A hierarchy of metacognitive capacities

Sophie Bavard (sophie.bavard@gmail.com)
Paris Brain Institute - ICM, Paris, France

Andrew McWilliams
Department of Experimental Psychology, University College London, London, UK

Flora Chartier
Paris Brain Institute - ICM, Paris, France

Stephen M. Fleming
Department of Experimental Psychology, University College London, London, UK

Marion Rouault
Paris Brain Institute - ICM, Paris, France

Abstract

How do we evaluate our overall performance on a task? While most research on metacognition focuses on local confidence - our ability to assess accuracy on a trial-by-trial basis - real-world decisions often rely on global confidence, a broader judgment of overall success, measured through self-performance estimates. This study introduces a novel method to investigate self-performance estimates in memory and perceptual tasks. Participants made decisions in blocks containing two item categories and then selected the category they believed they performed better on, reflecting self-performance estimates. By analyzing the relationship between self-performance estimates and factors such as local difficulty, accuracy, response times (RTs), and local confidence, our study shows that self-performance estimates rely on different cues depending on the domain: both accuracy and local confidence shaped self-performance estimates in memory, while only local confidence did in perception, without a contribution of RT or difficulty. These findings advance our characterization of metacognitive processes and pave the way for the development of Interventions to modify metacognition.

Keywords: metacognition; confidence; memory; perception; self-performance estimates

Introduction

Metacognition - our ability to monitor and evaluate our own cognitive processes - is essential for adaptive decision-making and learning (Fleming, 2024; Metcalfe & Shimamura, 1994). Within metacognition, confidence judgments play a crucial role in guiding how we allocate cognitive resources and adjust decision strategies (Fleming et al., 2012; Risko & Gilbert, 2016; Son & Metcalfe, 2000). While metacognition has been extensively studied through local confidence (trial-by-trial accuracy assessments), real-world decisions often require global confidence, a higher-order evaluation of overall abilities, measured through self-performance estimates. Despite its importance in controlling behavior and decision-making, the cognitive mechanisms shaping self-performance estimates remain largely unknown. Here, we study how self-performance estimates are formed in two vastly-studied cognitive domains: memory and perception. In both domains, participants completed decision-making tasks in blocks composed of two different item categories and expressed their self-performance estimates at the end of each block, by choosing in which category they believed they performed best. By investigating the relationship between self-performance estimates and local difficulty, accuracy, RT and local confidence, we show that self-performance estimates rely on different information across domains. In memory tasks, both accuracy and local confidence contributed to self-performance estimates, whereas in perceptual tasks, local confidence alone played a dominant role. Notably, neither fluctuations in RT nor objective difficulty influenced self-performance estimates beyond fluctuations in accuracy and local confidence, despite their well-established links to local confidence. By shifting the focus from local, trial-based monitoring to the construction of self-performance estimates, our study provides new insights into self-evaluation, decision-making, and interindividual differences, paving the way for applications in education, mental health, and metacognitive interventions.

Results

Experiment assessing local and global metacognition. To investigate the domain-specificity of global metacognition, we employed a cross-sectional design incorporating novel measures of local and global metacognition across two sessions separated by three weeks and two domains: memory and visual perception (McWilliams et al., 2023). In memory, participants were asked to memorize a category of items and subsequently select the familiar stimulus when paired with a distractor. In perception, participants were presented with an array of multiple identical red and blue shapes and then asked whether there were more red or blue shapes. For both domains, retrospective ("local") confidence judgments were elicited after each decision using a horizontal visual sliding scale, the ends of which were labelled "complete guess" and "absolutely certain". In each domain, participants were presented with 6 blocks of 20 interleaved trials, with half of the trials featuring stimuli from a specific item category. At the end of each block, participants were asked in which category they thought they performed best - a measure shown to reflect task-level ("global") confidence (Rouault et al., 2019).

Construction of self-performance estimates in each domain. At the end of each block, participants indicated in which category they believed they performed best. To investigate which factors contribute to the formation of these self-performance estimates in each domain, we performed a generalized linear mixed model (GLMM) predicting end-of-block category choice as a function of local difficulty, accuracy, RT and confidence (Figure 1). In the memory domain, we found a positive contribution of accuracy $(\text{estimate} = 0.075, \text{SE} = 0.028, t = 2.67, p = .0075)$ and confidence $(\text{estimate} = 0.23, \text{SE} = 0.034, t = 6.84, p < .0001)$ to self-performance estimates, without a significant contribution of other indices (all $|\text{estimate}| < 0.05, p > .10$). In contrast, in the perception domain, we only found a positive effect of confidence $(\text{estimate} = 0.095, \text{SE} = 0.028, t = 3.42, p = .00064)$ on self-performance estimates, without other local cues contributing (all $|\text{estimate}| < 0.05, p > .08$). Overall, these results indicate that different local contributors are involved in memory and perception when participants construct their self-performance estimates, but in both domains their decision confidence was a key contributor to their final category decision.

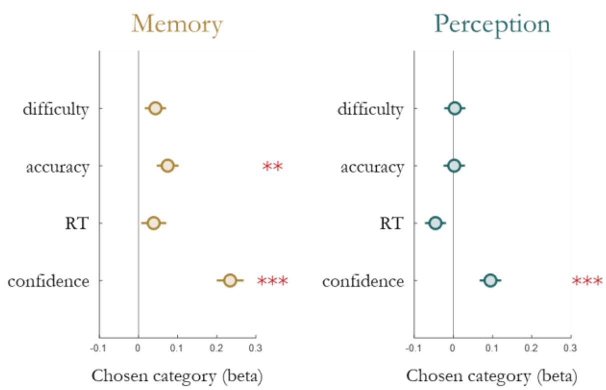

Figure 1. GLMM predicting self-performance estimates in the memory (left) and perception (right) domains as a function of the difference in local difficulty, accuracy, RT and local confidence between categories.

Between-domains comparison of metacognitive bias and metacognitive efficiency. First, metacognitive bias, defined as the discrepancy between mean confidence and mean accuracy, was positive in both memory $(\text{mean} \pm \text{SD} = 0.071 \pm 0.073,\ t(51) = 6.98, p < .0001)$ and perception $(\text{mean} \pm \text{SD} = 0.061 \pm 0.076,\ t(51) = 5.81, p < .0001)$, indicating overconfidence in both domains. We observed equivalent levels of overconfidence across the two domains (memory vs. perception, $t(51) = 1.22, p = .23$), and these levels were significantly correlated across participants (Spearman's $\rho(50) = .69, p < .0001$). This suggests that the degree of overconfidence may be a stable individual trait transcending domains, in line with prior work (Ais et al., 2016; Binnendyk et al., 2024). Second, metacognitive efficiency, defined as how well confidence discriminates between correct and incorrect decisions with respect to first-order accuracy, was quantified using the M-ratio meta-d'/d' (Fleming, 2017; Maniscalco & Lau, 2012). We found across participants an M-ratio of 1.17±0.16 in memory and 0.69±0.13 in perception. Metacognitive efficiency was significantly higher in the memory compared to the perception domain (95% HDI on difference in M-ratio: [0.37, 0.75]). However, no significant correlation in metacognitive efficiency was observed between domains (95% HDI on correlation coefficient: [-0.84, 0.75]). These findings support the idea that metacognitive efficiency is domain-specific rather than a trait at the individual level, in line with previous reports (Rouault et al., 2018).

Discussion

In recent years, there has been growing interest in understanding the mechanisms underlying metacognition and its role in decision-making across domains. The current study reveals several important insights into local and global metacognition. The domain-specific nature of metacognitive efficiency suggests that interventions aimed at improving metacognitive efficiency may need to be tailored to specific domains (Carpenter et al., 2019). At the same time, the stability of confidence and metacognitive bias across domains points to the existence of global metacognitive traits and would instead argue for intervention effects likely to transfer to new domains. These traits may serve as a foundation for individual differences in decision-making and learning, with potential applications in clinical and educational contexts.

References

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

Binnendyk, J., Li, S., Costello, T., Hale, R., Moore, D. A., & Pennycook, G. (2024). *Is Overconfidence a Trait? An Adversarial*

Collaboration. OSF. https://doi.org/10.31234/osf.io/awugz

Carpenter, J., Sherman, M. T., Kievit, R. A., Categoryh, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology. General*, *148*(1), 51-64. https://doi.org/10.1037/xge0000505

Fleming, S. M. (2017). HMeta-d : Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), nix007. https://doi.org/10.1093/nc/nix007

Fleming, S. M. (2024). Metacognition and Confidence : A Review and Synthesis. *Annual Review of Psychology*, *75*(Volume 75, 2024), 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

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

McWilliams, A., Bibby, H., Steinbeis, N., David, A. S., & Fleming, S. M. (2023). Age-related decreases in global metacognition are independent of local metacognition and task performance. *Cognition*, *235*, 105389. https://doi.org/10.1016/j.cognition.2023.105389

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition : Knowing about knowing* (p. xiii, 334). The MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1141. https://doi.org/10.1038/s41467-019-09075-3

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains : Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, *1*, e17. https://doi.org/10.1017/pen.2018.16

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204-221. https://doi.org/10.1037/0278-7393.26.1.204