Experimental Assessment of Choice-Congruent Bias in Human Reinforcement Learning

Alexandre Lietard (alexandre.lietard@kuleuven.be)
Brain and Cognition, KU Leuven, Tiensestraat 102
Leuven, 3000, Belgium
Kobe Desender (kobe.desender@kuleuven.be)
Brain and Cognition, KU Leuven, Tiensestraat 102
Leuven, 3000, Belgium

Abstract

Confidence constitutes a fundamental signal which aids adaptative processes in both learning and decision-making. Yet, it often deviates from optimality. In the domain of perceptual decision-making, one such deviation, known as the choice congruent bias, reflects the tendency to overweight evidence that supports the chosen option. As a result, confidence tends to increase with the amount of overall evidence, even when accuracy remains unchanged. Here, we examined whether the same biased computation occurs in value-based decisions, by testing whether greater overall evidence leads to higher confidence in a reinforcement learning task.

The results demonstrate that increasing the average reward successfully elevates confidence levels. Our computational modelling indicates, however, that this effect does not necessarily reflect a biased confidence computation. Therefore, while manipulating average reward may be a useful method for dissociating confidence from accuracy, it cannot serve as a direct test of the choice-congruent bias.

Keywords: Metacognition; Reinforcement learning; Choice-congruent bias; Confidence

Introduction

Confidence is a core metacognitive process that reflects the subjective probability of being correct and guides decisions and actions. Although adaptive, it often deviates from optimality due to noise or systematic biases (Shekar & Rahnev, 2024). One such bias, the choice-congruent bias (also named positive evidence bias), involves overweighting evidence favoring the chosen option (Zylberberg, Barttfeld, Sigman, 2012). Consequently, conditions with more total evidence can elicit higher confidence even when accuracy remains constant.

Extending this phenomenon beyond visual perception, Salem-Garcia, Palminteri and Lebreton (2023) suggested that confidence is also biased by the estimated value of the chosen option in value-based tasks. However, to date, no study has corroborated the existence of this bias from an experimental point of view in a reinforcement learning task. An experimental dissociation between confidence and accuracy could be useful for isolating the influence of confidence on value learning (see also Ting et al., 2020).

In the present study, we developed a novel experimental manipulation to test the choice-congruent bias. We employed a two-armed bandit task in which the amount of evidence was manipulated by subtly increasing the average reward of both options, increasing the overall level of evidence (Rollwage et al., 2020). Thus, we hypothesized that higher average rewards would lead to increased confidence, without affecting accuracy.

Methods

Fifty participants completed a gamified 2-armed bandit task, aiming to collect as much fruit as possible. Their goal was to learn which of two fields (left or right) had the higher average yield for a given fruit. On each trial, participants chose an option, rated their confidence in their choice (1 = "guessing" to 4 = "very confident"), and received a reward (i.e. the number of fruits obtained) (Fig. 1.A). Rewards were drawn from a normal distribution, one option having a mean above 50 and the other below 50.

In the low (overall) evidence condition, the lower-average option had an average reward of 40, and the higher-average option of 60. In the high (overall) evidence condition, both averages were increased by a small amount (46 for the lower option and 66 for the higher option), maintaining the same difference between the two options (Fig. 1.B). Each participant completed both conditions, performing 20 blocks of 15 trials per condition. Predictions were tested using behavioral measures as well as computational modeling.

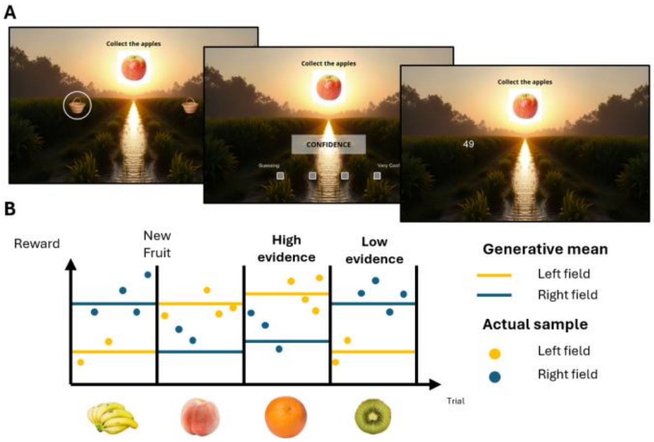

Fig. 1. Experimental Design: A. After selecting one of the two fields, participants rated their confidence and received a reward. **B**. Rewards were drawn from a normal distribution in high evidence (mean = 46–66) or low evidence (mean = 40–60) contexts.

To model our data, we assumed that participants used a Rescorla-Wagner (1972) learning rule to update the belief Q about the average of the option k at time t based on the reward R obtained:

$$Q_{t+1}(k) = Q_t(k) + \alpha(R_t - Q_t(k)) \quad (1)$$

with α corresponding to the learning rate. They then used a soft-max rule to decide which option to choose based on the average estimated associated with each option:

$$p_t(k) = \frac{exp(Q_t(k) \cdot \beta^{-1})}{\sum_{i=1}^{2} exp(Q_t(i) \cdot \beta^{-1})} \quad (2)$$

with β reflecting the decision noise. We modelled confidence as being either the value of the chosen option $Q(c)$, reflecting the choice-congruent bias or as the distance between the two options $Q(c) - Q(u)$.

Results

We first computed the average difference in confidence and accuracy between the two contexts. The analyses revealed a clear dissociation, with a reliable difference in confidence (45/48 participants, $t(47) = 8.05$, $p < .001$) between contexts, while no significant difference in accuracy was found ($t(47) = 1.74$, $p = .15$). Nevertheless, mixed models indicated that the difference between conditions interacted with the trial number for both confidence ($z = 6.39$, $p < .001$) and accuracy ($z = 5.05$, $p < .001$), indicating only a partial dissociation. To gain a more precise temporal analysis of the effect, we conducted a cluster-based permutation analysis on the context differences. These analyses revealed that the effect for accuracy emerged later, toward the end of the block, compared to confidence (Fig 2.)

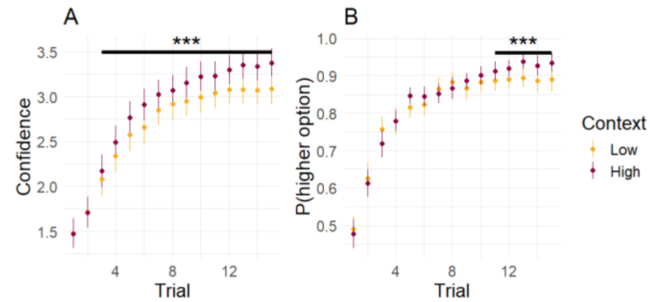

Fig. 2. Results: A. Confidence showed a difference between the two contexts, and this effect started at the beginning of the block. **B.** In contrast, the difference in choosing the higher option (i.e. accuracy) emerged only at the end of the block. *Note:* Error bars correspond to 95% confidence intervals. *** indicate the significance of the cluster (p<.001).

We next fitted our models to human data, testing variants of the original model that differed in how it updated the chosen/unchosen values and handled positive vs. negative prediction errors. The best-fitting model updated only the chosen value (ΔBIC = 21, ΔAIC = 15 compared to the second best). Using this model, we extracted participants' Q-values (expected reward for both options) and fitted their confidence judgments using three thresholds (corresponding to the four possible responses of the Likert scale). Surprisingly, the analysis showed that the distance rule, $Q(c) - Q(u)$, better explained the confidence of participants compared to the choice congruent bias rule $Q(c)$, $t(47) = 3.87$, $p < .001$.

In sum, our results show that this novel manipulation of evidence may help to dissociate confidence from accuracy in value-based decisions and can be used as a tool to isolate the effects of confidence. However, the computational modelling suggest that this dissociation between confidence and accuracy does not per se reflect a choice-congruent bias in the computation of confidence.

Acknowledgments

Authors are supported by Marie Skłodowska-Curie Actions through the COnfident DEcisions (CODE) doctoral network.

References

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (Vol. 2, pp. 64–99). Appleton-century-crofts.

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications*, *11*(1), 2634. https://doi.org/10.1038/s41467-020-16278-6

Salem-Garcia, N., Palminteri, S., & Lebreton, M. (2023). Linking confidence biases to reinforcement-learning processes: Psychological Review. *Psychological Review*, *130*(4), 1017–1043. https://doi.org/10.1037/rev0000424

Shekhar, M., & Rahnev, D. (2024). How do humans give confidence? A comprehensive comparison of process models of perceptual metacognition. *Journal of Experimental Psychology: General*, *153*(3), 656–688. https://doi.org/10.1037/xge0001524

Ting, C.-C., Palminteri, S., Engelmann, J. B., & Lebreton, M. (2020). Robust valence-induced biases on motor response and confidence in human reinforcement learning. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(6), 1184–1199. https://doi.org/10.3758/s13415-020-00826-0

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