

The relational structure of a reinforcement learning task is represented and generalised in the entorhinal cortex

Alon B Baram (alon.baram@merton.ox.ac.uk)¹
Timothy H Muller (timothymuller127@gmail.com)¹
Hamid Nili (nili.hamed@gmail.com)¹
Mona Garvert (garvert@cbs.mpg.de)²
Timothy E J Behrens (behrens@fmrib.ox.ac.uk)¹

¹ Wellcome Centre for Integrative Neuroimaging, University of Oxford,
John Radcliffe Hospital, Oxford OX3 9DU, UK

² Max-Planck-Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103, Leipzig, Germany

Abstract

The ability to appropriately generalise previously acquired knowledge to novel situations is a hallmark of human intelligence. A possible neural solution to this problem is to devote pools of neurons to represent the relations between entities in the environment explicitly, in a manner that is divorced from the entities themselves. Such an explicit representation can generalise to novel situations with the same relational structure. Grid cells, originally found in the entorhinal cortex, have been proposed as such an explicit representation of the relations between different locations in physical space. However, the neural representations underlying the generalisation of relational structures in abstract tasks remain poorly understood. Here we use fMRI in humans to show that the entorhinal cortex explicitly represents the relations between reward-predicting stimuli in a reinforcement learning task with different underlying correlation structures between the reward probabilities associated with different stimuli. Our results demonstrate that the same brain regions, perhaps with the same mechanisms, represent the relational structure of the task in both spatial and abstract decision-making tasks. This suggests that the brain uses a common coding framework for the structure of tasks across a wide range of domains.

Keywords: Cognitive maps; generalisation; schemas; Grid cells; reinforcement learning; decision making

Introduction

The term “cognitive map” was coined by Tolman (1948) to describe the relational internal model underlying the flexible inferences his rats were making in complex spatial mazes. However, the ability of animals and humans to use such internal models to generalise

knowledge to novel situations is not unique to the spatial domain (Behrens et al., 2018).

How might this relational knowledge be represented in the brain? One option is to encode the relations between entities (e.g. locations or objects) in the strength of the synapses between pools of neurons representing the different entities. However, this representation is not generalizable: it is tied to the identities of the specific entities. To allow for generalisation of the relational structure, its representation must be explicit – divorced (abstracted) from the sensory particularities of the task or the entities in question (Behrens et al., 2018).

The well-studied domain of spatial cognition has revealed a candidate explicit and generalisable representation of the structure of 2D spatial tasks: “grid” cells, originally found in the entorhinal cortex (EC), fire when an animal is in one of multiple locations on an equally spaced triangular lattice (Hafting, Fyhn, Molden, Moser, & Moser, 2005). Experimentally, grid cells maintain (generalise) their firing covariance structure across perceptually different rooms (Fyhn, Hafting, Treves, Moser, & Moser, 2007). This is only true when in both rooms the animal is required to perform the same task - free foraging. Crucially, the grid code changes when the structure of the task changes (Boccaro et al., 2019; Butler, Hardcastle, & Giocomo, 2019). Theoretically, grid-like firing patterns emerge as a low-dimensional representation of the covariance of place cells firing and of 2D open-field state transition matrices (Banino et al., 2018; Dordek, Soudry, Meir, & Derdikman, 2016; Stachenfeld, Botvinick, & Gershman, 2016), suggesting grid cells activity during free navigation encodes the statistical regularities common to 2D open-field environments. Taken together, this suggests that the knowledge embedded in grid cells generalises across environments and tasks with the



same relational structure, but might “remap” when the structure is different (for a review see Behrens et al., 2018). These are exactly the properties that are required from an explicit representation of relational structure.

We hypothesised that the same brain regions where grid cells can be found will code for the relational structure of a non-spatial reinforcement learning task. To test this, we designed a stimulus-outcome association task with two underlying correlation structures between the outcome probabilities associated with stimuli. We could thus test for a brain region that represents the same stimulus differently, depending on the nature of its relationships with another stimulus. Importantly, we also used a second stimuli-set, resulting in a 2x2 factorial design of stimuli set x structure. This enabled us to also test for the other requirement of an explicit structural representation: it should generalise across stimuli with the same relational structure, like grid cells generalise when an animal is free foraging in different 2D boxes (Fyhn et al., 2007). As we hypothesized, we found that the entorhinal cortex explicitly coded for the relational structure between stimuli when they were presented.

Results

Task and behaviour

We trained participants on a probabilistic stimulus-outcome association task with two sets of three stimuli. Only one of the stimuli sets was used in each block. In each trial, participants viewed one of the three stimuli in

pseudo-random order, and had to indicate their prediction for its associated binary outcome (a “good” or a “bad” outcome) by either accepting or rejecting the stimulus (Fig 1a). Thus, there was always one correct answer in each trial: participants should accept if they predict the outcome to be the “good” outcome, and should reject if they predict the outcome to be the “bad” outcome. Outcome identity was revealed in all trials, including rejection trials, even though the participant’s score did not change in these trials (Fig 1b). Predictions of the outcomes could be formed based on the recent history, as the probabilities of outcomes for each stimulus switched pseudo-randomly between 0.9 and 0.1 with an average switch probability of 0.15. Crucially, for a given stimuli set, the outcome probabilities associated with two of the stimuli were positively correlated (+Corr) in half of the blocks, and negatively correlated (-Corr) in the other half, such that participants could learn from the outcome on one correlated stimulus about the other (Fig 1c). The third stimulus served as a control, and had an independent outcome probability (0Corr). Thus, there were four block types, arranged in a 2x2 factorial design of stimuli set by correlation structure (Fig 1d). In the fMRI experiment there were two independent runs of the four block types, with a pseudo-random block order counterbalanced across participants. The current block-type was signaled by the background color of all stimuli in the block. Participants pre-learned the mapping between background color and correlation structure prior to scanning. Hence, the only learning performed during scanning was of reversals/outcome probabilities, not of the relational structure - knowledge of which was available from the first scanning trial.

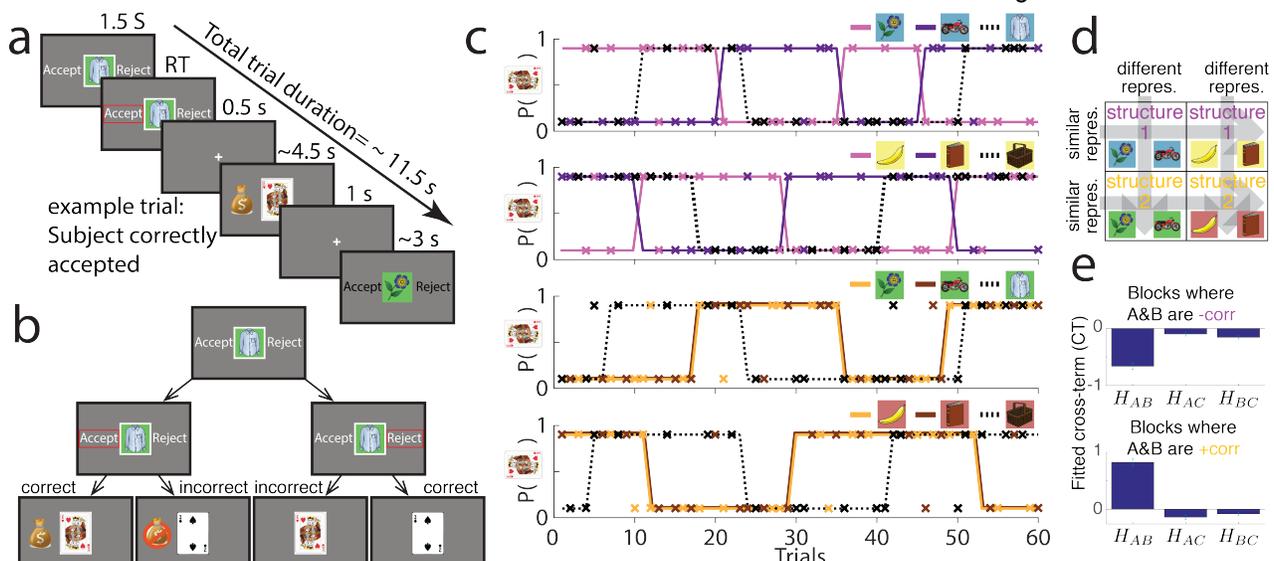


Figure 1: a. Example trial. b. Possible trial flows, depending on choice and predetermined outcome c. example schedules of the 4 block types. d. schematic of the factorial design. e. Fitted CTs, indicating whether participants treated feedback from pairs of stimuli as the same (CT=1), opposite (CT=-1) or independent (CT=0).

We modelled the subjects' behavior using an adapted delta-rule with cross-terms (CTs) that enable learning from one stimulus to another fitted to behaviour. The fitted CTs indicated that participants indeed used the correlation structure correctly (Fig 1e).

The reward network and the hippocampus use the relational structure

We first wanted to test whether known neural signals of reinforcement learning showed evidence of knowledge about the relational structure. We tested this by comparing how well a model that utilised relational structure explained neural signals relative to one that does not utilise structure (all fMRI analyses were performed on the correlated stimuli only, and ignored the control stimulus). In both models, we calculated the value of the chosen action (accept/reject) and the value prediction error on each trial of the two correlated stimuli. The first model was a naïve Rescorla-Wagner model (NAÏVE, cross-terms set to zero), and the second model utilised the relational structure (STRCT, cross-terms fit to behavior). The chosen value estimates were used to construct two regressors at the time of stimulus presentation, and the prediction error estimates were used in a similar way to construct two regressors at the time of outcome. Regressors from both models were entered into the same GLM (together with the main event regressors of stimulus presentation, button press and outcome times for all three stimuli). As estimates from both models were pitted against each other in the same GLM, any variance explained by a particular regressor was unique to that regressor, allowing us to compare the neural signals uniquely explained by each model.

A network of regions including the medial prefrontal cortex (mPFC), the amygdala (AMG), the anterior hippocampus (HPC) and the entorhinal cortex coded positively for the chosen action value from the STRCT model, while most of the orbital surface showed strong negative coding (Fig 2A). The difference between the STRCT and NAÏVE chosen value effects was positive in EC, HPC, medial AMG, dorsal mPFC, parietal cortex and the insula, and negative in the orbital surface (Fig 2B). The STRCT model value prediction error estimates correlated with activity in the ventral striatum, HPC and AMG (Fig 2C). The same regions coded for the STRCT model prediction error more than the NAÏVE model (data not shown). These results are an almost exact replication of (Hampton, Bossaerts, & O'Doherty, 2006), indicating the brain uses the relational structure to calculate value and learning signals.

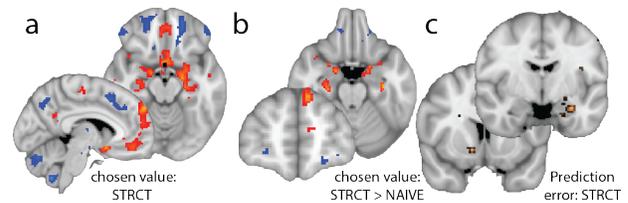


Figure 2: Univariate effects of RL parameters

Entorhinal cortex explicitly represents the relational structure of task events

An explicit neural representation of the relational structure of the task should be similar for stimuli which are part of the same relational structure, but dissimilar for stimuli under a different relational structure. We asked whether any region on the cortical surface displayed these properties at the times of stimuli presentations, using Representational Similarity Analysis (RSA, (Kriegeskorte, Mur, & Bandettini, 2008) with a searchlight approach (Kriegeskorte, Goebel, & Bandettini, 2006). A searchlight centered on a cortical voxel consisted of the 100 surrounding voxels with the smallest surface-wise geodesic distance from the central voxel. For each searchlight, we obtained 16 patterns of whitened regression coefficients of the responses to presentations of each of the two correlated stimuli in each of the 8 blocks. In other words, we obtained two patterns, one from each of the runs, for each of our 8 experimental conditions (a particular stimulus under a particular correlation structure). To define the “cross-run correlation distance” between conditions i and j ($d_{i,j}$) we first calculated the correlation distance ($1 - r$) between the condition i pattern from run 1 and condition j pattern from run 2, and then calculated the correlation distance between the condition j pattern from run 1 and condition i pattern from run 2. $d_{i,j}$ was defined as the mean of these two distances. Importantly, we never correlated conditions from the same block. This resulted in an 8 conditions by 8 conditions symmetric Representational Dissimilarity Matrix (RDM), summarising the representational geometry in the searchlight (e.g. Fig 3b). The ideal explicit structural representation can be formalised as an 8x8 model RDM, where the desired distances between conditions are determined by relational structure (Fig 3a). To test whether the data RDM of a given searchlight was consistent with the model RDM, we calculated the contrast between the means of the data RDM's hypothesised “dissimilar” and “similar” elements (white and black elements in Fig 3a, respectively). We then used permutation tests to ask whether this contrast was significantly positive across participants. We repeated this procedure for each

searchlight centre on the cortical surface, resulting in a cortical map of p-values.

The only cluster to survive multiple comparisons correction across a hemisphere was located focally in the right entorhinal cortex (Fig 3B and 3C, $P < 0.05$ FWE corrected on cluster level, cluster-forming threshold $P < 0.001$). This effect did not change when we repeated the analysis using model RDMs where same-stimuli or same stimuli set elements were ignored (data not shown). This suggests the effect was not driven by background color or low-level plasticity between stimuli that appear in the same block, but rather by an explicit representation of the relational structure between the stimuli in the task.

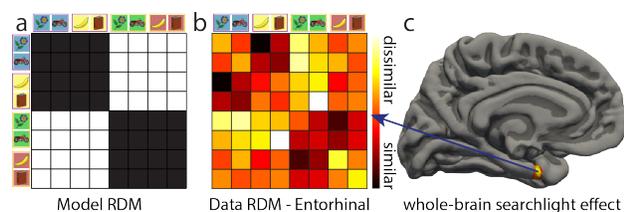


Figure 3: RSA effects of abstract structure

Discussion

Here, we show that the EC explicitly represents and generalises the relational structure of a non-spatial reinforcement learning task. This is the same area where grid cells, suggested to represent relational structure in spatial tasks, are found. Evidence of grid-like coding can also be found in non-spatial, 1D or 2D continuous tasks (Aronov, Nevers, & Tank, 2017; Constantinescu, O'Reilly, & Behrens, 2016), and the EC represents the statistical transition structure of a discrete state-space, even when participants are not aware of this structure (Garvert, Dolan, & Behrens, 2017). Taken together, our results suggest the same brain regions, perhaps with the same coding scheme, represent and generalise task structures in an explicit manner, across a wide variety of domains.

Acknowledgments

This work was supported by funding from a Wellcome Trust Senior Research Fellowship (WT104765MA) together with a James S. McDonnell Foundation Award (JSMF220020372) to T.E.J.B.

References

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature*, *543*, 719–722.
 Banino, A., Barry, C., Uribe, B., Blundell, C., Lillicrap, T.,

Mirowski, P., ... Kumaran, D. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature*, *557*, 429–433.
 Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018, October 24). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, Vol. 100, pp. 490–509. Cell Press.
 Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., Csicsvari, J., O'Neill, J., & Csicsvari, J. (2019). The entorhinal cognitive map is attracted to goals. *Science*, *363*, 1443–1447.
 Butler, W. N., Hardcastle, K., & Giocomo, L. M. (2019). Remembered reward locations restructure entorhinal spatial maps. *Science*, *363*, 1447–1452.
 Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science (New York, N.Y.)*, *352*, 1464–1468.
 Dordek, Y., Soudry, D., Meir, R., & Derdikman, D. (2016). Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *ELife*, *5*, 1–36.
 Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*, 190–194.
 Garvert, M. M., Dolan, R. J., & Behrens, T. E. J. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *ELife*, *6*, 1–20.
 Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.
 Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The Role of the Ventromedial Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans. *Journal of Neuroscience*, *26*, 8360–8367.
 Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*, 3863–3868.
 Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.
 Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2016). *The hippocampus as a predictive map*.
 Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.