Integrating explicit reliability for optimal choices: effect of trustworthiness on decisions and metadecisions

Keiji Ota (k.ota@qmul.ac.uk)
Centre for Brain and Behaviour, Queen Mary University of London, Mile End Road, London, UK

Anthony Ciston (aciston819@gmail.com)
Max Planck Institute for Human Cognitive and Brain Sciences, Postfach 500355, Leipzig, Germany

Patrick Haggard (p.haggard@ucl.ac.uk)
Institute of Cognitive Neuroscience, University College London, 17-19 Queen Square, London, UK

Thibault Gajdos Preuss (prénom.nom@univ-amu.fr)
Aix Marseille Univ, CNRS, CRPN, Marseille, France

Lucie Charles (l.charles@qmul.ac.uk)
Centre for Brain and Behaviour, Queen Mary University of London, Mile End Road, London, UK

Abstract

A key challenge in today's fast-paced digital world is to integrate information from diverse sources with different reliability. Beyond estimating the reliability of information based on prior knowledge, it is also fundamental to understand whether people can use explicit information about the reliability of the source. In particular, a question that remains underexplored is how people use probabilistic information about the likelihood of a source to give correct information. Here, we investigated how such explicit probabilistic estimates of reliability are encoded and integrated into decision processes. To do so, we developed a novel paradigm that required participants to combine evidence from sources with different explicit levels of reliability to estimate among two responses which one was more likely to be correct. Additionally, participants had to rate after each choice the extent to which they felt they were influenced by a given source of information. Through computational modelling, we found that participants misrepresented the reliability of sources, distorting the probability a source to give correct information. As a results, they gave too much weight to unreliable source and too little weight to sources that were reliably wrong sources. However, we found that subjective report of influence correctly predicted the effective influence a source had on the decision. These findings suggest that participants were at least partially aware of what bias their choices.

Keywords: Evidence accumulation; Bayesian decision theory; probability distortion; introspection of choice; metacognition

Introduction

In the era of misinformation and fake news, providing indices of information trustworthiness has appeared as an fundamental strategy to minimise their impact. Findings on the effectiveness of this methods to prevent the spread of false beliefs remains debated however. More fundamentally, how people use explicit information about source reliability when making decisions remains poorly understood. Research in economics has started to reveal the biases when reasoning with probabilities (Kahneman & Tversky, 1984). It remains unclear however whether such limitations also apply when probabilities represent the reliability of a source to give correct information.

Additionally, little is known about whether people can monitor how much weight they give to sources with different reliability when making a decision. People often misattribute the true reasons for their decisions (Epstein & Robertson, 2015), falling victim to "bias blindness" (Pronin, 2007). The question of whether people can introspect being biased by unreliable information however remains underexplored.

Methods

Task. Participants performed a decision-making task where they viewed successive samples (red and blue squares) supporting one of two responses (red or blue). Each sample was associated with a percentage corresponding to the reliability of the source providing the opinion (Fig. 1A) i.e. the probability that the source provides correct information about the correct colour. The sources reliabilities were chosen from one of the three levels: one unreliable source labelled as 50% and two other sources that could be reliably right (55% or 65%) or reliably wrong (45% or 35%), varied across 5 experiments. After deciding which response was more likely to be correct, participants reported how they felt their choice was influenced by a given source (Fig. 1A).

Computational modelling. We used a hierarchical model to fit participants' choices, probing the degree of distortion in the encoding of the reliability indices (Tversky & Khaneman, 1992) and the degree of recency in the evidence accumulation process. In the model, when presented with a sample of evidence (coloured square) of a given reliability, the evidence is first transformed into a log-odds value, where the reliability of evidence goes through a non-linear probability distortion function (Zhang et al., 2020; Gonzalez & Wu, 1999). Additionally, a sequential weight is applied to each log-odds according to its position in the evidence sequence. After summing these weighted log-odds for each colour, the model the colour with the highest log-odds value is chosen as the response.

Results

Choice Behaviour. Participants performed the task with good accuracy but the presence of reliably wrong information decreased their accuracy. Quantifying, for each source, how much the colour it favours

influenced the participant's choices, we confirmed that influence was proportional to source reliability – the evidence from a higher reliability source (65%) led to the steepest increase in choice likelihood (Fig. 1B). However, the participant's choices were also increasingly biased by the evidence from the unreliable source (50%), deviating from optimal decision-making strategy.

Computational account. The computational model suggested that participants treated the reliably right sources as more reliable than they actually were, for instance acting as if they believed that the source of 55% reliability was 73% reliable (Fig. 1C). The source of 50% reliability was also treated as having above-chance reliability (60%). The model also suggests that participants discounted the first few pieces of information compared to the last information (Fig. 1C). Finally, increasing the range of reliabilities presented actually led to a less distorted encoding of reliabilities.

Influence report. We confirmed that feeling of influence by a given source increased when a greater amount of evidence supported the choice (Fig. 1D). Importantly, this was true even for the unreliable sources (50%), suggesting participants were aware of being biased by those sources. Confirming this, we found a correlation between the distorted reliability obtained from the choice model and the subjective sense of influence (Fig. 1E), suggesting that participants who assigned a stronger weight to a source also reported a stronger sense of feeling influenced by that source.

Conclusion

We demonstrate that participants can use explicit reliability indices, but do so in a distorted way such that they overweight unreliable sources and underweighted old information. Additionally, participants have overall good introspective accuracy in reporting how strongly their choice is influenced by a given source. In particular, participants seem aware that they are biased by unreliable information. Taken together, these findings suggest that, although people are suboptimal in using explicit source reliability, they possess some metacognitive knowledge of how their decision are made.

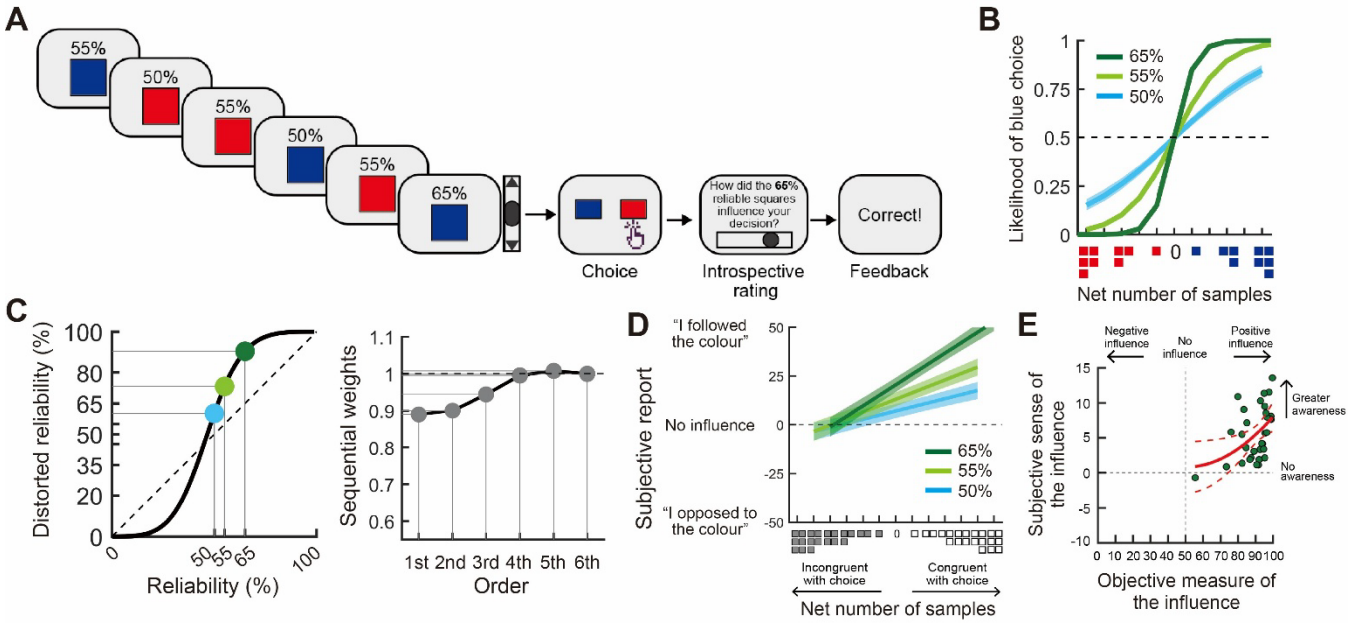

Figure 1. **(A)** Participants scrolled down a webpage to reveal a sequence of six sources, each predicting a colour. **(B)** The proportion of blue choices increases as the net number of blue samples displayed by each of three levels of reliability. **(C)** Distorted reliability percentage and sequential weights estimated from the computational model. **(D)** The subjective feeling of following the colour increases as the net number of congruent samples with choice. **(E)** The larger the influence of a given reliability source on choice, the larger the awareness of the influence.

Acknowledgements

This work was supported by an ESRC grant ES/V00378X/1 awarded to LC and PH.

References

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proc Natl Acad Sci U S A, 112*, E4512-4521.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cogn Psychol, 38*, 129-166.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341-350.

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends Cogn Sci, 11*, 37-43.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty *Journal of Risk and uncertainty, 5*, 297-323.

Zhang, H., Ren, X., & Maloney, L. T. (2020). The bounded rationality of probability distortion. *Proc Natl Acad Sci U S A, 117*, 22024-22034.