Trial-by-trial Fluctuations in Decision Criterion Shape Decision Confidence

Robin Vloeberghs (robin.vloeberghs@kuleuven.be)
KU Leuven (Tiensestraat 102, Leuven, Belgium)

Lara Navarrete (lara.navarrete@gmail.com)
KU Leuven (Tiensestraat 102, Leuven, Belgium)

Anne E Urai (a.e.urai@fsw.leidenuniv.nl)
Leiden University (Netherlands)

Kobe Desender (kobe.desender@kuleuven.be)
KU Leuven (Tiensestraat 102, Leuven, Belgium)

Abstract

Many choices we make are accompanied by a sense of confidence. Within Signal Detection Theory (SDT), confidence is traditionally conceptualized as the absolute distance between a decision variable and a decision criterion. Whereas this criterion is typically modeled as being stable over time, increasing evidence suggests that it undergoes trial-to-trial fluctuations. Based on theory and model simulations, we predict that fluctuations in the decision criterion shape confidence. Using the Hierarchical Model for Fluctuations in Criterion (hMFC) to obtain single-trial criterion estimates, we found robust evidence for this hypothesis across 15 datasets. When analyzing each dataset individually the effect was found in 13 out of 15 datasets, indicating a stable pattern across a variety of paradigms and confidence scales. Our results demonstrate that what has been previously interpreted as noise in confidence ratings, instead reflects variability driven by fluctuations in the decision criterion.

Keywords: decision making; decision confidence; signal detection theory; criterion fluctuations

Introduction

To understand the computational mechanisms underlying decision confidence researchers often rely on computational models. Signal detection theory (SDT) is one of the most influential computational frameworks of decision making (Green & Swets, 1966), and easily allows to model decision confidence. SDT assumes that observers generate an internal representation of the relevant stimulus information, typically referred to as the decision variable (DV). To make a binary decision, this DV is compared against an internal decision criterion (Figure 1; static criterion, gray line). Within SDT, decision confidence can be computed as the absolute distance between the decision variable and the decision criterion, with a larger distance indicating higher confidence (Hebart et al., 2016; Kepecs et al., 2008; Shekhar & Rahnev, 2021).

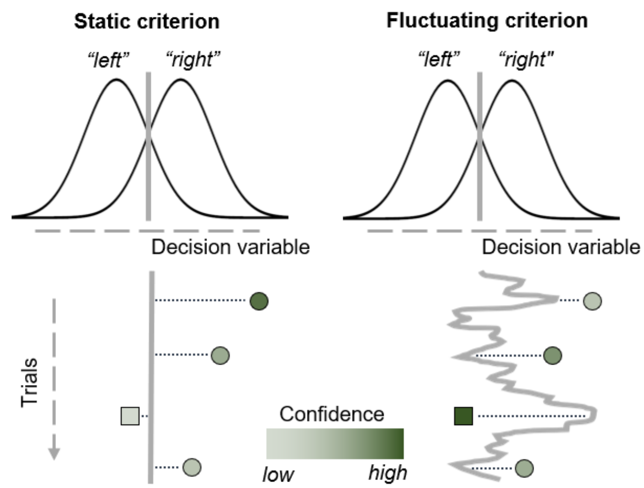

Figure 1. *For identical decision variables, predicted confidence differs depending on whether the criterion is static or fluctuating.*

By shifting the criterion, SDT naturally accounts for biases (i.e., when one response option is chosen more often than the other). For example, shifting the criterion to the right will result in overall more 'left' responses, while leaving the overall sensitivity unaffected. To estimate this decision criterion it is assumed that the criterion remains constant over all trials. However, increasing evidence suggests that computational variables, like the decision criterion, are not constant, but instead fluctuate over trials (Ashwood et al., 2022; Cowley et al., 2020; Gupta & Brody, 2022; Roy et al., 2021; Vloeberghs et al., 2024). Importantly, trial-by-trial fluctuations in the decision criterion should in theory also affect the computation of confidence, given that confidence is thought to reflect the distance between the criterion and the decision variable (Figure 1; fluctuating criterion). Therefore, the current study aims to examine whether confidence ratings are shaped by trial-to-trial fluctuations in the decision criterion.

Methods

To investigate how criterion fluctuations shape confidence we first simulated trials from a SDT observer with a fluctuating criterion. On each trial, a decision variable was sampled from a normal distribution with mean = [-3, -2, -1, 1, 2, 3] and standard deviation = 1. To obtain a binary response, the decision variable was compared to the decision criterion, which fluctuated trial-to-trial according to a first-order autoregressive process with an autoregressive coefficient = .9995 and noise standard deviation = .05. Confidence was quantified as the absolute distance between the decision variable and criterion.

Next, in a pre-registered analysis we selected 15 datasets from the Confidence Database (Rahnev et al., 2021), resulting in 382 subjects with 463.088 trials in total. Trial-to-trial fluctuations in decision criterion were estimated using Hierarchical Model for Fluctuations in Criterion (hMFC; Vloeberghs et al., 2024). Using (nested) linear mixed models, confidence was predicted based on stimulus strength, stimulus direction, and single-trial criterion estimates across all datasets and for each dataset individually.

Results

SDT simulations confirm that confidence is affected by criterion fluctuations (Figure 2A). More specifically, when the criterion fluctuates towards the right (yellow), on average the distance between a decision variable favoring the “left” stimulus and the criterion will increase, leading to higher confidence. On the contrary, when the criterion fluctuates towards the left (blue) the distance between a decision variable favoring the “left” stimulus and the criterion will decrease, leading to lower confidence. For “right” stimuli, this pattern flips.

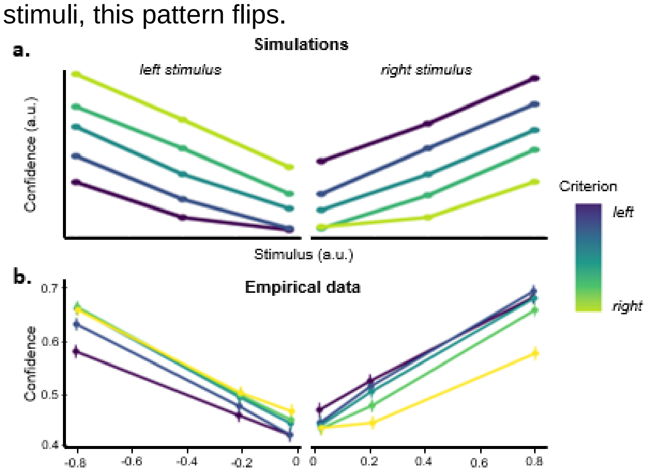

Figure 2: ***A.*** *Simulations from the SDT framework reveal a clear relation between trial-to-trial criterion fluctuations and decision confidence.* ***B.*** *Analysis of 15 datasets show a similar pattern as the simulations, confirming the hypothesis that criterion fluctuations shape decision confidence.*

In line with the model simulations, we find across 15 datasets that confidence is significantly predicted by an interaction between stimulus direction and criterion (β = -0.053, SE = 0.002, $F(1,458132)$ = 529.376, p < .001), showing that confidence is shaped by criterion fluctuations (Figure 2B). Moreover, when analyzing the datasets individually this crucial two-way interaction between stimulus direction and criterion was significant in 13 out of 15 datasets, demonstrating a robust effect independent of the specific paradigm or confidence scale (Figure 3).

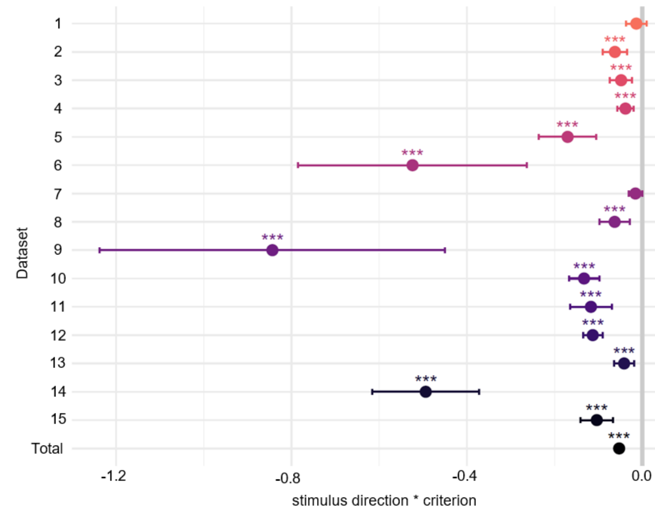

Figure 3. *Parameter estimates for the crucial interaction between stimulus direction and decision criterion and their 95% confidence intervals, separately for each dataset. The estimate plotted at the bottom, labelled “Total”, shows the parameter estimate from the model fitted across all datasets (taking into account the nesting of subjects within datasets).*

Conclusion

Our study provides robust evidence that confidence is shaped by trial-to-trial fluctuations in the decision criterion. This finding shows that variability in confidence judgements, which is typically assumed to reflect noise, actually reflects genuine computation of confidence as distance-to-criterion. One important consequence of quantifying confidence as the distance between the decision variable and a fluctuating criterion, is that confidence will naturally become autocorrelated across trials, a phenomenon referred to as the confidence leak (Rahnev et al., 2015). Whereas in the literature these two observations (variability and autocorrelation) are typically explained via different mechanisms, fluctuations in decision criterion could provide a parsimonious account for both explaining variability and autocorrelation in confidence ratings. With this work new insights are provided into the computational underpinnings of decision confidence.

References

Ashwood, Z. C., Roy, N. A., Stone, I. R., Urai, A. E., Churchland, A. K., Pouget, A., & Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), Article 2. https://doi.org/10.1038/s41593-021-01007-z

Cowley, B. R., Snyder, A. C., Acar, K., Williamson, R. C., Yu, B. M., & Smith, M. A. (2020). Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron*, *108*(3), 551-567.e8. https://doi.org/10.1016/j.neuron.2020.07.021

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Gupta, D., & Brody, C. D. (2022). *Limitations of a proposed correction for slow drifts in decision criterion* (No. arXiv:2205.10912). arXiv. https://doi.org/10.48550/arXiv.2205.10912

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2016). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. *Cerebral Cortex*, *26*(1), 118–130. https://doi.org/10.1093/cercor/bhu181

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200

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

Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D., & Pillow, J. W. (2021). Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron*, *109*(4), 597-610.e6. https://doi.org/10.1016/j.neuron.2020.12.004

Shekhar, M., & Rahnev, D. (2021). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review*, *128*(1), 45–70. https://doi.org/10.1037/rev0000249

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. https://doi.org/10.1038/ncomms14637

Vloeberghs, R., Urai, A. E., Desender, K., & Linderman, S. W. (2024). *A Bayesian Hierarchical Model of Trial-To-Trial Fluctuations in Decision Criterion* (p. 2024.07.30.605869). bioRxiv. https://doi.org/10.1101/2024.07.30.605869

Xue, K., Shekhar, M., & Rahnev, D. (2024). Challenging the Bayesian confidence hypothesis in perceptual decision-making. *Proceedings of the National Academy of Sciences*, *121*(48), e2410487121. https://doi.org/10.1073/pnas.2410487121