
Exploring the Influence of Semantic Labeling on Task Representation and Reinforcement Learning

Mina Habibi*¹, Pieter Verbeke¹, Mehdi Senoussi², and Senne Braem³

¹Department of Experimental Psychology, Henri Dunantlaan 2 Ghent, Belgium – Belgique

²Université Toulouse II – CLLE-ERSS (UMR 5263) – France

³Department of Experimental Psychology, Ghent University – Belgique

Résumé

Introduction: In reinforcement learning, humans demonstrate remarkable adaptability in high-dimensional environments, beginning with feature-based learning and transitioning to a more accurate object-based strategy. This study investigates how access to semantically meaningful response labels during the initial learning of three separate tasks can influence task representations and learning strategies in a subsequent value-based decision-making task. By examining the effects of semantic labels, we aim to determine whether labeling increases the separability of task representations, affects knowledge transfer, and facilitates earlier transitions from a feature-based to a more accurate object-based learning strategy. **Method:** Participants ($N = 117$) were divided into two groups: a Semantic-labeling group ($n = 59$, age = 29.2 ± 3.81), which received distinct semantic labels for tasks, and a No-labeling group ($n = 58$, age = 28.87 ± 4.56). First, participants learned to associate one of three features of stimuli with corresponding tasks, where one feature was linked to one task. In the second phase, participants needed to combine all three features for a value-based decision-making task, where only one feature was partially informative; while two other features individually could not predict the value. To explore the underlying task representations during initial task learning, we fitted neural networks to the human data. We computed the correlation between task representation separation in this initial learning phase and accuracy during the subsequent value-based decision-making task. Additionally, we analyzed choice data and computed Bayesian Information Criterion (BIC) values to assess model fit for feature-based and object-based learning strategies. **Results:** The Semantic-labeling group exhibited greater task representation separation compared to the No-labeling group ($p = 0.04$). However, the correlation between task representation distance and performance in the VBDM task did not differ significantly between groups. A significant difference in the correlation between performance accuracy in the initial learning phase and in the VBDM phase was found between the two groups (Fisher's $z = 2.56$, $p = .01$). The Semantic-labeling group showed a strong positive correlation ($r = 0.52$), while the No-labeling group showed a weaker correlation ($r = 0.08$). BIC analysis suggested that both groups favored the feature-based strategy (Semantic-labeling: $p < .0001$; No-labeling: $p = .024$), with no clear transition to object-based learning observed. However, in the Semantic-labeling group, the feature-based model showed equal weight distribution across the three features, whereas the No-labeling group assigned greater weight to the partially informative feature. **Discussion:** These findings indicate that semantic labeling enhances task representation, and improves task performance and knowledge transfer. Although no transition to

*Intervenant

an object-based learning strategy was observed in either group, the equal weight distribution for features in the Semantic-labeling group suggests a potential for a faster transition to the object-based strategy with extended training. **Conclusion:** Overall, the results emphasize the importance of semantic labeling in supporting effective task learning, representation, and knowledge transfer, highlighting its potential applications in improving AI algorithms to enhance knowledge transfer and continual learning. **Keywords: Semantic Labeling, Task Representations, Reinforcement Learning**