

Value-guided abstraction maps arbitration in a dynamic environment

Hugo Six¹, Aurelio Cortese^{1,2,3}

¹ *Computational Neuroscience Labs, ATR Institute International, Kyoto, Japan*

² *Department of Brain Science & Engineering, Sungkyunkwan University, Suwon, South Korea*

³ *Centre for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea*

Introduction

Abstraction enables humans to make decisions efficiently by focusing on relevant information while filtering out distractions. It has been previously shown that abstraction occurs during learning when only a subsample of the available information is relevant to the decision at hand (Cortese et al., 2021). An open question is how the brain dynamically selects and utilises abstract representations in decision-making while the optimal abstraction changes over time. Since metacognitive ability allows for uncertainty appraisal and error correction (Fleming & De Martino, 2014), it could be a good candidate for piloting abstract representation arbitration.

Methods

We designed a new two-step decision-making task in which participants learned the value of specific abstractions through trials and errors. Specifically, participants had to first select a set of two out of three features while also indicating their confidence in the decision and then select an associated item (a fruit). Only trials where participants correctly solved both steps were rewarded. The relevant features and the associations defined by them changed at random times following distinct probabilities. Therefore, participants had to arbitrate whether errors were due to a change in the feature rule or in the association.

Results

Using behavioural data, we first validated participants' ability to solve this challenging task and provide a congruent confidence judgment about the relevant abstract representation. Next, we used a Mixture-of-Experts (MoE) computational model (Cortese et al., 2021; De Martino & Cortese, 2023) to simulate agents solving the task, showing how value decreases differently depending on errors arising at the first or the second step (i.e., errors in rule selection vs errors in selecting the associated item). Using the model fitted to participants' data, we further demonstrate that trial-by-trial value signals for the chosen abstract representation closely correlate with trial-by-trial subjective confidence, enabling us to compute the entropy of the value distribution for each trial, thereby creating an uncertainty metric. After using MRI data to validate the decodability of visual features in the early visual cortex, we show that this decodability is enhanced in high-confidence trials compared to low-confidence trials. Finally, using Decoded neurofeedback in a within-participant design, we induced neural activity patterns associated with high and low entropy of the abstraction value distribution. We observed that in the low-entropy condition, participants took longer to change their choice of feature pairs after error feedback due to a change in relevant abstraction, but not after error feedback due to a change in the feature-fruit association.

Conclusions

Our current results suggest that humans track values of competing abstract representations in a dynamic environment and use these values to arbitrate between them when an error is encountered. Moreover, when neural patterns associated with abstraction uncertainty, which are highly correlated with metacognitive ability, were induced using decoded neurofeedback, we observed an alteration in participants' flexibility in updating such values. Finally, our results suggest a potential causal link between metacognition and the dynamic control of abstraction during learning.

References (optional, maximum 6)

Cortese, A., Yamamoto, A., Hashemzadeh, M., Sepulveda, P., Kawato, M., & De Martino, B.

(2021). Value signals guide abstraction during learning. *eLife*, 10.

<https://doi.org/10.7554/eLife.68943>

De Martino, B., & Cortese, A. (2023). Goals, usefulness and abstraction in value-based choice. *Trends in Cognitive Sciences*, 27(1), 65–80.

<https://doi.org/10.1016/j.tics.2022.11.001>

Fleming, S. M., & De Martino, B. (2014). Metacognition and Confidence in Value-Based

Choice. In *The Cognitive Neuroscience of Metacognition* (pp. 169–187). Springer Berlin

Heidelberg. https://doi.org/10.1007/978-3-642-45190-4_8