Brain Development of Selective Attention Enhances Credit Assignment

Wei Chen (chen-wei@g.ecc.u-tokyo.ac.jp)
Chinese Institute for Brain Research
Beijing, China
The University of Tokyo
Tokyo, Japan

Yukun Qu (yukun.qu@mail.bnu.edu.cn)
Beijing Normal University
Beijing, China

Xiaoyu Zhou (xyzhou.psy@gmail.com)
Beijing Normal University
Beijing, China

Yuko Yotsumoto (cyuko@g.ecc.u-tokyo.ac.jp)
The University of Tokyo
Tokyo, Japan

Yunzhe Liu (yunzhe.liu@bnu.edu.cn)
Chinese Institute for Brain Research, Beijing Normal University
Beijing, China

Abstract

A fundamental challenge of interacting with natural environment is to know what to learn. Previous studies have suggested that the learning of stimulus-outcome associations is related to the function of lateral orbitofrontal cortex (lOFC). The ability of credit assignment may be enhanced by lOFC through increasing the specificity of learning. The present study validates such idea by investigating the development of brain in relation to the change of cognitive processes in a reinforcement learning task. We designed a novel behavioral paradigm in which correctly assigning credit is critical for decision making. We quantitatively assessed participants' learning propensities using reinforcement learning models, and found that age-related credit assignment processes are influenced both by focusing attention to identify task-relevant features, and by shifting attention away from irrelevant features. Our structural MRI data showed that only the development of later is specifically mediated by the change of thickness in lOFC.

Keywords: development; lOFC; credit assignment; cognitive map;

Introduction

Interacting with the environment requires us to constantly search for causality from numerous environmental cues to guide our actions correctly (Jocham et al., 2016; Lake, Ullman, Tenenbaum, & Gershman, 2017; Santoro, Frankland, & Richards, 2016). Existing research has found that the ability to assign rewards to specific stimulus-outcome pairs is associated with the learning signal from lateral orbitofrontal cortex (lOFC) to hippocampus (Boorman, Rajendran, O'Reilly, & Behrens, 2016; M. Noonan et al., 2010), and that lesioning to this region impairs credit assignment (M. P. Noonan, Chau, Rushworth, & Fellows, 2017). Recently, lOFC has also been suggested to play a role in creating cognitive map by determining the precision of credit assignment (Costa et al., 2023; Radulescu, Niv, & Ballard, 2019). In the current study, we are interested in how this ability develops with age (Schlichting, Guarino, Roome, & Preston, 2022). We utilize reinforcement learning framework as a naturalistic trial-and-error learning mechanism to serve as our tool (Dayan & Niv, 2008; Sutton & Barto, 2018). It is particularly suited also because of its proven correlations with neurological findings of cognitive processes during development (Schultz, Dayan, & Montague, 1997; Dayan & Daw, 2008; Nussenbaum & Hartley, 2019).

Methods

We developed a cognitive task with the 2-alternative-forced-choice paradigm. In each trial, participants chose between two visual compounds, both marked by a same contextual symbol - either ice or fire (Figure 1a), guiding the participant's choice based on perceived ranking of the compounds within that context. The visual stimuli are composed of three feature categories. Two categories are randomly selected to be associated with ice and fire, respectively, while the third category is considered an 'irrelevant' category. Each context (ice or fire) will assign an independent ordinary scale to the corresponding category's features (Figure 1b). This means that, for example, if the category 'decoration' is paired with 'ice', then when the ice symbol is under the stimuli in a given trial, the magnitude of visual compound would depend on the rank of each decoration, unknown to participants.

Each participant would complete a 72-trial experiment ses-

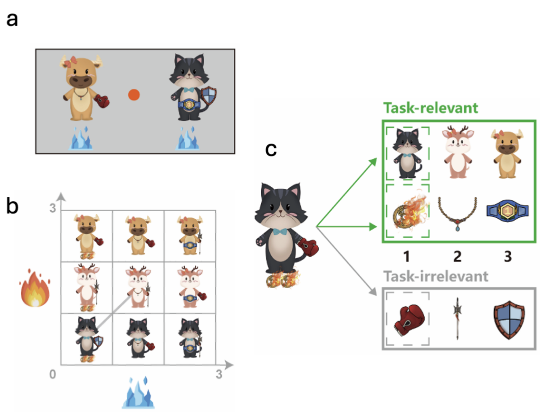

Figure 1: Sample task structure

sion, in the session, the participant would experience a pair of stimuli that contains adjacent ranked items in the category that the context is associated with. In other words, this is a pairwise learning task for each of the two task-relevant categories out of three. To achieve the best performance, participants were required to simultaneously learn the feature category associated with each context, as well as the ordinal rank of items within these feature categories to make informed selections.

To quantitatively measure the learning process, we adopted a formal reinforcement learning model based on successor representation, which decomposes value representation to expected discount future states and expected reward in each state. This model representation is particularly suited for tasks that balance habit and deliberate behaviour. We fitted our model into participants' choice data with learning-related free parameters. Participants with a performance below the chance level (16 out of 134) were excluded from all analyses. Our model captures a number of distinct cognitive processes: decision certainty, reflecting the use of expected value differences in decision-making; a dual-purpose learning rate, capturing both updates of reward-prediction error and updates in successor representation per trial; an attention parameter, quantifying focus on assigning credit to task-relevant features; and a confusion parameter, quantifying the blocking of irrelevant features during credit assignment - During learning, there are instances where two stimuli may contain identical item(s) in a feature category. For example, in a trial where the relevant dimension is ice, the stimuli may contain the same item from fire-related and/or irrelevant feature category (Figure 1c).

Results

Credit Assignment Develops with Age

We first collected behavioral data of 118 participants aged 8 to 18 and fitted the reinforcement learning model to each participant's data with a maximum likelihood estimate through gradient descent.

Our resulting model captures core cognitive processes in the task, based on temporal difference learning with successor representation updating:

$$\Delta \mathrm{SR}_{ij}^{(d)} = \delta_{ij} + \gamma \cdot \mathrm{SR}_{jn}^{(d)} - \mathrm{SR}_{ij}^{(d)}$$
$$\mathrm{SR}_{ij}^{(d)} \leftarrow \mathrm{SR}_{ij}^{(d)} + \alpha \cdot \Delta \mathrm{SR}_{ij}^{(d)}$$

where $\mathrm{SR}^{(d)}$ is the successor matrix for context d, α is the learning rate for the successor matrix, γ is the discount factor of SR updating, δ_{ij} is the Kronecker delta, indicating whether item i and j within a category are the same, and n is the index of the next possible item in the ranking states.

Given two contexts (ice and fire in our case), d_1 and d_2, the updates to the attention weights $w_{d,p}$ for feature categories p and q when current context is d_1 are as follows:

$$\begin{aligned} &\text{if } \exists p : m1_p = m2_p \text{ and } m1_q \neq m2_q \text{ for some } q, \\ &\text{in context } d_1, \quad w_{d_1,q} \leftarrow w_{d_1,q} + \text{Attention}, \\ &\text{in context } d_2, \quad w_{d_2,q} \leftarrow w_{d_2,q} + \text{Confusion}. \end{aligned}$$

Here, $m1$ and $m2$ indicates two stimuli in a trial, $w_{d,p}$ and $w_{d,q}$ represent the attention weight for feature category p and q in context d, correspondingly. Then we have the reward prediction updated for feature q in stimulus:

$$R_q \leftarrow R_q + \alpha \cdot (\text{observed reward} - R_q) \cdot w_{d,q}$$

Then we have Q-value calculation as:

$$Q(s,d) = \sum_q \left(\mathrm{SR}_{sq}^{(d)} \cdot R_q \cdot w_{d,q} \right)$$

where R_q is the predicted reward for feature q. The reward prediction itself is updated as:

And finally, the softmax decision rule is applied to convert learned value into choice probabilities:

$$P(a|s) = \frac{e^{\beta \cdot Q(s,a)}}{\sum_{a'} e^{\beta \cdot Q(s,a')}}$$

where β is the inverse temperature parameter that scales the sharpness of the decision policy.

We find that age-related changes in the learning mechanism are mainly characterized by bidirectional changes in attention (r = 0.29; p = 8.3e-3) and confusion (r = -0.21; p = 1.8e-2) parameters, instead of learning rate (r=0.14; p=0.13), which did not show significant correlations with age.

Development of lOFC Underpins Age-related Increase of Credit Assignment

To understand the brain's functionality change relates to the change of selective attention mechanism in our task. We performed structural MRI on 67 participants. We first performed a correlational analysis on 34 regions of the whole brain (averaged for the left and right hemispheres) based on Freesurfur's Destrieux Atlas (Destrieux, Fischl, Dale, & Halgren, 2010). After Bonferroni correction of multiple comparisons, we found that, among many brain regions that experienced a rapid change during development, participant's confusion parameter was only specifically correlated with the decreased thickness of lOFC ($r=0.39$; $p=6.5e-4$), while attention parameter was correlated with the decreased thickness of precuneus ($r=-0.41$; $p=3.2e-4$) and insula ($r=-0.42$; $p=2.2e-4$). Furthermore, through a mediation analysis (Figure 2), we found that, age's predictability of confusion can be fully mediated by the thickness change of lOFC (direct effect coef=-2.8e-2, p=0.49; indirect effect coef=-3.7e-2, p=8.4e-3). However, the mediation effect is not significant for attention-correlated brain regions (direct effect coef=0.26, p=0.06; indirect precuneus coef=0.11, p=0.24; indirect insula coef=6.4e-2, p=0.37).

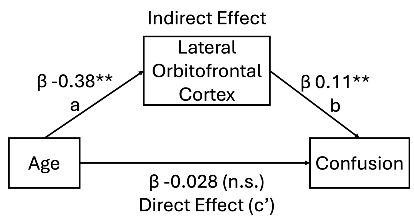

Figure 2: Mediation Analysis

Discussion

In conclusion, through a novel cognitive task, we demonstrated that the development of lOFC is fully mediates participants' improved ability of blocking irrelevant features in a credit assignment task. We postulate that the reduced thickness of lOFC may enhance credit assignment by synaptic pruning during the brain development. Our finding well supports the result of previous research in rodent indicating lOFC's role in defining the specificity of cognitive maps during learning (Costa et al., 2023).

However, we acknowledge that the current method is limited to pairwise learning paradigm and may be validated by an independent testing phase for the learned behavior, and differences between age can be better understand by a Maximum-A-Posterior estimation of group level parameters.

References

Boorman, E. D., Rajendran, V. G., O'Reilly, J. X., & Behrens, T. E. (2016). Two anatomically and computationally distinct learning signals predict changes to stimulus-outcome associations in hippocampus. *Neuron*, *89*(6), 1343–1354.

Costa, K. M., Scholz, R., Lloyd, K., Moreno-Castilla, P., Gardner, M. P., Dayan, P., & Schoenbaum, G. (2023). The role of the lateral orbitofrontal cortex in creating cognitive maps. *Nature neuroscience*, *26*(1), 107–115.

Dayan, P., & Daw, N. D. (2008, December). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429–453. doi: 10.3758/CABN.8.4.429

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*(2), 185-196. (Cognitive neuroscience) doi: https://doi.org/10.1016/j.conb.2008.08.003

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1-15. doi: https://doi.org/10.1016/j.neuroimage.2010.06.010

Jocham, G., Brodersen, K., Constantinescu, A., Kahn, M., Ianni, A. M., Walton, M., ... Behrens, T. (2016). Reward-guided learning with and without causal attribution. *Neuron*, *90*(1), 177-190. doi: https://doi.org/10.1016/j.neuron.2016.02.018

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*, e253.

Noonan, M., Walton, M., Behrens, T., Sallet, J., Buckley, M., & Rushworth, M. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, *107*(47), 20547–20552.

Noonan, M. P., Chau, B. K., Rushworth, M. F., & Fellows, L. K. (2017). Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *Journal of Neuroscience*, *37*(29), 7023–7035.

Nussenbaum, K., & Hartley, C. A. (2019, December). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, *40*, 100733. doi: 10.1016/j.dcn.2019.100733

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic reinforcement learning: the role of structure and attention. *Trends in cognitive sciences*, *23*(4), 278–292.

Santoro, A., Frankland, P. W., & Richards, B. A. (2016). Memory transformation enhances reinforcement learning in dynamic environments. *Journal of Neuroscience*, *36*(48), 12228–12242.

Schlichting, M. L., Guarino, K. F., Roome, H. E., & Preston, A. R. (2022). Developmental differences in memory reactivation relate to encoding and inference in the human brain. *Nature Human Behaviour*, *6*(3), 415–428.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.