	-0.07 [-.24, .09]	-0.96	0.21	0.35	0.35
Exp 2	22	-0.10 [-.22, -.03]	-1.58	1.40	0.65	0.13
Exp 3	58	-0.05 [-.14, .04]	-1.14	0.15	0.26	0.26

Table S9

Mixed-effects model comparison for condition x metacognitive accuracy interaction

Experiments	df	χ^2	ΔAIC	*p* Values
Exp 1	6	8.35	3.65	0.21
Exp 2	4	0.49	-7.51	0.97
Exp 3	3	11.25	5.25	0.01

Table S10

Results of two-tailed (Exp 1) and one-tailed (Exp 2 and 3) t-tests for difference in average curiosity for metacognition between correct and wrong metacognitive judgments

Experiments	df	Mean [95% CI]	*t* -value	Cohen's *d*	Bayes Factor	*p*
Exp 1	15	0.09 [.03, .15]	3.04	0.76	6.40	0.01
Exp 2	20	0.01 [-.03 Inf]	0.29	0.06	0.29	0.78
Exp 3	47	0.06 [.02 Inf]	2.58	0.37	5.96	0.01

Descriptive Bayesian observer model

To understand in more detail the nature of nested judgments from Type-1 perception to Type-3 curiosity, we developed a Type-2 metacognitively-ideal Bayesian observer model. According to the model, for each pair of trials, the observer first estimates the orientation of each stimulus (perceptual decision), and then compares the evidence for each estimated orientation, selecting the one with greater evidence (metacognitive decision). For Type-3 curiosity, the observer then compares perceptual evidence to metacognitive evidence to infer which type of feedback would *subjectively* lead to the greater information gain. This strategy would allow the observer to maximise information intake depending on the context, considering that both perceptual and metacognitive evidence are valuable. It is important to note that our observer model is ideal at the metacognition (Type-2) level, but not at the curiosity (Type-3) level: it does not make any assumption on the value function of the Type-3 decision for a given observer, but rather simply assume that an individual has some level of intrinsic appetite for metacognitive feedback. This approach was chosen because contrary to a Type-3 confidence judgement (i.e., the confidence in the quality of a confidence judgement, see Recht et al., 2022; Zheng et al., 2023), there isn't a definitive normative ground yet for Type-3 curiosity.

The evidence at each order being always in Type-1 sensory evidence units, this approach provides a form of common currency across decision orders. Here, we use a nested observer model to describe the perceptual decision for each trial in a given pair, the metacognitive decision, and Type-3 curiosity. Critically, the observer makes multiple, recursive inferences on perceptual evidence (Type-1) and metacognitive evidence (Type-2) and finally compares these estimates to fine-tune curious exploration (Type-3). Drawing from earlier research (e.g., Keshvari et al., 2012; van den Berg et al., 2012, 2014), we use Fisher information (J) to measure Type-1 precision and quantify Type-2/Type-3 evidence. This metric sets a minimal threshold for the variance of any unbiased estimator: it is a measure of the amount of information in a specific random variable (Ly et al., 2005). Importantly, this measure of precision is in Type-1 units (or sensory units), making it comparable between cognitive orders (from Type-1 to Type-3).

Perception (Type-1 decision)

Stimulus distribution

On each trial, a grid of oriented Gabor corrupted by noise is presented to the observer. The goal of the observer is to estimate the average orientation of the grid. The variance of the grid depends on the condition. Here, for simplicity, we consider a stimulus within a relatively stable context (i.e., one condition). In our paradigm, the stimulus' average orientation is sampled from a circular uniform distribution (between 0 and pi, which can be rescaled); here, we will directly consider it in a full circle space.

$$\langle \theta \rangle \sim \mathrm{Uniform}(0,\ 2pi) \text{ (Eq. 1)}$$

where $\langle \theta \rangle$ is the average angle for a given grid.

The final composite stimulus is made of 20 samples from a Von Mises distribution (with the concentration parameter depending on the condition, $\kappa_H = 3$, $\kappa_E = 12$). Here, we consider a fixed concentration parameter:

$$\theta_1, \theta_2, \ldots, \theta_{20} \sim \text{Von Mises}(\langle\theta\rangle, \kappa_{Exp}) \text{ (Eq. 2)},$$

where κ_{Exp} is the concentration parameter (the inverse of the variance) for the stimuli grid in the considered condition. Here, we consider that $\langle\theta\rangle$ represents the ground truth for the external environment.

Measurement distribution

For the following steps, we use the Fisher information (J) as a measure of sensory precision in place of the concentration parameter. The relation between Fisher information and κ is given by:

$$J(\kappa) = \kappa \frac{I_1(\kappa)}{I_0(\kappa)} \text{ (Eq. 3)},$$

where I_0 and I_1 are the modified Bessel functions of orders 0 and 1, respectively. Equation 3 can be numerically inverted to estimate the concentration of an arbitrary precision level ($\kappa(J)$). Note that in the following paragraphs, we will distinguish between κ_{Exp} (the parameter used to generate our stimulus from the experimenter's perspective) and κ, the parameter describing the internal precision of the observer.

The sensory measurement itself is an estimate of the average orientation of the grid. We assume that it is unbiased but subject to (circular) Gaussian sensory noise. The measurement distribution for Type-1 is therefore centred on the correct value $\langle\theta\rangle$:

$$m_{\text{Type-1}} \sim \text{Von Mises}\left(\langle\theta\rangle, \kappa(J)\right) \text{ (Eq. 4)},$$

where $m_{\text{Type-1}}$ is the measurement in a given trial, $\langle\theta\rangle$ the composite stimulus orientation and J the overall precision for Type-1 judgement. We assume that the precision is variable from trial to trial (following previous work, e.g., van den Berg et al. (2012)). This variability results from multiple sources, and we remain largely neutral to the specific sources (i.e., external noise, vigilance/attention…). To model the distribution of precision levels across trials, we opt for a Gamma distribution, principally because it has been shown to capture noise in working memory (e.g., Schneegans, Taylor & Bays, 2020; van den Berg et al., 2012) and attention tasks involving confidence judgments (e.g., Recht et al., 2021), but also because it is a natural choice for a positive-only continuous variable like precision.

$$J_{\text{Type-1}} \sim \text{Gamma}(\tau, \gamma) \quad \text{(Eq. 5)},$$

where J is the precision in a given trial, τ is the shape $(\tau > 0)$ and γ the scale $(\gamma > 0)$. The average precision is $\bar{J} = \tau\gamma$.

Inference process

Next, we define the prior distribution, the likelihood function and the posterior distribution for a Bayesian observer.

For simplicity, we consider that the observer has an accurate representation of the statistics of the experimental environment (as provided in the instruction); the prior is therefore uniform over the orientation space (rescaled to a full circle for simplicity).

$$P(\langle\theta\rangle) = \frac{1}{2pi}, \quad \text{for } \langle\theta\rangle \in [0,\ 2pi] \quad \text{(Eq. 6)}$$

The likelihood, or the probability of the measurements given the stimulus, takes the same shape and variance as the sensory noise distribution (Eq. 4):

$$P(m_{\text{Type-1}}|\langle\theta\rangle) = \text{Von Mises}(\langle\theta\rangle; m_{\text{Type-1}}, \kappa(J))$$

(Eq. 7)

Finally, the observer inverts the generative model via Bayes' rule to obtain the posterior distribution, and the final measurement estimation is inferred from the posterior via a cost function:

$$p(\langle\theta\rangle|m_{\text{Type-1}}) \propto p(m_{\text{Type-1}}|\langle\theta\rangle)p(\langle\theta\rangle)$$ (Eq. 8)

$$L_2(\langle\hat{\theta}\rangle, \langle\theta\rangle)$$

Here, we consider a squared error loss function (L_2) equivalent to computing the (circular) mean of the posterior. Given our assumption of a uniform prior, the posterior distribution and its precision are reducible to the likelihood parameters; the precision of the estimate is therefore J. Note that under non-uniform priors, the Type-1 evidence might differ from the precision of the likelihood function and has to be estimated numerically.

Metacognition (Type-2 decision)

Every two trials, the observer is required to compare their performance in the first and second trial to make a confidence judgement. The Type-2 evidence is therefore the signed difference in evidence (or precision) read out from the two posterior distributions (trial A and trial B in the pair). We remain neutral to the exact readout of confidence from the posterior, any monotonic estimate of the inverse posterior uncertainty could work. We consider that it reliably reflects the precision of the likelihood, J. We assume the readout to be ideal and therefore to incur no extra noise.

$$m_{\text{Type-2}} = J_A - J_B \text{ (Eq. 9)},$$

where J_A and J_B are the precision levels for the first and second trial in a pair.

We also considered a biased observer model, following the strong metacognitive bias observed empirically (see **Figure 2c and e**). The metacognitive bias is implemented as a consistent shift in evidence imbalance between first and second trial in a pair.

$$m_{\text{Type-2}} = J_A - J_B - c \text{ (Eq. 10)},$$

Where c is the bias in precision, a positive value corresponding to a bias towards trial B.

Finally, the Type-2 response (2AFC) is defined using the sign of the Type-2 evidence:

$$R_{\text{Type-2}} = \begin{cases} \text{Trial A} & \text{if } m_{\text{Type-2}} > 0 \\ \text{Trial B} & \text{otherwise} \end{cases} \text{ (Eq. 11)}$$

and the Type-2 evidence for the considered response is the absolute value of the Type-2 measurement (i.e., the estimated difference in precision between the two trials):

$$E_{\text{Type-2}} = |m_{\text{Type-2}}| \text{ (Eq. 12)}$$

Curiosity (Type-3 decision)

We define curious exploration as the result of an arbitrage between Type-1 and Type-2 evidence. The Type-3 decision involves choosing from 4 distinct alternatives. Discarding the "no" trials, three alternatives remain available, and we consider the evidence for each of these alternatives here in order to make a judgement. The evidence for the first trial and the second trial was defined previously (J_A and J_B). Considering that the vast majority of Type-1 exploration was made for the selected trial (see **Figure 2a**), we only consider the evidence for the trial that was selected for the sake of simplicity (that is, we consider that the observer will never request Type-1 feedback for the rejected trial):

$$E_{\text{Type-1}} = J_{\text{Selected Type-1}} \text{ (eq. 12)}$$

Now, the observer has to compare the precision of Type-1 (J_{Type-1}) to the precision of Type-2 (J_{Type-2}) and select the option with the lower precision. The Type-3 measurement for a given curiosity decision is therefore the difference in evidence between the selected Type-1 and the Type-2 decision:

$$m_{\text{Type-3}} = E_{\text{Type-1}} - \beta E_{\text{Type-2}} \text{ (Eq. 13),}$$

where β is a Type-3 evidence weight accounting for the tendency to select one type of feedback more often than the other. Without this scaling factor, the observer would systematically select Type-2, because the evidence is by definition lower for Type-2 than Type-1 (Type-2 evidence being calculated as the difference in Type-1 evidence).

Finally, the Type-3 response is defined using the sign of the Type-3 evidence:

$$R_{\text{Type-3}} = \begin{cases} \text{Type-2 feedback} & \text{if } m_{\text{Type-3}} > 0 \\ \text{Type-1 feedback} & \text{otherwise} \end{cases} \quad \text{(Eq. 14)}$$

The Type-3 evidence for the considered response is the absolute value of the Type-3 measurement (i.e., the estimated difference in precision between the two trials):

$$E_{\text{Type-3}} = |m_{\text{Type-3}}| \quad \text{(Eq. 15)}$$

Note that the evidence across orders (E_{Type-1}, E_{Type-2}, E_{Type-3}) are all in the same Fisher information unit (J).

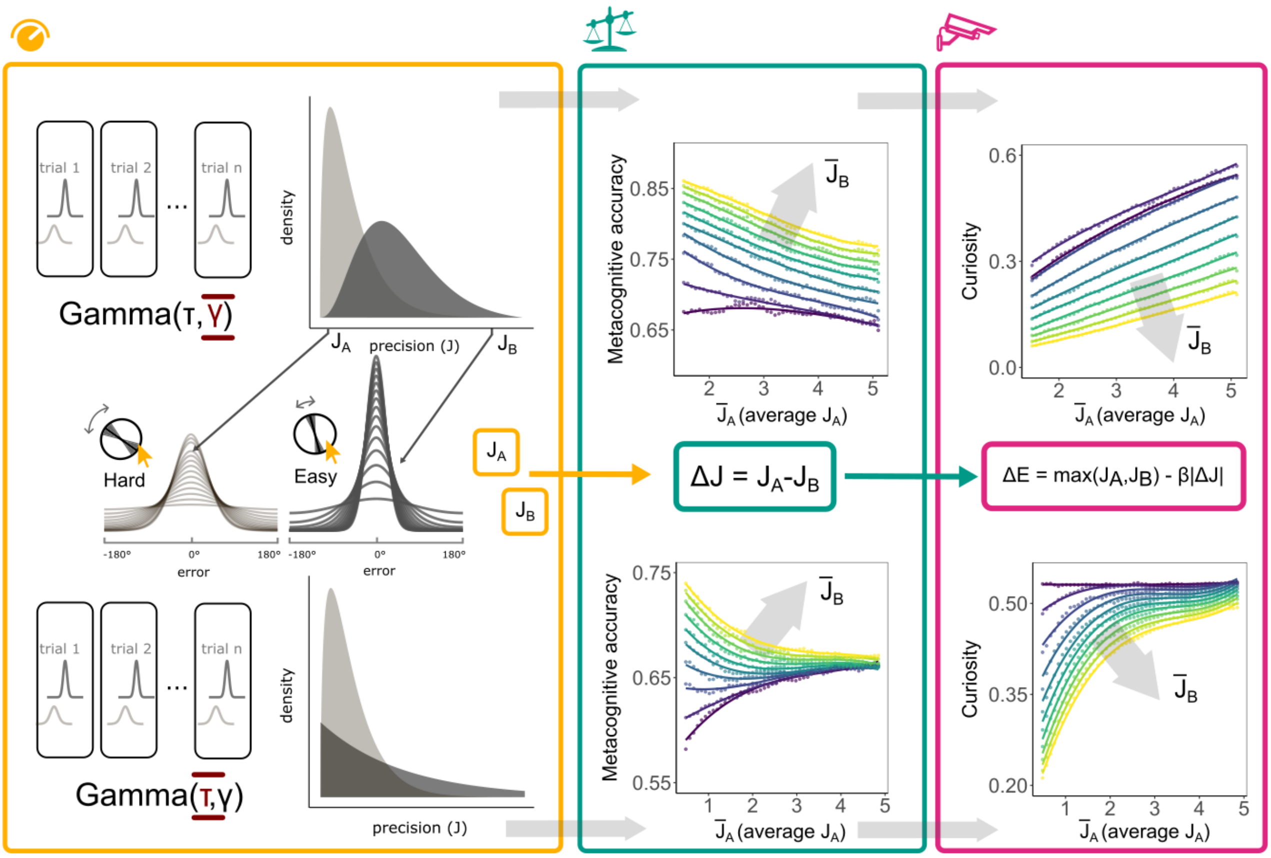

Figure S1. Model predictions. **(a) Left 'perception' panel (yellow):** For each pair of trials, the response error is drawn from a circular normal distribution (Von Mises), centred on the correct orientation as shown in the middle figures. The internal precision of the representation varies across trials according to a Gamma distribution. The top-right figure illustrates how the Gamma distribution changes with the scale parameter (γ), while the bottom-right figure shows changes as a function of the shape parameter (τ). Light and dark grey represent hard and easy trials, respectively. **Middle 'metacognition' panel (blue):** An ideal observer assesses and compares the precision of each trial, preferring the one with greater precision. The blue panel plots the proportion of correct metacognitive judgments against the difference in average perceptual precision. The change in precision is achieved either through the scale or shape of the Gamma distribution (top or bottom figures, respectively). The x-axis displays the average precision for Trial A in a pair, with changes mediated either by scale (top figure) or shape adjustments (bottom). The colour gradient shows how metacognitive accuracy shifts with an increase in Trial B's average precision relative to Trial A, either via change in scale (top panel) or shape parameter (bottom). Note that the 'proportion correct' indicates the highest level of performance attainable given the limits of perceptual sensitivity. **Right 'Curiosity' panel (pink)**: the observer then decides to monitor metacognition (or perception otherwise) by comparing the perceptual evidence of the selected trial ($\max(J_A,J_B)$) to the metacognitive evidence (the absolute of the difference in precision), deciding request feedback for the decision with lower evidence. Note that for the arbitration to be possible, it is necessary to scale metacognitive evidence by a proportionality factor considered fixed (β). The figure plots the rate of metacognition feedback request as a function of average perceptual precision (β = 1.42, see Model fitting procedure section below for details).

Model fitting procedure

Perception

The model was fit exclusively on the perceptual error. We aggregated the data from all experiments, leading to 1147 trials per condition in total. For each trial, we calculated the response error as the signed angular difference between the participant's response and the true orientation:

$$\epsilon = \langle\hat{\theta}\rangle - \langle\theta\rangle \text{ (Eq. 16)}$$

Where $\langle\hat{\theta}\rangle$ is the response of the participant, and $\langle\theta\rangle$ the true orientation of the stimulus. The probability of an error is given by an infinite mixture of Von Mises, each with a different concentration (κ) following a Gamma distribution:

$$p(\epsilon) = \int Gamma(J(\kappa), \tau, \gamma) \mathrm{VonMises}(\epsilon; \kappa)\, d\kappa.$$

(Eq. 17)

Where κ is the concentration of a Von Mises, τ and γ are the shape and scale parameters of the Gamma distribution.

Metacognition

We did not fit any extra parameter for metacognition in the ideal observer model. For the biased observer model, we added a bias term to the metacognitive evidence (**Figure S2**).

Curiosity

To estimate a tangible curiosity rate (the probability of requesting metacognitive feedback), rather than fitting the parameter on one data point, we decided to select a rate that would lead to an average of 50% for as many τ and γ combinations (i.e., evidence distributions) as possible within a reasonable range of possible values given our experimental data. To this end, we tested 1,000 equally spaced biases between 0.1 and 5, across a variety of different Gamma distributions ($\tau \in [0.5, 2]$, $\gamma \in [0.5, 6]$, with a factorial combination of 12 distinct parameter values), and took the bias that led to a rate of 50% the most frequently ($\beta = 1.42$). Of note, this approach was not meant to provide a representative value in absolute terms, but rather to allow for curiosity to vary as a function of perceptual and metacognitive evidence.

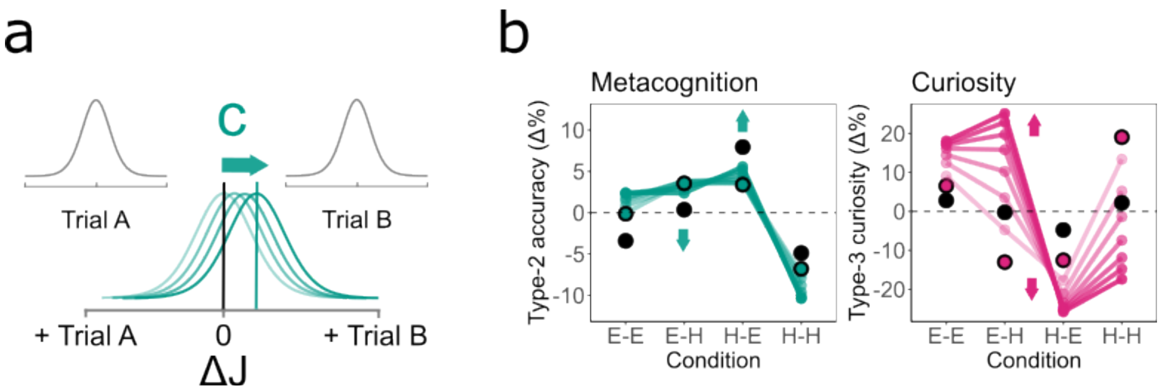

Figure S2. Effect of metacognitive bias. (a) Adding a bias term (c) to the observer metacognitive decision is equivalent to systematically shift the evidence (Δ_J), here in favour of the second trial in a pair. (b) Increasing metacognitive bias leads to a marked imbalance between the Easy-Hard and Hard-Easy conditions, despite the two having the same average evidence. (c) The metacognitive bias has a cascading effect and also affects curiosity in an opposite pattern (lower metacognitive accuracy = greater curiosity). Similarly to metacognition, the curiosity predictions from the model do also capture the relative shift in between conditions, but suffers from a notable overshoot in terms of magnitude.

The Python code for the observer model is available on the OSF repository: https://osf.io/45wvb/

Simulations (Figures 3a and S1)

To generate the plots on **Figure 3a**, and **Figure S1**, we fixed either the scale or the shape parameter of our perceptual model (making it equal to the estimated value from our aggregated observer, $\tau = 0.96$ and $\gamma = 5.08$ for the easy trials, $\tau = 1.52$ and $\gamma = 0.54$ for the hard trials, we used the values for the easy trials only), and estimated the number of times where both $J_A < J_B$ and $\epsilon_A > \epsilon_B$ were verified, the first being the decision signal, and the second the objective outcome of the decision (the average precision can be calculated as the product of the shape and scale parameters). Note that because of sampling variability, it is possible for the observer to be correct about the difference in precision, despite the metacognitive decision being wrong (because the errors' difference does not match the difference in precisions).

Instructions provided to the participants

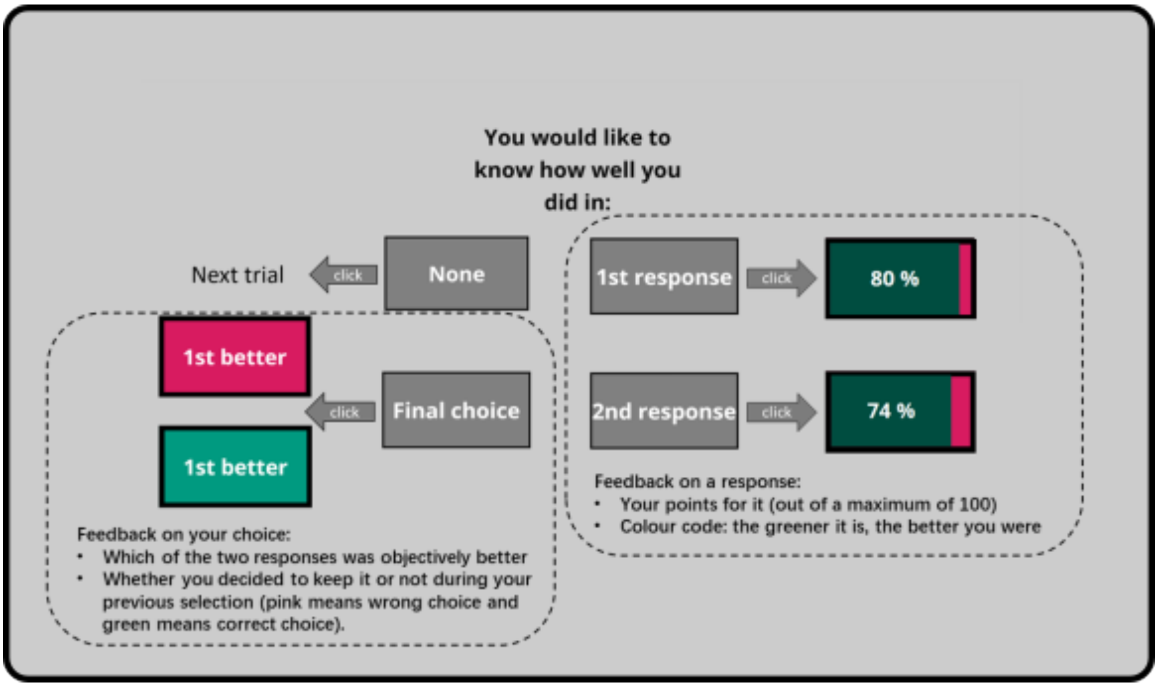

Figure S3: Caption of the instructions provided to the participants prior to the study commencing: The order of the options' display was randomised across trials.

Additional notes on the pre-registration for Exp 2 and 3

Experiment 2

Upon finding a medium effect size in Experiment 1 (the pre-registration erroneously mentioned 0.57, the correct value being 0.75) via the one-sample (two-sided) paired *t*-test on the difference in Type-2 feedback request rate between correct and incorrect Type-2 judgments, we determined that a minimum of 24 participants was required to ensure an 85% power to detect the effect (for a Cohen's *d* of 0.57 in a one-tailed t-test). A Bayesian stopping rule was applied to determine the final sample size (less than one third or greater than three for the Bayesian version of the aforementioned *t*-test). After collecting data from 25 participants, we reached the evidence threshold for the null (0.29 < 1/3) and therefore stopped data collection. For Exp 2 and 3, our design initially involved three more "global" subjective estimates of performance, recorded during each break using rating scales, for exploratory analyses. We did not analyse these ratings in the present study.

Link to pre-registration: https://aspredicted.org/ii3gb.pdf

Experiment 3

We used the same target rule as for Experiment 2. In addition to the unambiguous Bayes factor (BF_{10}) for the aforementioned t-test, we also added a secondary BF target of reaching an unambiguous BF for the condition effect. As initially pre-registered, we aimed to use both anovaBF for estimating the Bayes Factor (BF) on the condition effect and ttestBF for estimating the Bayes Factor on the effect of metacognitive accuracy on feedback requests. These BF estimates were used during data collection with a Bayesian stopping rule. However, we identified two key errors: first, anovaBF is not compatible with our mixed-effects logistic model analysis, since it only considers aggregated averages over trials, while the pre-registered logistic regression model involves a trial-by-trial prediction; second, we used a two-tailed test in ttestBF during data collection, which has lower power compared to the expected one-tailed test given our pre-registered directional prediction. At N = 60 (our pre-registered upper bound), we did not cross the BF target for the anovaBF test (BF_{10} = 0.43), but we did for the one-tailed ttestBF (BF_{10} = 5.96), a two-sided t-test leading to a BF_{10} of 3.00.

Link to pre-registration: https://aspredicted.org/y932m.pdf

References

Keshvari, S., Van den Berg, R., & Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS One*, *7*(6), e40216. https://doi.org/10.1371/journal.pone.0040216

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P., & Wagenmakers, E. J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, *80*, 40-55.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: computation of bayes factors for common designs.

Recht, S., de Gardelle, V., & Mamassian, P. (2021). Metacognitive blindness in temporal selection during the deployment of spatial attention. *Cognition*, *216*, 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). https://doi.org/10.1093/nc/niac014

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*, *117*(34), 20959-20968. https://doi.org/10.1073/pnas.200430611

Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological review*, *121*(1), 124.

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*, *109*(22), 8780-8785.

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