Bidirectional Interactions between Local and Global Confidence Explain the Pervasive Effect of Confidence Biases

Van Marcke, Hélène[1,2]; Braem, Senne[2] & Desender, Kobe[1]
1: Brain and Cognition, Faculty of Psychology and Educational Sciences, KU Leuven
2:Department of Experimental Psychology, Ghent University

Abstract

Research into decision confidence, i.e. the subjective feeling about the correctness of a decision, indicates important links between pervasive confidence biases and a variety of real-life outcomes and psychiatric symptoms. While decision confidence is traditionally studied in isolation, recent theoretical accounts posit that such "local" confidence about decisions interacts with global confidence, i.e. the general feeling about the ability to perform a task. Here, we provide empirical and modelling evidence for such bidirectional influences. Using a manipulation of global confidence, we measured both constructs in a perceptual decision paradigm. We found that local confidence is indeed informed by global confidence in addition to accuracy and RTs, while global confidence is informed by local confidence rather than accuracy and RTs. By explicitly modelling global confidence in a signal-detection theory framework, we provide computational evidence for bidirectional interactions between local and global confidence that explain the pervasive nature of confidence biases.

Keywords: metacognition; decision confidence; global confidence; confidence biases; SDT

Introduction

Every decision we make is accompanied by a sense of confidence about its correctness. Metacognition research, which studies the higher-order thoughts we have about our own thinking, typically studies such decision confidence in isolation. However, a recent theoretical proposal (Seow et al., 2021) posits that metacognition is best described as a hierarchy, in which local confidence interacts with the general feeling one has about their ability to perform a task, i.e. global confidence. In the current study, we aimed to provide empirical evidence for such bidirectional influences between local and global confidence.

Method

Participants (N=40) performed two perceptual decision-making tasks, but received manipulated feedback (within-subject) during the training phase of each task. By respectively telling participants they scored among the best/worst participants, this previously validated paradigm (Van Marcke et al., 2024) causally induced positive/negative feelings about the ability to perform the task without affecting objective performance. Participants then proceeded with a manipulation-free testing phase, rating their local confidence in each decision and their global confidence every 10 trials. By doing so, we could track naturally occurring interactions between local and global confidence, while the feedback manipulation allowed us to compare these interactions both in the context of underconfidence (i.e. negative feedback) and overconfidence (i.e. positive feedback).

Results

Behavioural results. As expected, participants were more/less confident after positive/negative feedback both locally and globally ($p<.001$), despite unaffected accuracy and reaction times ($p>.05$), indicating a selective, direct influence of our manipulations on second-order but not first-order cognition. Next, we tested the hypothesised bidirectional influences between both confidence levels statistically. Top-down, local confidence was best predicted by a model including global confidence as a predicting factor ($p<.01$) on top of the expected predictive value of feedback, accuracy and reaction time ($p<.05$). Bottom-up, we found that the most reliable predictors for global confidence were the local confidence ratings of the preceding 10 trials (all $p<.05$) rather than accuracy or reaction times. A recency effect indicated that the predictive value of a local confidence rating was higher for trials closer to upcoming global confidence queries.

Computational results. To provide mechanistic insights into these bidirectional influences, we explicitly modelled global confidence in a signal detection theory framework as expected d' (*ed'*) (Rahnev et al., 2015), i.e. an observer's subjective belief about the distance in underlying choice distributions and hence, difficulty of the task. An observer assuming a high *ed'* thus internally represents the underlying distributions for each choice alternative as having little overlap, which conceptually is associated with holding high global confidence as the observer believes the task to be easy (see Fig. 1A). To model recurrent interactions with local confidence, our model:

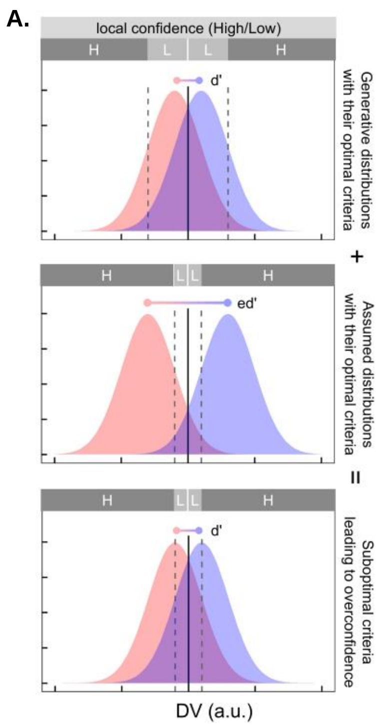
(i) assumes *ed'* to be a running average of local confidence, with a weighting factor θ determining the strength of updating:

$$ed'_t = (1-\theta) * ed'_{t-1} + \theta * conf_{t-1} \qquad \textit{where t=trial}$$

(ii) sets the criteria used to compute local confidence as a function of *ed'* (Fig. 1A, middle):

$$c_{conf,t} = c_{conf,t-1} * \frac{d'}{ed'} \qquad \textit{where t=trial}$$

This leads to over-/under-confidence when *ed'* results in more/less liberal confidence criteria than optimal given the actual difficulty (d'), respectively. Indeed, such a model successfully replicated all of our behavioural findings, most importantly our finding that high/low local confidence results in respective high/low global confidence, which in turn leads to high/low local confidence on upcoming trials (Fig. 1B). This model outperformed alternative models which (i) updated *ed'* based on the difference between $c_{conf,t}$ and average confidence (Rahnev et al., 2015) or (ii) set *ed'* equal to the mean confidence of the past 10 trials.



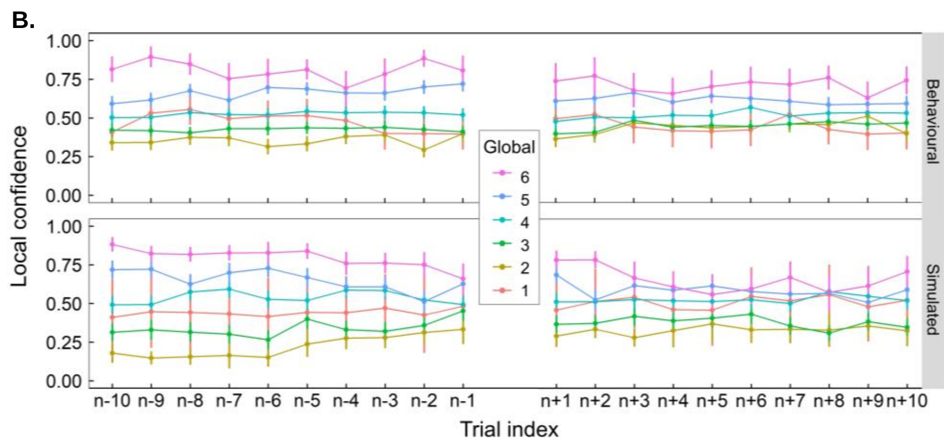


Figure 1: A. Given a difficult task (top), operating under the assumption the task is easier than it is (ed' > d', middle) will result in less stringent confidence criteria than warranted by d' (bottom), thus explaining overconfidence. **B.** In behavioural (top) as well as model-simulated (bottom) data, higher global confidence ratings were associated with higher local confidence on the ten trials preceding the global rating (n-10 to n-1), and in turn also resulted in higher local confidence on the following ten trials (n+1 to n+10), demonstrating bidirectional influences.

Discussion

Combining statistical and computational methods, our study provides empirical and modelling evidence for continuous, bidirectional local-global confidence interactions. Moreover, by demonstrating how local under- or overconfidence can be maintained over time by self-sustaining interactions with global confidence, we provide mechanistical insight into the computational dynamics underlying under- and overconfidence. This is a first step towards uncovering the interplay between local decision confidence and higher-level forms of confidence such as global or even self-confidence, which helps bridging the gap between empirical work on confidence and clinical practice.

References

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychological Science*, *26*(11), 1664–1680. https://doi.org/10.1177/0956797615595037

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

Van Marcke, H., Denmat, P. L., Verguts, T., & Desender, K. (2024). Manipulating Prior Beliefs Causally Induces Under- and Overconfidence. *Psychological Science*, *35*(4), 09567976241231572. https://doi.org/10.1177/09567976241231572