Tracking of dynamic neural representations during goal-directed action

Jet Lageman (j.lageman@vu.nl)
Institute Brain and Behaviour Amsterdam, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081 HV Amsterdam, Netherlands

Atsushi Kikumoto (atsushi_kikumoto@brown.edu)
Department of Cognitive and Psychological Sciences, Brown University, 190 Thayer Street Providence, RI 02912, United States of America

David Badre (david_badre@brown.edu)
Department of Cognitive and Psychological Sciences, Brown University, 190 Thayer Street Providence, RI 02912, United States of America

Johannes J. Fahrenfort (j.j.fahrenfort@vu.nl)
Institute Brain and Behaviour Amsterdam, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081 HV Amsterdam, Netherlands

Heleen A. Slagter (h.a.slagter@vu.nl)
Institute Brain and Behaviour Amsterdam, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081 HV Amsterdam, Netherlands

Abstract

Theoretical accounts of goal-directed action disagree on whether actions are represented by their motor commands or sensory outcomes. Here we use a novel analysis technique combining linear decoding of EEG data with representational similarity analysis to shed light on this matter. Preliminary findings (n=5) show that it is possible to track neural conjunctive representations that integrate goals, actions, and sensory outcomes over time during goal-directed action, which can be related to learning outcomes. These findings help pave the way towards a mechanistic understanding of how the brain plans actions to reach goals.

Keywords: action planning; active inference; EEG; representational similarity analysis.

Introduction

Goal-directed action is commonly defined as grounded in knowledge about action-outcome contingencies (Dickinson & Balleine, 1994). However, there are different theoretical viewpoints about how the brain plans goal-directed actions to attain desired outcomes. Motor command theories state that actions are represented and planned in terms of motor output, by an inverse model that converts sensory goals to motor commands (Todorov, 2004; Wolpert & Kawato, 1998). These motor commands are then fed to a forward model to compute their expected sensory outcomes (Franklin & Wolpert, 2011). According to this viewpoint, learning mainly involves the optimization of the inverse mapping between goals and actions (Wolpert & Flanagan, 2010). On the other hand, active inference states that goal-directed actions are represented and planned in terms of their sensory consequences (Adams et al., 2013). This viewpoint predicts that learning mainly involves the integration of actions and their sensory outcomes (Proietti et al., 2023), similar to the formation of 'event files' that include goals, actions, and outcomes (Hommel, 2004).

The aim of the current study is to determine whether goal-directed action planning fundamentally incorporates sensory outcomes by examining if action learning is associated with the development of highly integrated, conjunctive representations using EEG. Participants performed goal-directed actions in order to generate specific sensory outcomes (Figure 1). Using a novel analysis technique combining linear decoding with representational similarity analysis, we could track the representational strength of several task features over time, as well as their non-linear integration into high-dimensional conjunctive representations (Badre et al., 2021; Kikumoto & Mayr, 2020). These conjunctive representations have previously been shown to play a critical role in flexibly responding to stimuli. Here, we extend this work by investigating their role in goal-directed action-outcome learning.

Methods & Results

Experimental paradigm Participants had to generate specific sensory outcomes (Figure 1). These outcomes were defined by 2 independent features: their shape (circle or star) and spatial frequency (high or low). In every block, each unique outcome was mapped to one of 4 locations on the screen. Participants could generate outcomes by moving a white square horizontally (left/right) or vertically (up/down) towards the location of the outcome they aimed to generate. Importantly, they first needed to learn the location-outcome mapping (i.e., the relationships between stimuli, actions, and outcomes) in a given block before they could generate the correct outcomes. Each block, the location-outcome mapping would switch between 4 possible mappings, such that

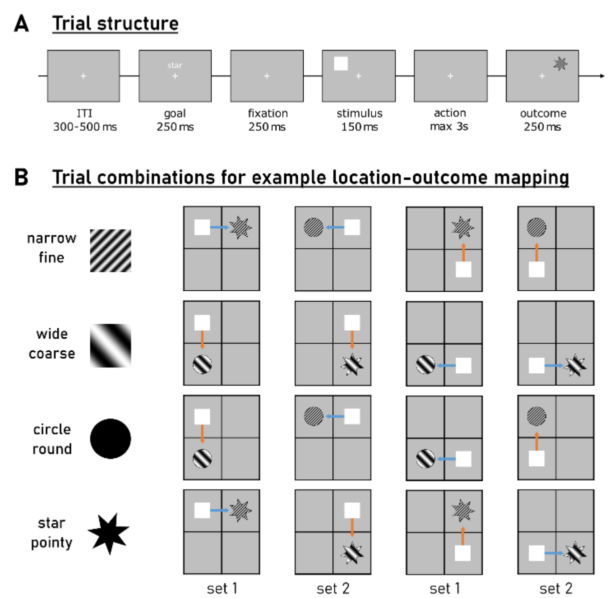

Figure 1. **A** Trial structure. **B** Overview of all 16 trial combinations for one location-outcome mapping. To orthogonalize stimulus location and outcome location, these combinations are divided into 2 sets.

each outcome appeared equally often in each location over the course of the experiment. On each trial, participants were given a goal feature to generate (e.g., circle), after which a white square occurred on one of the 4 screen locations. The participant then had to generate an outcome with the goal feature by pressing one of two buttons on the keyboard ('X' or 'M', corresponding to 'horizontal' or 'vertical', action mapping counterbalanced between participants).

Preliminary data We so far collected behavioral and EEG data from 5 participants (all female, mean age = 21.6; data collection ongoing, final sample n=40) over 2 sessions per participant. Each session consisted of 2 practice blocks of 64 trials and 12 experimental blocks of at least 160 trials, or until performance in the last 16 trials exceeded 85%.

EEG recording EEG activity was recorded from 64 electrodes (10-20 system). Data was digitized at 512 Hz and down-sampled to 128 Hz. For preprocessing, data was re-referenced to the average of the mastoids, high-pass filtered at 0.01 Hz, and eye-blink artifacts were removed using ICA. Finally, data was epoched between -600 ms and 800 ms around stimulus presentation.

Data analysis A combination of linear decoding and representational similarity analysis (Kriegeskorte et al., 2008) was used to track the dynamic representation of task features and their conjunction over time, following the procedure by Kikumoto & Mayr (2020). Using the ADAM toolbox (Fahrenfort et al., 2018), we trained a linear discriminant classifier to discriminate between each of the

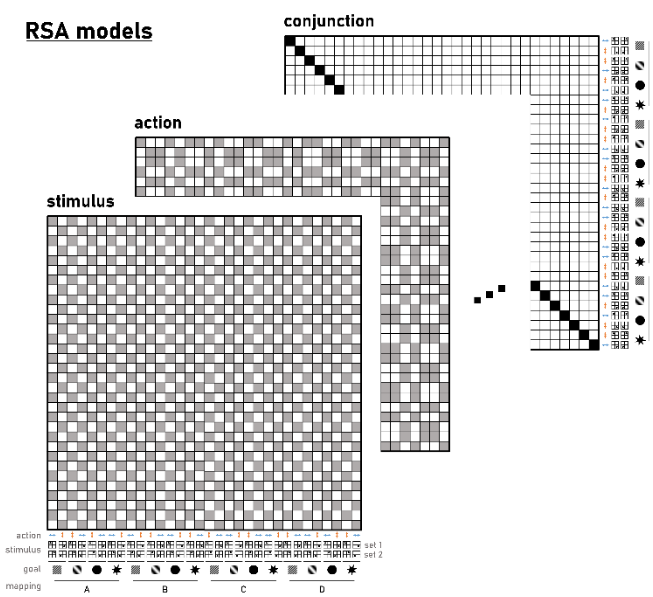

Figure 2. RSA models.

32 different trial combinations in each set with 5-fold cross-validation. This classification resulted in a 32-by-32 confusion matrix for each trial and timepoint, which were log-transformed before serving as the input for RSA.

We constructed RSA models for the goal, stimulus location, action, outcome identity, location-outcome mapping, and their conjunction (Figure 2). In addition, we controlled for possible differences in difficulty between conjunctions by adding a predictor with subject-specific conjunction RTs. Results from this analysis indicate that we can successfully track the representational strength of individual task features and their conjunction over time (Figure 3).

We plan to relate RSA predictor values to trial-to-trial performance and RT with multilevel linear modeling in a larger sample, to determine whether the strength of conjunction representations predicts learning performance.

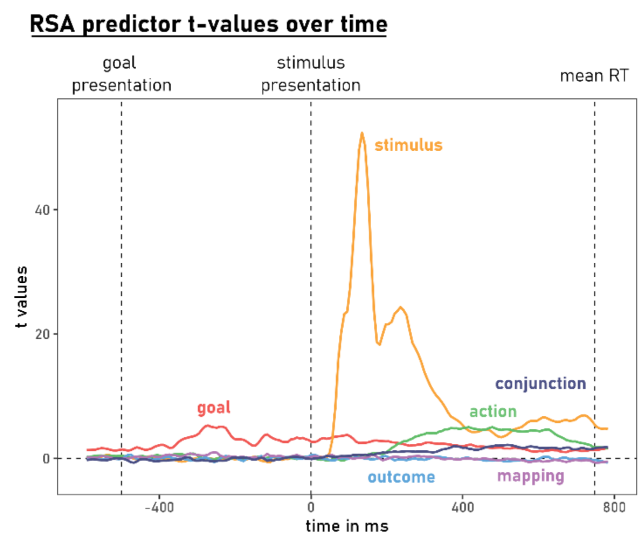

Figure 3. Results of combined LDA-RSA analysis indicate that we can track the strength of neural representations over time.

Conclusion

The current study aimed to investigate whether goal-directed actions are represented by motor commands or sensory outcomes. First results show that by combining linear decoding with RSA, we can track dynamic neural representations during goal-directed action. This will allow us to arbitrate between competing hypotheses from motor command theories and active inference, enhancing understanding of how humans learn to reach their goals.

Acknowledgements

This work was supported by an ERC consolidator grant (PlasticityOfMind; 101002584) to H.A.S.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*, 611-643.

Badre, D., Bhandari, A., Keglovits, H., & Kikumoto, A. (2021). The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences*, *38*, 20-28.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior, 22*(1), 1-18.

Fahrenfort, J. J., Van Driel, J., Van Gaal, S., & Olivers, C. N. (2018). From ERPs to MVPA using the Amsterdam decoding and modeling toolbox (ADAM). *Frontiers in Neuroscience*, *12*, 368.

Franklin, D. W., & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*(3), 425-442.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494-500.

Kikumoto, A., & Mayr, U. (2020). Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. *Proceedings of the National Academy of Sciences*, *117*(19), 10603-10608.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 249.

Proietti, R., Pezzulo, G., & Tessari, A. (2023). An active inference model of hierarchical action understanding, learning and imitation. *Physics of Life Reviews*, *46*, 92-118.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907-915.

Wolpert, D. M., & Flanagan, J. R. (2010). Motor learning. *Current Biology*, *20*(11), R467-R472.

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7-8), 1317-1329.