Meta-cognitive Efficiency in Learned Value-based Choice

Sara Ershadmanesh (sara.ershadmanesh@tuebingen.mpg.de)
Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Ali Gholamzadeh (ali.gholamzadeh@tuebingen.mpg.de)
Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
University of Tübingen, Tübingen, Germany

Kobe Desender (kobe.desender@kuleuven.be)
Brain and Cognition, KU Leuven, Belgium
Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
Department of Experimental Psychology, Ghent University, Ghent, Belgium

Peter Dayan (dayan@tue.mpg.de)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
University of Tübingen, Tübingen, Germany

Abstract

Meta-cognition, our ability to assess the quality of our own decisions, is important for regulating choices and has been extensively studied through assessing and modeling reports of confidence. However, these studies focus on immediate choices rather than sequential ones. The latter pose particular problems for meta-cognitive efficiency assessments, because the underlying difficulty of decision-making changes as the problems evolve. Here, we focus on sensitivity and bias of meta-cognitive judgments in learning/decision-making tasks in which outcome values must be learned across trials. We repurposed the central idea underlying the M-ratio, a popular meta-cognitive assessment measure in perceptual decision-making. We built a Forward model of confidence, characterizing the subjects' choices and generating 'first order' confidence from the modelled probability of being correct; and a Backward model of confidence, which generates choices whose first-order confidence best matches the subjects' confidence reports. The performance of Forward and Backward models play the roles of d' and meta-d' in our measure of meta-cognitive efficiency, MetaRL-Ratio. We found that the performance of the Backward model was consistent with previous measures of meta-cognitive sensitivity and the MetaRL-Ratio differentiated simulated low and high meta-cognitive competence. This study suggests that MetaRL-Ratio is a promising tool for assessing meta-cognitive efficiency in the value-based learning/decision-making.

Keywords: Meta-cognitive efficiency; Forward/Backward model of confidence; Value-based learning/decision-making

Materials and Methods

Cognitive task. Participants performed an online two-armed restless bandit task (Sutton et al., 1998). They were told that the goal of the task was to gain as much money as possible on virtual slot machines. On each trial, participants had to choose between two slot machines, presented on their computer screen, one of which had a higher average reward per choice. They were told that every 18-22 trials a switch would take place and the other slot machine would now give the higher reward. After each choice the participants had to indicate how confident they were about their decision on a continuous scale running from 'this was a guess' to 'very certain'. After they rated their confidence, a numerical reward based on their choice appeared on the screen and the next trial started. In one condition of experiment, the rewards following a worse or better choice were drawn from normal distributions $\mathcal{N}(40,8)$ or $\mathcal{N}(60,8)$ respectively (with a single set of rewards for all participants, but with randomly shuffled orders). The task included 20 blocks of 18-22 trials, for a total of 400 trials.

Computational modelling

Forward model of confidence. We use maximum likelihood estimation to fit each subject's empirical choices. The best account came from an extended Q-learning-based reinforcement learning model (Figure 1, blue) with softmax exploration and three learning rates, one each for positive and negative prediction errors for the chosen option, and one more (typically negative) for the unchosen option, which multiplied the full prediction error of the chosen option. We evaluated how much reward the best-fitting forward model would gather on average when run autonomously on the task using the same (potential) reward sequences as the participants in the experiment ($\bar{R}_s^f$ for subject s). To model confidence, we scaled the output of the softmax using parameters L_c and H_c, which represent the confidence biases for each subject, and were fit by minimizing the Euclidean distance between the scaled and empirical confidence (using scipy.optimize.minimize).

Backward model of confidence. For the case of perceptual decision-making, the meta-cognitive sensitivity measure meta$-d'$ comes from treating the empirical reports of confidence as the result of a probabilistic choice process (as in a first-order decision-making model), and quantifying the effective perceptual sensitivity of that model. This can be seen as going backwards from confidence to choice. We therefore defined a Backward model (figure 1, red) in which we characterize subjects' choices as coming from the same RL process as the Forward model, but with decision-making parameters fit to make the (similarly scaled) choice probabilities match the empirical confidence judgments as best as possible (rather to match the empirical choices). We then evaluated how much reward the best-fitting Backward model would gather when run autonomously on the task ($\bar{R}_s^b$ for subject s), as for the Forward model.

Measure of meta-cognitive efficiency. Inspired by the M-ratio for perceptual-decision-making, we defined the MetaRL-Ratio $= \bar{R}_s^b / \bar{R}_s^f$ as a measure of meta-cognitive efficiency. A salient feature of this ratio is that it accounts correctly for the varying difficulty of the underlying RL problem, generated here by the unsignalled changes in the qualities of the bandits. Such changes make difficulties in the perceptual case (Rahnev & Fleming, 2019). Standard meta-cognitive sensitiv-

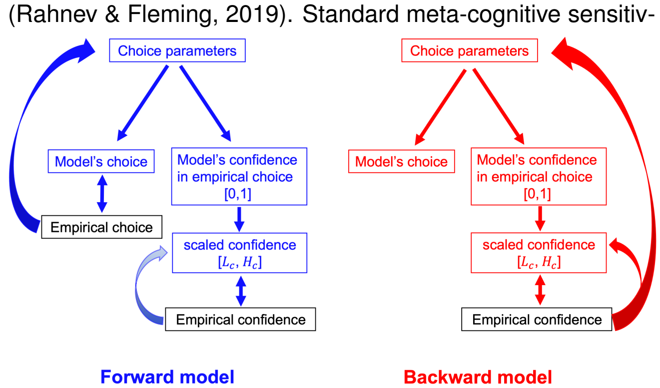

Figure 1. In the Forward model (blue), we used the choices made by subjects to determine the choice parameters of our model. The scaled confidence of model in empirical choices via confidence bounds parameters, L_c and H_c, was fitted to the empirical confidence reports. In the Backward model (red), we determined choice parameters by matching the optimally-rescaled first-order confidence of the choice model to the empirical confidence reports.

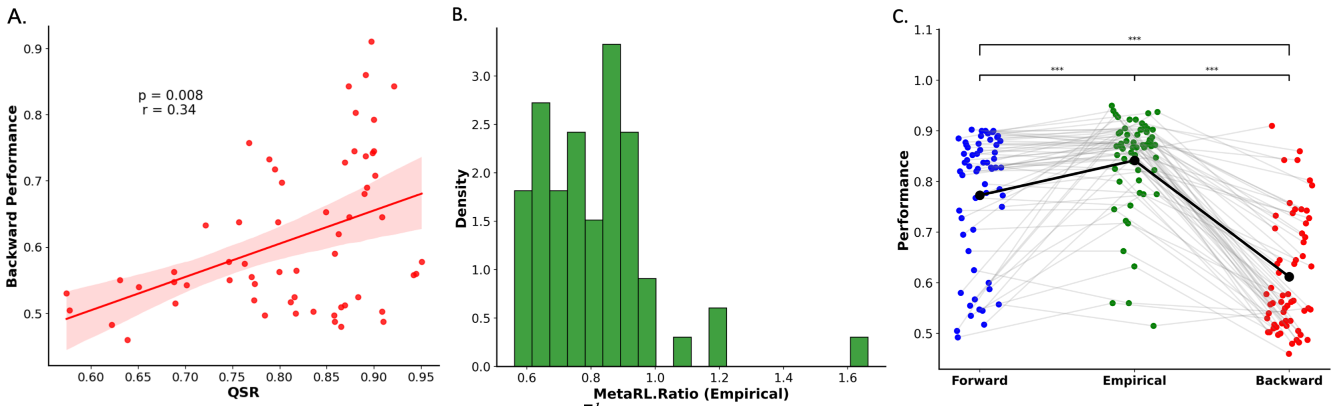

Figure 2. A) The performance of the Backward model, $\bar{R}_s^b$ was positively correlated with the empirical QSR. B) Histogram of the MetaRL-Ratio for the empirical data. C) Comparison of empirical performance (green) with those of the Forward ($\bar{R}_s^f$, blue) and Backward ($\bar{R}_s^b$, red) models. Thin black lines connect subjects; thick lines connect the means.

ity measures, such as the QSR (Carpenter et al., 2019), do not admit natural measures of efficiency. We then conducted various tests, including whether $\bar{R}_s^b$ is correlated with QSR; whether the MetaRL-Ratio is generally less than 1 (a form of meta-cognitive inefficiency that is often found in the perceptual case; Maniscalco & Lau, 2012; Fleming, 2017), and whether, if we simulate artificial meta-cognitively incompetent agents ('low'; by sampling confidence values at random) and competent agents ('high'; by setting confidence to a large/small value for choices associated with the better/worse bandit respectively), the MetaRL-Ratio can discriminate appropriately.

Results

Consistency of Backward performance with QSR. Our measure of meta-cognitive sensitivity, $\bar{R}_s^b$, was positively correlated with the Brier score QSR ($t(58) = 2.747$, $r = 0.340$, $p = 0.008$, 95%CI $= [0.093, 0.546]$, Pearson correlation; figure 2A).

Meta-cognitive inefficiency. The Backward model generally underperformed the Forward model ($Z = 67.0,\ p = 4.3e^{-10}$; figure 2C), MetaRL-Ratio was lower than 1 for 56 out of 60 subjects (figure 2B). The empirical performance was higher than Forward performance ($Z = 234.0\ ,\ p = 2.5e^{-6}$) and also higher than the Backward performance ($Z = 4.0,\ p = 2.931e^{-11}$) (all statistics were from Wilcoxon-signed-rank test between two groups).

Discrimination between high versus low meta-cognitive agents. By construction, simulated 'low' and 'high' meta-cognitive agents have the same Forward performance $\bar{R}_s^f$. The Backward performance $\bar{R}_s^b$ was worse than this for the 'low' agents , mean across 30 sampling for confidence rates, ($Z = 110,\ p = 3.10e^{-9}$), and better for the 'high' agents ($Z = 185.5, p = 7.85e^{-8}$), with the MetaRL-Ratio for the former duly being lower than for the latter ($Z = 12.0,\ p = 2.98e^{-11}$) (figure 3).

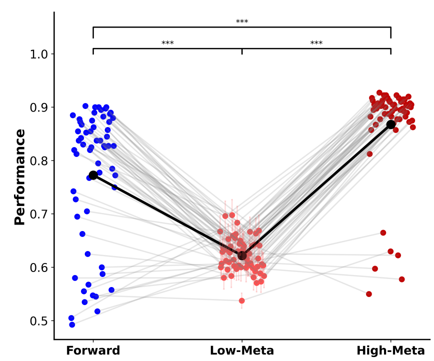

Figure 3. The performance of the Forward model (blue) was between the performance of the Backward model for simulated 'low' (light red) and 'high' (dark red) meta-cognitive agents. The Backward performance for low meta-cognitive agent was lower than the Backward performance for high meta-cognitive agent.

Conclusion

We report novel measures of meta-cognitive sensitivity and efficiency for RL problems. The Backward model, which attempts to simulate the choices implied by subjects' confidence judgments, was inspired by the meta$-d'$ in perceptual decision-making, and is a measure of sensitivity. The MetaRL-Ratio, inspired by the M-Ratio, corrects for the first-order performance of the subjects. We evaluate both Forward and Backward model by their reward rates when run autonomously on the task – this is slightly farther removed from the quality of the empirical performance of the subjects than the perceptual equivalent, d'. However, it correctly accommodates the ever-changing degree of difficulty of the underlying task (as the bandits switch their relative values in an unsignalled manner). We showed that $\bar{R}_s^b$ covaries with a standard measure of meta-cognitive sensitivity; and that the MetaRL-Ratio is appropriate to discriminate between suitably simulated subjects. We also found that our subjects were generally meta-cognitively inefficient. We will next extend this measure to a broader range of RL problems, including those in common use for assessing cognitive disorders.

References

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, *148*(1), 51.

Fleming, S. M. (2017). Hmeta-d: hierarchical bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of consciousness*, *2017*(1), nix007.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, *21*(1), 422–430.

Rahnev, D., & Fleming, S. M. (2019). How experimental procedures influence estimates of metacognitive ability. *Neuroscience of consciousness*, *2019*(1), niz009.

Sutton, R. S., Barto, A. G., et al. (1998). *Introduction to reinforcement learning* (Vol. 135). MIT press Cambridge.