Learning and use of reward-related representations across cortex and time

Ai Phuong S. Tong (aiphuong.s.tong@gmail.com)
Oxford Centre for Human Brain Activity
University of Oxford, Oxford, Oxfordshire OX3 7JX, UK

Vishnu Sreekumar (impromptu.pianist@gmail.com)
International Institute of Information Technology Hyderabad
Professor CR Rao Rd, Gachibowli, Hyderabad, Telangana 500032, India

Kareem A. Zaghloul (kareem.zaghloul@nih.gov)
National Institutes of Health,
10 Center Dr Bethesda, MD 20892, USA

Mark W. Woolrich (mark.woolrich@ohba.ox.ac.uk)
Oxford Centre for Human Brain Activity
University of Oxford, Oxford, Oxfordshire OX3 7JX, UK

Abstract:

Reward-related representations are found distributed throughout many human subcortical and neocortical regions that support different neural processes. These representations get used at different points in time for related tasks. However, the way these representations get re-used and strengthen over time is not well understood. To investigate this, we recorded from the temporal lobe and prefrontal cortex with intracranial electrocorticography (ECoG) while human subjects learnt two-choice decisions between two scenes. Subjects were able to straightforwardly re-use knowledge only when reward contingencies stayed the same between the two scenes. Using a Bayesian learner, we inferred reward expectations from choice behavior, and then measured representations of reward expectation in ECoG data. Reward expectations were uniquely represented in distributed regions across human cortex. The representations of reward expectation in the medial temporal lobe and orbitofrontal cortex were re-used between the two scenes, only when subjects could straightforwardly transfer knowledge between the two scenes. Finally, in a separate region, the anterior temporal lobe, the strength of reward representations as measured by similarity between scenes, increased as learning increased. Our findings suggest that patterns of activity representing reward information are integrated into multiple brain regions, get re-used in similar situations, and increase in fidelity with learning.

Keywords: reward expectation; reward representation similarity; reversal learning; electrocorticography

Introduction

Complex behaviors rely on structured representations of information (Niv, 2019). Representations in the prefrontal cortex have been shown to demonstrate this structure through the mixed selectivity of neurons that show adaptive coding and highly diverse responses that change over time (Fusi, Miller, & Rigotti, 2016; Rigotti et al., 2013). Importantly, the same representations can be reused in related situations. For example, the overlap in responses of neurons to variables have been shown to be common between experiences that are linked by time, item, or context (Cai et al., 2016; Zeithamova & Preston, 2010).

There is evidence that reward-related representations can be found in many distributed regions of subcortex and neocortex, including the orbitofrontal, ventromedial and dorsolateral prefrontal, cingulate, and parietal cortex (Elliott, 2000; Kahnt, Heinzle, Park, & Haynes, 2010; Rushworth & Behrens, 2008; Vickery, Chun, & Lee, 2011). While there is good evidence for representations of reward expectation found distributed across cortex, there is limited knowledge of when and how these representations get learnt and re-used.

Results

Here, we investigated the capacity of different regions to support reusable representations of reward expectation. To do this, we recorded simultaneous electrocorticography (ECoG) in temporal and prefrontal cortices while human subjects (n=11) performed a task in which they learnt two-choice decisions between two different scenes, corresponding to two different spatial

environments within which the choices were presented (Fig. 1a).

Modulation of Reward Expectation

In an environment where reward outcomes change, an agent dynamically learns to adapt the learning rate with each choice (Austerweil, 2015; Bartolo & Averbeck, 2020; Behrens, Woolrich, Walton, & Rushworth, 2007). Thus, we modeled the prediction of possible reward with a Bayesian framework where the reward expectation updates with each choice in proportion to the uncertainty in the belief of reward (Fig. 1b). We tracked learning at each trial using the absolute difference between the expectation for reward and the true reward probability (Fig. 1c,d). With each task over blocks, the reward expectation error progressively improved or decreased, indicating more efficient learning, when there was no reward reversal and not when there was a reward reversal (Fig. 2e; mixed effects across-subject t-test, rev, t(10) = .059, p = .954; no rev, t(10) = -2.589, p = .027).

Re-use of Representations in MTL and OFC

We hypothesized that when reward information is relevant to the task it should be actively represented in a way to influence choices and learning so we next examined whether the reward expectation is represented in the brain. We found that representations of reward expectation increased across multiple electrodes distributed over multiple brain regions within one second prior to when the choice is made (Fig. 2a).

We next hypothesized that only brain regions that represent reward expectation and integrated this with the representation of items would have similar representations of reward expectation when the knowledge of reward associated with the item can be transferred between scene 1 and scene 2. Across brain regions, the pre-choice representation of reward expectation, 1.0 sec prior to when the choice was made, in the medial temporal lobe (MTL) and orbitofrontal cortex (OFC) was significantly similar between scene 1 and scene 2 for the item set for which there was no reward reversal, but not for the item set for which there was a reward reversal (Fig. 2b).

Strengthening of Representations in ATL

We next investigated whether the similarity and strengthening, or increase in similarity, of the representations of reward expectation between scene 1 and scene 2 played a role in learning. To quantify learning in each block we measured how much reward expectation error decreased between scene 1 and scene 2. Next, we correlated this measure with the similarity in representations between scene 1 and scene 2 as it varied over blocks. A significant positive correlation was found in the ATL, such that as learning increases then so did the similarity between scene 1 and 2, but only for the item set for which there was no reversal in reward, or when knowledge can be straightforwardly shared between scenes, and not for the item set for which there was a reversal in reward (Fig. 2c,d).

Methods

Behavioral task Subjects were trained through a series of "blocks" with instructions about the task shown at the beginning. Human subjects were tasked to learn the most rewarding choices of two different "item sets", where each item set consists of two items that the subject must choose between and whose reward probabilities are coupled and sum to one. One item set corresponded to a building versus a face, and the other item set corresponded to a cutlery versus an animal (Fig. 1b). Subjects were tasked to learn the high reward probability item in two different scene contexts, a beach and a forest. In each block, either beach or forest was selected as the scene context for the first 60 trials, which we refer to as "context 1"; then the second 60 trials were carried in the other scene context, which we refer to as "context 2." Subjects were instructed that there was a reward reversal between the two different scenes in a block for one item set, but no reward reversal for the other.

Reward Expectation We applied Bayes's rule to compute the trial-by-trial *posterior P(r | h)*, the belief in the reward rate r after observing the history of choices and outcomes from the environment, h, from the product of the *prior* belief in the reward rate, *P (r)*, and the *likelihood* of the observed history of choices and outcomes had it been produced by the reward rate, *P(h | r).*

Representation similarity Representations can change dynamically over time in each context for each item set in each block. Thus, we performed a trial-wise regression of neural power on reward expectation separately for each context and item set in each experimental block to derive a [electrode x time] reward representation matrix, $W(k,i)$, from:

$$Y_t(k,i) \sim W(k,i) \cdot X_t + \beta(k,i) \cdot C_t,$$

where Y_t is the 30-80 Hz and 80-120 Hz neural power over 30 trials, t, for electrode, k, in a trial timepoint, i. X_t is the corresponding relative reward expectation over the 30 trials, t. C_t is the choice over 30 trials, t.

Figures

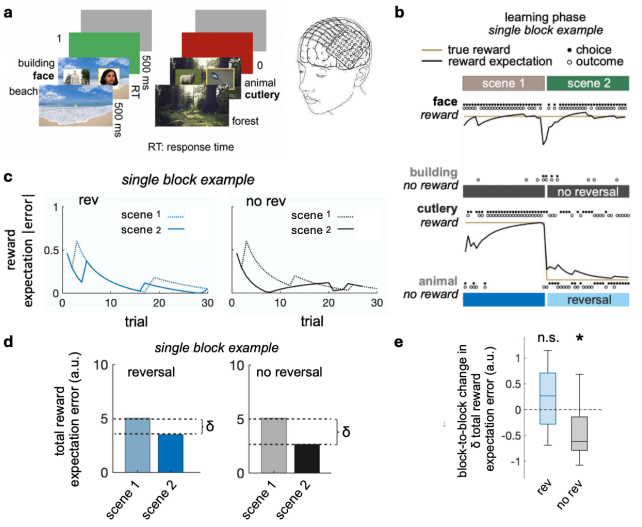

Figure 1: Reward expectations are modulated during reversal learning. **a**, Example item sets and scenes in two trials. **b**, Example of reward expectation inferred over trials in one block. **c**, Example of reward expectation error over trials. d, Calculation example of between-scene change, δ, in total reward expectation error. **e**, Learning measured by the block-to-block change in δ reward expectation error.

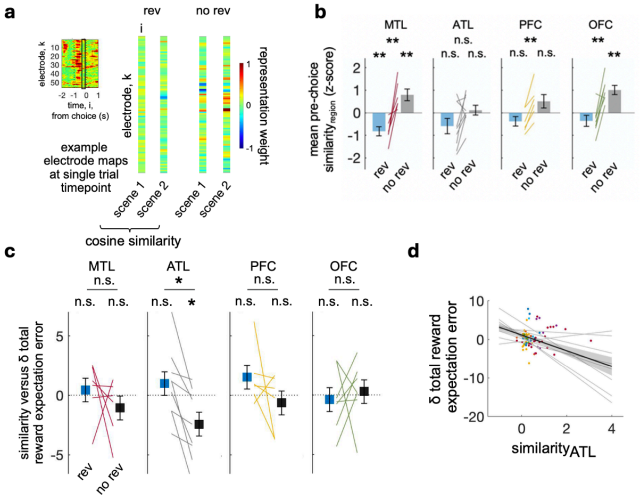

Figure 2: Reward expectation representations are used and strengthen with learning. **a**, Representation of reward expectation. **b**, Representation similarity across regions (MTL, medial temporal lobe; ATL, anterior temporal lobe; PFC, lateral prefrontal cortex; OFC, orbitofrontal cortex). **c**, Mean correlations between representation similarity and learning, as measured by δ total reward expectation error. **d**, Comparison of ATL similarity and learning.

Acknowledgments

We thank Laurence Hunt, Sanjay Manohar, Ben Seymour, Cam Higgins, Andrew Quinn, members of the Woolrich lab (Chet Gohil, Mats van Es) and members of the Zaghloul lab (Weizhen Xie, Julio Chapeton) for feedback and helpful discussions. This project was supported by funding from the National Institutes of Health Oxford-Cambridge Scholars Program (to A.P.S.T.).

References

Austerweil, J. L., Gershman, S.J., Griffiths, T.L. (2015). Chapter 9: Structure and Flexibility in Bayesian Models of Cognition. *The Oxford Handbook of Computational and Mathematical Psychology*, 187-208.

Bartolo, R., & Averbeck, B. B. (2020). Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron, 106*(6), 1044-1054 e1044.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci, 10*(9), 1214-1221.

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., . . . Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature, 534*(7605), 115-118.

Elliott, R., Friston, K.J., Dolan, R.J. (2000). Dissociable Neural Responses in Human Reward Systems. *Journal of Neuroscience, 20*, 6159–6165.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol, 37*, 66-74.

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J. D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proc Natl Acad Sci U S A, 107*(13), 6010-6015.

Niv, Y. (2019). Learning task-state representations. *Nat Neurosci, 22*(10), 1544-1553.

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature, 497*(7451), 585-590.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci, 11*(4), 389-397. doi:10.1038/nn2066

Vickery, T. J., Chun, M. M., & Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron, 72*(1), 166-177.

Zeithamova, D., & Preston, A. R. (2010). Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *J Neurosci, 30*(44), 14676-14684.