# Dis-confirmatory evidence drives confidence

**Annika Boldt (a.boldt@ucl.ac.uk)**
Institute of Cognitive Neuroscience (UCL), Alexandra House, 17-19 Queen Square
London WC1N 3AZ, United Kingdom

**Kobe Desender (kobe.desender@kuleuven.be)**
Faculty of Psychology and Educational Sciences, Tiensestraat 102 - box 3711
3000 Leuven, Belgium

**Abstract:**

**Assessing one's own decisions is a crucial component of cognition, enabling adaptive behavior. According to normative models, the brain employs available evidence optimally to evaluate decisions. If decision confidence (judging the accuracy of a decision) were to be assessed, the normative strategy would involve giving equal weight to each sample. However, recent findings instead suggest that the brain tends to overweight decision-congruent information when forming confidence judgements, a finding referred to as the "positive-evidence bias" (PEB). Here, we report the re-analysis of nine datasets involving human participants who judged the average color of eight shapes and rated their confidence in this decision. Our findings suggest that participants overly relied on evidence that conflicted with their choice, contrary to both the normative model and the PEB. We furthermore fitted a log-posterior-ratio (LPR) model to our data and show that our findings can instead be explained by a robust averaging principle.**

**Keywords: confidence; metacognition; perceptual decision making; positive-evidence bias**

## Introduction

Humans have a remarkable ability to express confidence in their internal cognitive processes, known as metacognition. Studies have shown that confidence and accuracy only moderately correlate, which sparked an interest in investigating how confidence is formed. Recent studies suggest that – rather than weighting all information equally – confidence judgments are based predominantly on decision-congruent evidence, ignoring decision-incongruent evidence (PEB; e.g., Zylberberg, Barttfeld, & Sigman, 2012; Peters et al., 2017; Khalvati, Kiani, & Rao, 2021; Maniscalco et al., 2021; Samaha & Denison, 2022). Here, we investigated the PEB using a color discrimination task in which evidence is continuously distributed across a continuum and stimuli need to be categorized as either red or blue (De Gardelle & Summerfield, 2011). This task has previously been used to illustrate that humans utilize *robust averaging*: They discard more extreme evidence and let their choices be driven mostly by the more 'inlying' samples (De Gardelle & Summerfield, 2011).

Taken together, based on previous findings from the confidence field, we should expect to find that confidence is driven by category-congruent evidence (e.g., the red-most samples out of a red stimulus matter more). On the other hand, according to the robust averaging principle the opposite should be the case: category-incongruent evidence should drive confidence because those samples lie closest to the category boundary where sensitivity is the highest.

We analyzed data from nine experiments (total $N$ = 176; Boldt, de Gardelle, & Yeung, 2017; Boldt, 2015; Boldt, Schiffer, Waszak, & Yeung, 2019; Desender, Boldt, Verguts, & Donner, 2019; Desender, Boldt, & Yeung, 2018; Desender, Murphy, Boldt, & Yeung, 2019; Boldt, 2015) and furthermore fitted six extensions of the previously proposed log posterior ratio (LPR) model (De Gardelle & Summerfield, 2011). The results of our study indicate that a model which assumes that evidence near the category boundary is most important fitted best, in line with the robust averaging principle.

## Methods

All nine datasets were collected using the same color averaging task (De Gardelle & Summerfield, 2011): On each trial, people were presented with an array of colored shapes grouped around a central fixation dot (Fig. 1A). Participants had to press one of two buttons depending on whether they thought the average color of the stimulus was red or blue. We used two orthogonal difficulty manipulations, changing both the mean and variance of the color values from which the stimulus was composed (Fig. 1B). Importantly, individual elements of a stimulus could cross the category boundary (e.g., a red stimulus could contain a blue shape). After their response, participants judged their confidence in the decision being correct (Fig. 1A).

We fitted a mixed effect model predicting people's confidence from the sorted color values, with participants nested under experiments. This allowed us to examine which areas of the feature space exhibited

the strongest influence on confidence: decision-congruent, as the PEB would predict, or decision-incongruent, as the robust averaging principle would predict.

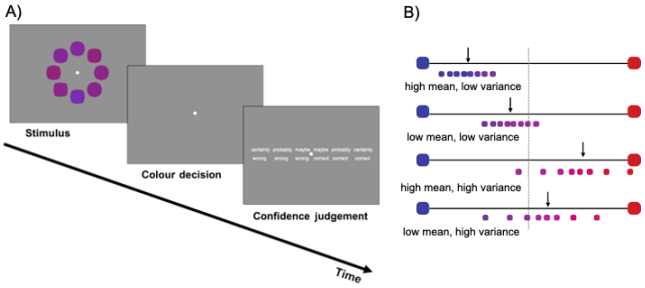


Figure 1: A) Trials were composed of a color decision followed by a confidence judgement. B) Four example stimuli, showing how difficulty was manipulated orthogonally (mean/variance color).

We furthermore fitted six different variants of the LPR model to our data (Tab. 1). As in the original model, the free parameters included a *slope* of the sigmoidal transfer function through which the color values were passed prior to being averaged, as well as a noise parameter (*decnoise*) that was added to achieve a level of randomness in choices. In addition, we tested the existence of an additional source of noise for metacognition (*metanoise*) as well as a parameter that scales the level of this noise depending on stimulus variability aimed to explain an effect from previous studies (*scaleconfnoise*; e.g., Boldt et al., 2017).

Table 1: Model comparison.

| Mod. | slope | dec noise (dn) | meta noise | scale conf noise | df | M(AIC) |
|---|---|---|---|---|---|---|
| 1 | X | X | | | 2 | -76.59 |
| 2 | X | X | X | X | 4 | -79.30 |
| 3 | X | X | =dn | X | 3 | -76.91 |
| 4 | X | X | X | | 3 | -79.31 |
| 5 | X | X | =dn | | 2 | -76.59 |
| 6 | | X | | X | 2 | -56.12 |

## Results

We found that decision-incongruent evidence had a larger influence on confidence compared to decision-congruent evidence (Fig. 2). This effect was also present on the decision level (not shown here).

A model in which metacognitive decisions were distorted using a separate source of noise (*metanoise*) fitted the data best (Mod. 4). Importantly, all models were fit to the average error rate and confidence for each experimental cell. Nevertheless, the model reproduced the qualitative pattern of the congruency effect (Fig. 2), however, only if the *slope* parameter was included.

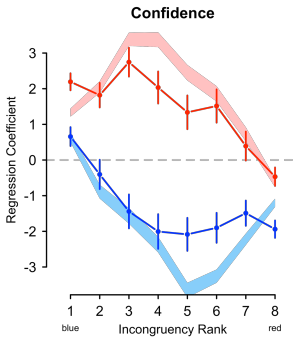


Figure 2: Regression weights predicting confidence as a function of sorted color elements, separate for the two categories (red/blue). Lines show empirical data, shaded bands show CIs from the best-fitting model.

## Discussion

In direct opposition to previous findings from the confidence literature, we observed that people overly relied on decision-incongruent evidence when forming confidence. This can be explained by the robust averaging principle: As in many naturalistic perceptual decisions, our stimuli were not artificially restricted from crossing the category boundary and therefore the decision-incongruent evidence happened to lie closest to the region of highest perceptual sensitivity. We formalized this insight using an extension of an LPR model (De Gardelle & Summerfield, 2011).

Our findings are diametrically opposed to normative models of decision confidence as well as the PEB. In other studies, participants could have learned to suppress information that can be identified to come from a separate latent source (Gershman & Niv, 2013). Our findings have implications into the question of how confidence is formed and shed light on the source of cognitive biases such as overconfidence.

## Acknowledgments

This research was funded by an ESRC AQM Studentship to AB, and by the Wellcome Trust, who awarded a Sir Henry Wellcome Postdoctoral Fellowship (206480/Z/17/Z) to AB. We would like to thank Sam Gilbert, Megan Peters and Brian Maniscalco for useful discussions.

## References


Boldt, A. (2015). *Metacognition in decision making*. University of Oxford. Retrieved from https://ora.ox.ac.uk:443/objects/uuid:5d9b2036-cc42-4515-b40e-97bb3ddb1d78

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1520–1531. https://doi.org/10.1037/xhp0000404

Boldt, A., Schiffer, A.-M., Waszak, F., & Yeung, N. (2019). Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making. *Scientific Reports*, *9*(1), 4031. https://doi.org/10.1038/s41598-019-40681-9

De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13341–13346. https://doi.org/10.1073/pnas.1104517108

Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *ELife*, *8*. https://doi.org/10.7554/eLife.43499

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

Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. (2019). A Postdecisional Neural Marker of Confidence Predicts Information-Seeking in Decision-Making. *The Journal of Neuroscience*, *39*(17), 3309–3319. https://doi.org/10.1523/JNEUROSCI.2620-18.2019

Gershman, S. J., & Niv, Y. (2013). Perceptual estimation obeys Occam's razor. *Frontiers in Psychology*, *4*(September), 623. https://doi.org/10.3389/fpsyg.2013.00623

Khalvati, K., Kiani, R., & Rao, R. P. N. (2021). Bayesian inference with incomplete knowledge explains perceptual confidence and its deviations from accuracy. *Nature Communications*, *12*(1), 5704. https://doi.org/10.1038/s41467-021-25419-4

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S. H., Basso, M. A., Lau, H., & Peters, M. A. K. (2021). Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *PLOS Computational Biology*, *17*(3), e1008779. https://doi.org/10.1371/journal.pcbi.1008779

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., … Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*, *1*(7), 1–8. https://doi.org/10.1038/s41562-017-0139

Samaha, J., & Denison, R. (2022). The positive evidence bias in perceptual confidence is unlikely post-decisional. *Neuroscience of Consciousness*, *2022*(1). https://doi.org/10.1093/nc/niac010

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*(September), 79. https://doi.org/10.3389/fnint.2012.00079