Monetary incentives bias confidence judgements less in the absence of choice

Nahuel Salem-Garcia (nahuel.salem.garcia@ens.psl.eu)
Laboratoire de Neurosciences Cognitive et Computationnelles,
Institut National de la Santé et de la Recherche Médicale,
Département d'Etudes Cognitives, Ecole Normale Superieure - PSL University,
Paris 75005, France
Swiss Center for Affective Science, Faculty of Psychology and Educational Sciences, University of Geneva,
Geneva 1202, Switzerland

Sebastien Massoni (sebastien.massoni@gmail.com)*
Bureau for Economic Theory and Applications,
Université de Lorraine,
Université de Strasbourg,
CNRS,
Nancy 54000, France

Maël Lebreton (mael.lebreton@gmail.com)*
Paris-Jourdan Sciences Économiques,
Paris School of Economics,
Paris 75014, France
Swiss Center for Affective Science, Faculty of Psychology and Educational Sciences, University of Geneva
Geneva 1202, Switzerland

Valentin Wyart (valentin.wyart@gmail.com)*
Laboratoire de Neurosciences Cognitive et Computationnelles,
Institut National de la Santé et de la Recherche Médicale,
Département d'Etudes Cognitives, Ecole Normale Superieure - PSL University,
Paris 75005, France

*: equal senior authorship

Abstract

Humans have a sense of confidence that tracks the probability of having made a correct decision based on uncertain evidence. This sense, though relatively accurate, presents numerous biases, especially being affected by the subjective value of expected outcomes, and neglecting evidence for unchosen options. Here, we ask how these affective and choice-related biases may interact, by manipulating monetary incentives and choice agency in a visual categorization task with confidence judgements. We then compare the results with predictions from computational models. We show that the incentive effect on confidence is not only present when participants judge their own choices, but also when observing choices imposed by the computer. However, in the latter case, the effect is attenuated. We conclude that the incentive effect emerges from attention, not action, exaggerating subjective evidence in favor of an option, and that choice ownership intensifies this effect.

Keywords: metacognition; confidence; motivated cognition; computational modeling

Introduction

Humans have a sense of confidence that tracks the probability of having made a correct choice (Aitchison et al., 2015; Sanders et al., 2016), which they can use to modulate behavior (Meyniel et al., 2015; Rollwage et al., 2020). However, confidence is subject to an incentive effect: prospective gains (losses) increase (decrease) confidence (Lebreton et al., 2018; Salem-Garcia et al., 2023). It is unknown whether this incentive effect is specific to self-monitoring processes (about one's own choices) or, on the contrary, it is a general feature of probabilistic computations (about external states). In fact, the act of choosing may be the root of biased beliefs (Sharot et al., 2010; Michel and Peters, 2021; Chambon et al., 2020), and confidence judgements in particular have been proposed to be affected by an interaction of subjective value of outcomes and choice (Dayan and Daw, 2008; Salem-Garcia et al., 2023).

Here, we compare the incentive effect on free versus observed decisions in a human behavioral experiment, and conclude that the incentive effect does not require a self-made choice, though it is exacerbated by it.

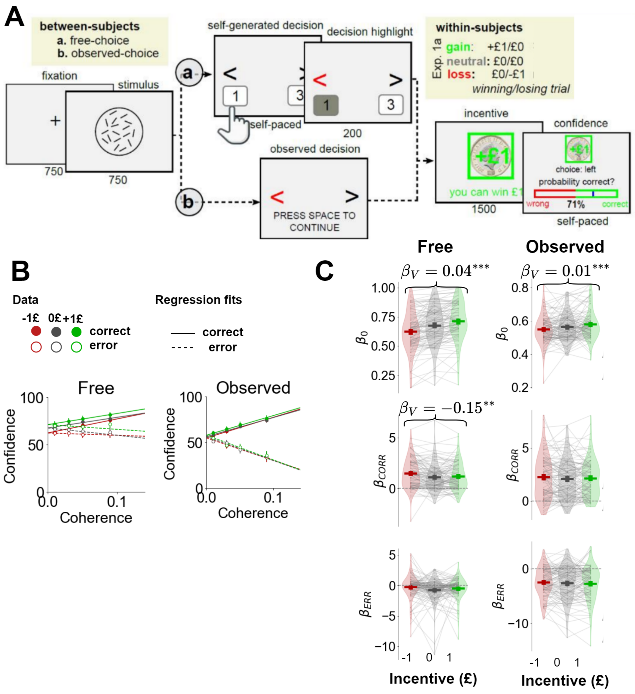

Figure 1: Experiment details and behavioral results. A. Task design. B. Confidence ratings over stimulus coherence, correctness, and incentive. C. Regression coefficients for formula $confidence \sim coherence : correct + coherence : error + 1$ (β_0:intercept, $beta_{CORR}$: slope over coherence for correct choices, β_{ERR}: slope over coherence for error choices). Incentive effects on the coefficients are reported when significant (β_V).

Methods and Results

Experiment In a perceptual categorization task, 200 participants repeatedly saw clouds of moving dots on a screen, made or observed decisions about the general motion direction, and rated their confidence in the decision being correct (Figure 1 A. Within-participants, we varied the potential outcomes: depending on the trial, participant could earn £ 1 (gain condition), nothing (neutral condition) or avoid losing £ 1 (loss condition) for an accurate confidence judgment, (see Lebreton et al., 2018, for a similar design). This was shown after the choice, and before the probability rating. Between-participants, we varied the agency of the decisions: 100 participants made the decisions (Free condition) and 100 participants observed decisions made by the computer (Observed condition).The observed decisions were taken based on the average performance of participants in a pilot study. In both conditions, participants were informed that the average expected accuracy of decisions was 75%.

Models We consider an actor-and-observer Signal Detection Theory model, which takes two perceptual

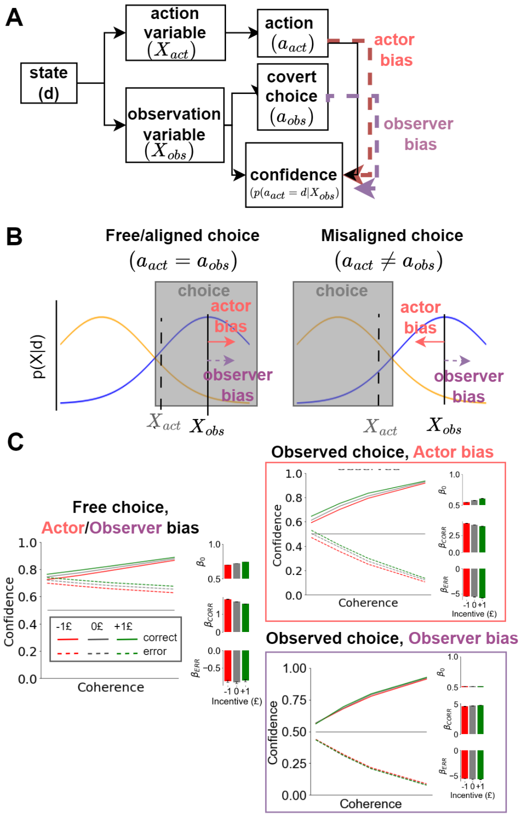

Figure 2: A.Model of choice and confidence. B.Effects of actor-centric and observer-centric biases on evidence used for confidence (v_n values of 0.5, 1, and 1.5 for -1,0,and +1 £ incentives respectively). C. Predicted incentive effects of actor- and observer-centric biases in the different conditions on confidence pattern across stimulus coherence and choice correctness. Line plots show simulation behavior, and bar plots the coefficients (β_0:intercept, $beta_{CORR}$: slope over coherence for correct choices, β_{ERR}: slope over coherence for error choices)

samples of noisy evidence about the direction of motion that we call X_{act} and X_{obs}. In free choices these are identical, and in observed choices, they are independent. Actor and observer make choices a_{act} and a_{obs} based on the sign of the evidence, though only the actor's choice is expressed.

Confidence is computed based on a combination of the observer evidence and a bias term:

$$X_{conf} = X_{obs} + \upsilon_n a_* \quad (1)$$

Where υ_n is a parameter quantifying the bias for incentive n, and a_* is either a_{act} or a_{obs} for an actor-centric and observer-centric model respectively. Confidence is computed as the Bayesian probability of the actor's choice being correct given X_{conf} as in Fleming and Daw (2017). [1]

Results In our experimental data, we regressed confidence on coherence separately for each incentive level and participant. We found incentive value affects confidence independently of coherence (affecting the regression intercept) in both Free and Observed choices. This effect is smaller in the observed condition (Figure 1 B and C). In model simulations, we found that only an actor-centric bias predicts an intercept effect in both conditions (Figure 2 C). We also wondered whether the smaller incentive effect in observed choices could be due to structural features of confidence in free vs observed decisions (e.g. higher variance in observed). However, the model predicts a similar magnitude of effect, suggesting the difference may be due to an agency-driven cognitive bias.

Conclusions

We showed that incentive bias in confidence persists in the absence of free choice, validating an actor-centric model of bias (incentives affect evidence towards actual choice, regardless of agency) and falsifying an observer-centric bias (incentives affect evidence towards a covert choice).We also show that this bias is attenuated in observed choices. We suggest the interpretation that the incentive bias emerges from directed attention to a specific option, but that self-made choice engages this attentional effect more effectively than externally provided cues.

Acknowledgements

The study was supported by research funds allocated to SM (programme FUTURE LEADER of Lorraine Université d'Excellence within the French National Research Agency Investissements d'avenir ANR-15-IDEX-04-LUE). ML and NSG acknowledge the support of an FNSF Ambizione grant (PZ00P3_174127) and of an ERC Starting Grant (INFORL-948671) awarded to ML. NSG is currently supported by a Postdoc.Mobility Grant from FNSF (P500PS_214314).

References

Aitchison, L., Bang, D., Bahrami, B., and Latham, P. E. (2015). Doubly Bayesian Analysis of Confidence in Perceptual Decision-Making. *PLoS Computational Biology*, 11(10):e1004519.

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., and Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, 4(10):1067–1079. Number: 10 Publisher: Nature Publishing Group.

Dayan, P. and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4):429–453.

Fleming, S. M. and Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological Review*, 124(1):91–114. arXiv: 1106.2252v2 ISBN: 6505615628.

Lebreton, M., Langdon, S., Slieker, M. J., Nooitgedacht, J. S., Goudriaan, A. E., Denys, D., van Holst, R. J., and Luigjes, J. (2018). Two sides of the same coin: Monetary incentives concurrently improve and bias confidence judgments. *Science Advances*, 4(5).

Meyniel, F., Schlunegger, D., and Dehaene, S. (2015). The Sense of Confidence during Probabilistic Learning: A Normative Account. *PLOS Computational Biology*, 11(6):e1004305.

Michel, M. and Peters, M. A. K. (2021). Confirmation bias without rhyme or reason. *Synthese*, 199(1):2757–2772.

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., and Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications*, 11(1):2634. Number: 1 Publisher: Nature Publishing Group.

Salem-Garcia, N., Palminteri, S., and Lebreton, M. (2023). Linking confidence biases to reinforcement-learning processes. *Psychological Review*, 130(4):1017–1043.

Sanders, J., Hangya, B., and Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. *Neuron*, 90(3):499–506.

Sharot, T., Velasquez, C. M., and Dolan, R. J. (2010). Do Decisions Shape Preference?: Evidence From Blind Choice. *Psychological Science*, 21(9):1231–1235. Publisher: SAGE Publications Inc.

[1] For simplicity, we assume the estimated noise is the true observer's noise ($\hat{\sigma} = \sigma_{obs}$), and that the estimated mean is the expected value of X for an accuracy of 75% ($\hat{\mu} = \Phi^{-1}(0.75)\hat{\sigma}$, where Φ is the standard cumulative density function of the normal distribution).