Model abstraction for model-based reinforcement learning in the human orbitofrontal cortex

Yu Takagi (yu.takagi@psych.ox.ac.uk) [1]
University of Oxford, Oxford Centre for Human Brain Activity, United Kingdom

Wako Yoshida (yoshida-w@sys.i.kyoto-u.ac.jp) [2]
Kyoto University, Department of Information Science, Japan
ATR, Brain Information Communication Research Laboratory Group, Japan

Saori C. Tanaka (xsaori@atr.jp) [2]
ATR, Brain Information Communication Research Laboratory Group, Japan

[1] Corresponding author, [2] These authors contributed equally

Abstract:

How the human brain represents multiple models of the environment for decision-making (model-based reinforcement learning, MB-RL) is not well understood. We hypothesized that models are efficiently represented based on the similarity among them, to reduce redundancy, and this technique is called 'model abstraction' in the field of AI research. We designed a novel sequential learning task in which participants were required to simultaneously learn multiple models with a hidden latent structure, and studied corresponding brain activity using fMRI. By using an MVPA, we found that human OFC encodes the models reflecting the similarity among them. The degree of this 'model abstraction' ability was correlated with individual behavioral performance. Our results suggest that the human brains represent multiple models in a compact space, and this allows us to efficiently learn complex environments.

Keywords: fMRI; Decision-making; Model-based RL; Mental simulation; Orbitofrontal cortex

Introduction

Humans can flexibly deal with complex environments by using mental models, which are often learned through experience. This ability comes under the umbrella of model-based reinforcement learning (MB-RL), and allows decision-making based on an internal representation of the environment and a knowledge of how actions may lead to various consequences. Although previous studies have shown that humans use a MB-RL strategy in laboratory decision-making tasks, the tasks in these studies have relied on a single, explicit environmental model (Daw, Gershman, Seymour, Dayan, & Dolan, 2011). In the real world, however, the environment may permit many models, and somehow the human brains can represent them in an efficient manner, despite limited memory capacity. This ability is called 'model abstraction' in the field of artificial intelligence, and is an important topic in AI research. Recent studies showed that humans brains represent abstract structures of items (Tang et al., 2019) or state representations (Schuck et al., 2016). Here, we hypothesized that humans also utilize such dimensionality reduction when representing environmental models for decision-making.

To test this, we conducted a functional magnetic resonance imaging (fMRI) with a novel sequential learning task. In this task, participants performance relied on simultaneously learning multiple models with a hidden latent structure and updating them to follow alternations of the models. We employed a multi-voxel pattern analysis (MVPA) to investigate whether the brain activity patterns can be explained by the model's latent structure. Furthermore, to confirm that the participants employed a model-based strategy (Doll, Duncan, Simon, Shohamy, & Daw, 2015), we used MVPA to quantitatively define the degree of mental simulation, which is crucial for model-based strategy, and showed that it is correlated with individual behavioural performance of model learning.

Methods

fMRI experiment

21 participants performed a modified version of the multi-step sequential learning task (Momennejad et al., 2017) in an fMRI scanner. Participant consent was obtained in accordance with a protocol reviewed and approved by the Ethics Committee of the Advanced Telecommunications Research Institute International. A 3-T Siemens Prisma scanner with a 12-channel head coil was used to perform T2*-weighted echo planar imaging. The scanning parameters were: TR, 1000 ms;

TE, 30 ms; FA, 60°; FOV, 192 × 192 mm; 72 slices; and a 2.0-mm slice thickness without gap.

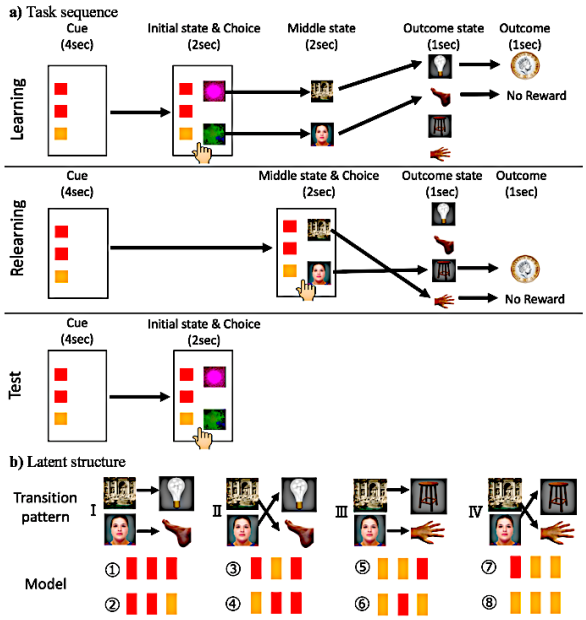


Figure 1: Task design. a) The task consists of three phases. The semantic categories of stimuli in the middle and outcome states were used for the decoding. b) Example of the latent relationship. In the relearning phase, we switched model-transition relationship for 4 models so that every transition pattern includes one swapped model: e.g., swapping the model 1 for model 5, and swapping model 3 for model 7.

In this task, there were three phases: learning, relearning, and test phases. First, in the learning phase, participants learned two-step transitions leading to a monetary outcome by trial and error. At the beginning of each trial, a triplet of colored boxes (yellow or red) were presented for 4 secs ('Cue' in Fig. 1). After that, two fractal images appeared (initial state) and the participants were asked to choose one within 2 secs. Then, a face or a scene image (middle state) was shown for 2 secs followed by a body or an object image (outcome state) with or without a reward (coin image). All state transitions were deterministic.

The transition rules from the initial and middle state were fixed through the experiment, while the transitions rules from the middle and the outcome state were defined by the pattern of triplet of boxes; here, we referred to this as a 'model'. There were eight different models but there was a latent structure and their corresponding transition patterns were overlapped (Fig. 1b). Namely, there were four 'transition patterns', I-IV, and two models belong to each transition pattern. We also define the model similarity in terms of the image categories in the outcome states and the transitions from the middle to the outcome states, i.e., transition pattern I and II are similar, while transition pattern I and III are dissimilar. Participants were instructed that the pattern of triplet of boxes defines the transition from the middle to the outcome state, but not the latent structure. Participants performed 2 sessions, and each session consists of 64 trials; i.e., they performed 16 trials per model. The model-transition relationship and the order of the trials were pseudo-randomized across participants.

In the relearning phase, participants were asked to make a choice on the middle state not the initial state. Participants were solely exposed to the triplet of boxes for 4 secs, and then chose one of two images. For half of 8 models, the mapping between the model and transition pattern was switched so that every transition pattern will have swapped model and relearning was required (Fig. 1b). Participants performed 64 trials, i.e., 8 trials for each model in this phase.

Finally, in the test phase, participants made a choice on the initial state after the display of triplet of boxes for 4secs. Unlike the learning phase, no state transition and reward feedback were displayed to the participants. The participants were asked to choose a fractal image which leads a rewarded outcome state, and paid based on the amount of rewards accumulated in all phases. There were 32 trials in this phase.

fMRI decoding analysis

We examined whether the participants employed (1) the latent structure of models and (2) a model-based strategy by using neural decoding techniques. For both analyses, we extracted patterns of blood-oxygen-level-dependent (BOLD) response during the 'Cue' in three phases separately, and applied support vector machine (SVM; Cortes & Vapnik, 1995) for classification (Fig. 2).

To identify the structure of model representations in the brain, we defined a 'model abstraction score' which evaluates the similarity between the brain activity patterns of the models. For each participant, we conducted an SVM classification for 8 models and obtained the outputs, namely 8 decision values (DVs), for all trials. Here, DVs can be considered as a neurally-defined similarity between different models, because models which have similar neural representation should have high DVs. We then calculated a 'confusion matrix' D (Fig. 2a) of which element is defined as:

$$D(I,J) = \mathrm{E}\left[\sum_{i\in I}\sum_{j\in J} DV_{i,j}\right]$$

where $DV_{i,j}$ is the decision value for model *j* obtained from the brain data of trials with model *i*. Note that the diagonal components, $D(I,I)$, did not include the value of the DVs for the same model, e.g. classifier's output for model *i* from input data of trials with model *i*. We compare this confusion matrix and a 'template matrix', which represents the latent structure of the models and has ones for the models with same transition pattern, otherwise zeros. We defined a 'model abstraction score' as Pearson's correlation coefficient between the confusion matrix and the template matrix. As the model's latent structure, i.e., template matrices, are different in the learning phase and in the relearning and test phases, we used the original template for learning phase, whereas switched template for relearning and test phases.

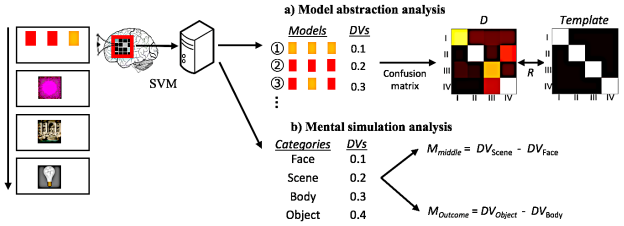


Figure 2: MVPA analysis. **a)** For model abstraction analysis, an eight-class SVM was applied to single-trial brain activity patterns during the Cue presentation. Outputs of the classifier for each model construct a confusion matrix, and the similarity between the confusion matrix and a template matrix is calculated as a model abstraction score. **b)** For mental simulation analysis, the outputs of four-class SVM with image semantics were used to estimate the mental simulation score.

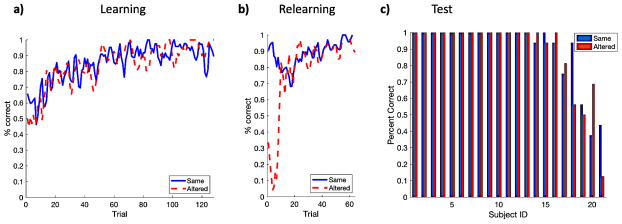
To investigate whether the participants used a model-based strategy, we evaluated the degree of mental simulation from decoding accuracies. First, to construct a decoder for image semantics, we conducted an experiment where the participants observed 60 different images with each of four semantic categories (face, scene, body and object) used in the main experiment. We next constructed a 4-class SVM decoder from activity patterns during the Cue step. Outputs of the SVM, DVs, can be considered as a degree of mental simulation for the future state. We defined a 'mental simulation score' *M* as the difference of DVs between visited and non-visited states for each step (Fig. 2b):

$$M_{\text{middle}} = \text{DV}_{\text{visited middle state}} - \text{DV}_{\text{non-visited middle state}}$$
$$M_{\text{outcome}} = \text{DV}_{\text{visited outcome state}} - \text{DV}_{\text{non-visited outcome state}}$$

Results

Through trial and error, participants acquired the correct choices over the course of the task (Fig. 3).

Figure 3: Behavioral results. **a)** For the learning and **b)** the relearning phases, 3-trial moving average of the behavioral performances are shown. **c)** For the test phase, percentage of correct responses are shown for each participant.



Abstract model representation in OFC

We tested how the models are represented in the brain. We used the orbitofrontal cortex (OFC) for the region of interest because previous work has suggested that OFC is involved in representation of environmental models (Schuck et al., 2016). We defined OFC by Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) and conducted a decoding analysis using all trials in each phase.

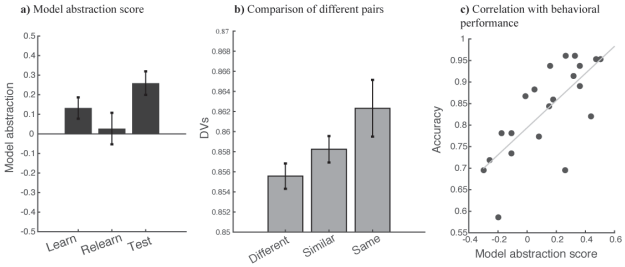


Figure 4: Abstract model representation in OFC. **a)** Model abstraction scores were significantly correlated with the confusion matrices in the learning and test phases. **b)** The confusion matrix in the test phase showed that the model similarity is preserved in the brain activity patterns. **c)** Individual model abstraction scores are correlated with the behavioural performance in the learning phase.

The model abstraction score was significantly positive both in the learning phase (Fig. 4a; $P < 0.03$; two-tailed one-sample *t*-test) and in the test phase ($P < 0.0004$; two-tailed one-sample *t*-test). Strikingly, the confusion matrix in the test phase, where the participants were required to engage mental simulation studiously, shows that the model similarity is structurally preserved in the brain (Fig. 4b). It is noteworthy that OFC has the greatest correlation with the template matrix, i.e., the highest model abstraction score, among all AAL regions.

We next examined whether the degree of model abstraction was related to participant's behavioural performance. Figure 4c shows that the individuals with higher model abstraction score achieved significantly

better choice accuracy in the learning phase (Spearman's $R = 0.74$; $P < 0.0014$). In the relearning and test phases, the model abstraction scores were not correlated with the choice accuracies ($P > 0.05$ for both phases).

Mental simulation leads to better performance

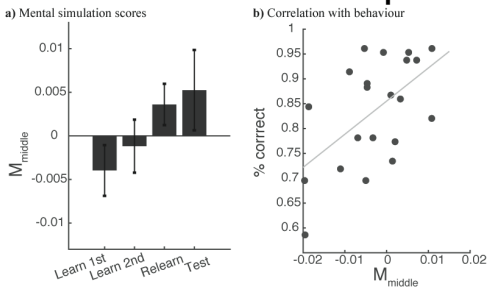


Figure 5: Mental simulation analysis. **a)** M_{middle} for each phase. **b)** Correlation between M_{middle} and the behavioural performance in the learning phase. $M_{outcome}$ were not correlated with behavioral performances ($P > 0.05$).

Next, we asked whether participants employed a model-based strategy. We obtained reasonable 4-class cross-validated classification accuracies of 65.4 ± 0.03 in occipital and fusiform regions defined by AAL (mean ± s.t.d). To test whether mental simulation score was increased through the experiment, we used a simple linear regression model that aimed to explain the modulation of experimental phase in terms of mental simulation score. The regression slope (β) associated with the experimental phase was significantly positive (Fig. 5a; $\beta = 0.0032$; $P < 0.032$), suggesting that the participants could do mental simulation especially in the later phases. We further confirmed that the individuals with higher mental simulation score achieved significantly higher choice accuracy in the learning stage (Fig. 5b, Spearman's $R = 0.5$; $P < 0.022$).

Discussion

In this study, we developed a three-phase sequential learning task in which participants store multiple models simultaneously, and investigated how human brain represents these models by fMRI decoding analyses.

We found that models are represented in OFC by representing the latent structure underlying the models. The degree of model abstraction was correlated with individual behavioral performance. It has been suggested that the human brain abstracts relationships between items (Tang et al., 2019) or states (Schuck et al., 2016). and our study suggests that the human brain also performs abstraction in the context of model space for MB-RL. We also found that participants engaged in mental simulation, wherein the degree of simulation was correlated with behavioral performance.

Our results suggest that humans efficiently store environmental models for decision-making using model abstraction. Since recent studies have found abnormalities in MB-RL in patients with psychiatric disorders (Gillan, Kosinski, Whelan, Phelps, & Daw, 2016), it is possible that disruption of model abstraction could play a contributory role in their pathogenesis.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science (16H06396, 16K21720, 16H06395), and Versus Arthritis (21537). We thank Ben Seymour for proofreading the manuscript.

References

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215.

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, (March), 1–9.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal- directed control. *ELife*, *5*(e11305).

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680.

Schuck, N. W., Cai, M. B., Wilson, R. C., Niv, Y., Schuck, N. W., Cai, M. B., … Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space Article Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412.

Tang, E., Mattar, M. G., Giusti, C., Lydon-staley, D. M., Thompson-schill, S. L., & Bassett, D. S. (2019). Effective learning is accompanied by high-dimensional and efficient representations of neural activity. *Nature Neuroscience*, *22*(June), 1000–1009.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289.