Exploring the Neural Basis of Stimulus-Driven Overconfidence

Ema Zezelic (ema.zezelic@uni-tuebingen.de)
Hertie-Institute for Clinical Brain Research, Centre for Integrative Neuroscience & MEG Center,
University of Tuebingen, Tuebingen, Germany

Florian Sandhaeger
Hertie-Institute for Clinical Brain Research, Centre for Integrative Neuroscience & MEG Center,
University of Tuebingen, Tuebingen, Germany

Katrina R Quinn
Hertie-Institute for Clinical Brain Research, Centre for Integrative Neuroscience & MEG Center,
University of Tuebingen, Tuebingen, Germany

Markus Siegel
Hertie-Institute for Clinical Brain Research, Centre for Integrative Neuroscience & MEG Center,
University of Tuebingen, Tuebingen, Germany

Abstract

In decision-making, confidence is the ability to judge how likely our choices are to be correct or incorrect. Whilst this internal sense of accuracy can be a useful tool with which to adapt our behaviour, it is not always reliable. In fact, certain visual manipulations can lead observers to feel more confident even when the likelihood of correct choices remains the same. Here we used such a manipulation combined with magnetoencephalography, to investigate the neural basis of the "positive-evidence bias". Participants performed a visual decision-making task with confidence judgements, in which we induced overconfidence in one of two conditions, while keeping accuracy between the conditions the same. We found evidence that the observed behavioral overconfidence could be explained by increased separation and variance of neural evidence representations and is not necessarily due to a higher-level cognitive bias.

Keywords: confidence; perceptual decision-making; magnetoencephalography; decoding

Introduction

In decision-making, confidence is the ability to judge how likely our choices are to be correct or incorrect. An important aspect of confidence is that it is correlated with accuracy. If we are highly confident in our decisions, those decisions are more likely to be accurate. While this internal sense of accuracy can be a useful tool with which to guide our future behaviour, it is not always reliable. Certain visual manipulations can make us feel more confident than we should be. One example of such a manipulation is known as the "positive-evidence bias". Specifically, by varying the magnitude and ratio between evidence that supports the choice ('positive evidence') and the evidence for the opposing choice ('negative evidence'), humans and non-human primates can perform perceptual decision-making tasks with similar levels of accuracy, but different levels of confidence (Odegaard et al., 2018; Samaha & Denison, 2022). High magnitudes of positive and negative evidence leave participants more confident than if the magnitude were smaller while keeping the ratio between positive and negative evidence similar.

When it comes to behavioural studies investigating this manipulation, there are two common explanations. Overconfidence arises either due to: (1) a higher-level cognitive bias e.g. observers only focus on confirmatory evidence when constructing confidence judgments, or (2) an increase in separation and variance of evidence distributions, without a change in the confidence criterion. In the latter case higher confidence is a result of the choice distribution being shifted towards more extreme values (Shekhar & Rahnev, 2024).

To investigate the whole-brain neural dynamics and uncover the signals related to choice and confidence, we used magnetoencephalography (MEG) paired with multivariate decoding and signal detection theory (SDT) modelling. We measured means and variances of choice distributions in neural and behavioural data and found evidence that the increased separation and variance of these distributions can explain behavioural overconfidence. These results contribute to a mechanistic understanding of the neural basis of confidence in human decision-making (Rahnev et al., 2022).

Methods

We recorded MEG data from 37 subjects using a 271-channel whole-head MEG system. Participants performed a visual decision-making task with binary confidence judgements in which they had to judge whether orientation stimuli were tilted 45° or -45°. The task consisted of 960 trials from two randomly interleaved conditions: (1) High Positive Evidence (HPE), with strong evidence for both orientations and (2) Low Positive Evidence (LPE), with low evidence for both orientations. All stimuli had evidence for both orientations, but there was always more positive evidence. We fixed the parameters determining the strength of evidence to ensure 66% accuracy in both conditions. Participants had to simultaneously choose the orientation and the confidence level by pressing one of four buttons. A target screen, showing which button corresponded to which orientation and confidence level, appeared after the stimulus presentation.

Using multivariate decoding (cvManova, Allefeld & Haynes, 2014; Sandhaeger et al., 2023), we investigated neural information related to choice, confidence and stimulus orientation.

Results

While the accuracy between the conditions was not significantly different, participants were significantly more confident in the HPE condition ($p_{confidence} = 2.5 \times 10^{-6}$ for one-tailed t-test; see Fig. 1).

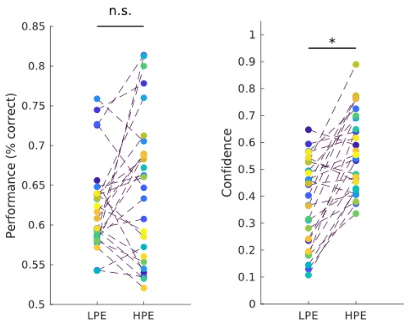

Figure 1: Accuracy and confidence plotted separately for LPE and HPE conditions. Each pair of coloured dots represents a single participant.

Using multivariate decoding, we found significant neural information related to choice, confidence, and stimulus orientation using all trials (see Fig. 2). We found that both, the mean and variance of the choice information distribution were significantly higher in the HPE than in the LPE condition ($p_{means} = 0.0097$, $p_{variances} = 0.0222$ for one-tailed t-test; see Fig. 3).

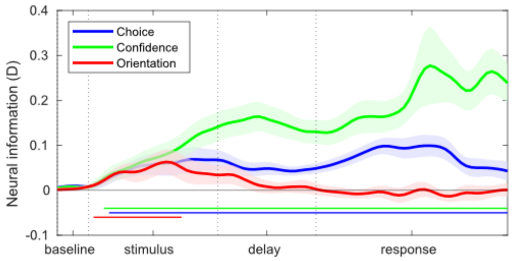

Figure 2: Decoded neural information related to choice, confidence, and stimulus orientation using trials of both conditions for training and testing. Horizontal bars indicate clusters of significant temporal information ($p < 0.05$).

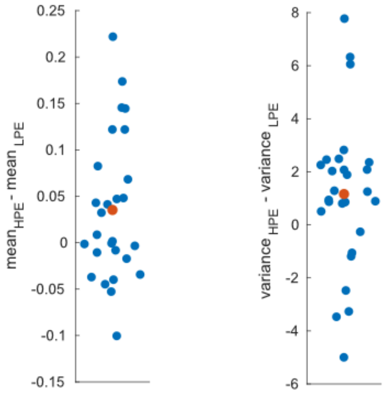

Figure 3: Difference of mean and variance of choice information distributions between HPE and LPE conditions. Each blue dot represents one participant, while the orange dots represent the mean value.

Conclusions

We found that increasing both positive and negative evidence in orientation stimuli induces overconfidence for similar levels of accuracy. We show that both the mean and variance of choice information are higher in the HPE than in the LPE condition across participants. This provides evidence that “positive-evidence bias” can be explained by low-level changes in the representation of sensory evidence rather than by a higher-level cognitive bias.

Acknowledgments

This research is supported by the European Research Council (ERC CoG 864491).

References

Allefeld, C., & Haynes, J. D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by crossvalidated MANOVA. Neuroimage, 89, 345-357. https://doi.org/10.1016/j.neuroimage.2013.11.043

Odegaard, B., Grimaldi, P., Cho, S. H., Peters, M. a. K., Lau, H., & Basso, M. A. (2018). Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proceedings of the National Academy of Sciences*, *115*(7). https://doi.org/10.1073/pnas.1711628115

Rahnev, D., Balsdon, T., Charles, L., De Gardelle, V., Denison, R., Desender, K., Faivre, N., Filevich, E., Fleming, S. M., Jehee, J., Lau, H., Lee, A. L. F., Locke, S. M., Mamassian, P., Odegaard, B., Peters, M., Reyes, G., Rouault, M., Sackur, J., . . . Zylberberg, A. (2022). Consensus goals in the field of visual metacognition. *Perspectives on Psychological Science*, *17*(6), 1746–1765. https://doi.org/10.1177/17456916221075615

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

Sandhaeger, F., Omejc, N., Pape, A. A., & Siegel, M. (2023). Abstract perceptual choice signals during action-linked decisions in the human brain. Plos Biology, 21(10), e3002324. https://doi.org/10.1371/journal.pbio.3002324

Shekhar, M., & Rahnev, D. (2024). Human-like dissociations between confidence and accuracy in convolutional neural networks. *PLoS Computational Biology*, *20*(11), e1012578. https://doi.org/10.1371/journal.pcbi.1012578