Orbitofrontal cortex distributes reinforcement to the decision that caused it

Kay H Brodersen[1,2] · Laurence T Hunt[3] · Ekaterina I Lomakina[1,2] · Matthew F S Rushworth[3] · Timothy E J Behrens[3,4]

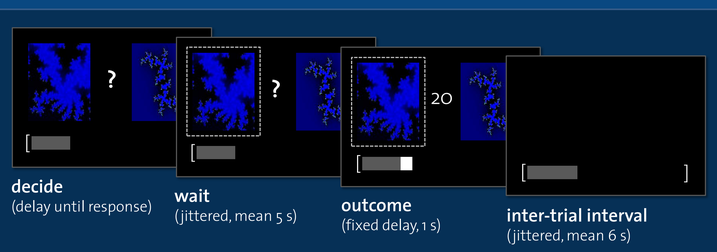
[1] Department of Computer Science, ETH Zurich, Switzerland [2] Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Switzerland
[3] Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Department of Clinical Neurology, John Radcliffe Hospital, University of Oxford, United Kingdom [4] Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom

1 Summary

- The theory of reinforcement learning proposes that reward-maximizing behaviour is based on the ability to associate an observed outcome with the decision that caused it.
- In a recent experiment, we demonstrated that a lesion to the lateral orbitofrontal cortex (LOFC) keeps reward processing intact but is fatal to the ability to correctly associate rewards with preceding decisions.
- Using fMRI in humans, we set out to explain this effect by examining the role of the LOFC in distributing reinforcement to the decision that caused it.
- We found that LOFC (as well as ventral striatum and right anterior prefrontal cortex) displays activity that does not simply code for rewards or reward prediction errors but critically distinguishes the contingent choices that caused those rewards.

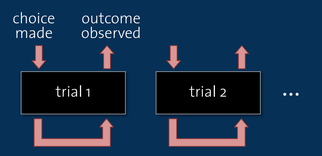
2 Experimental design

In a novel probabilistic decision-making paradigm, 24 healthy participants had to learn, by trial and error, the reward probabilities of two options, while undergoing 3T fMRI.

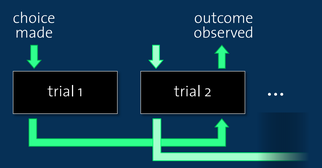


On each trial, participants had to choose between two options, which had different probabilities of leading to a reward. The same two options were presented on each trial within a block.

The experiment comprised 120 trials, split up into 8 blocks. At the beginning of each block, subjects were given one of the following types of instruction:



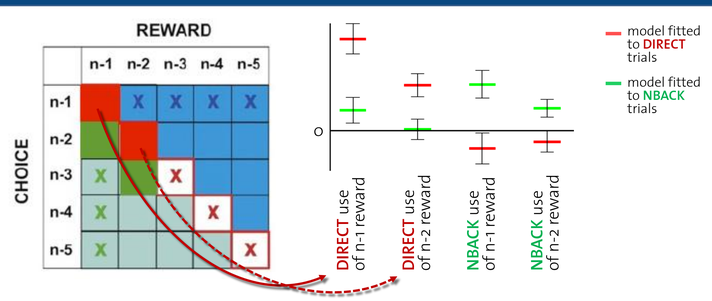
In a DIRECT block, whenever a rewarded option was chosen, its reward was presented on the *same* trial.



In an NBACK block, rewards were delayed by one trial such that a reward had to be linked to the choice made on the *previous* trial.

3 Behavioural validation of the design

As intended, decisions in DIRECT and NBACK trials could be best explained by decision-making models that linked each reward to the choice made on the *same* or *previous* trial, respectively.



Left | A simple computational model explains current choices as a function of previous choices and rewards. We considered choices and rewards of up to three trials into the past (corresponding to the four red and green filled cells).

Right | Model parameters were estimated separately for the two conditions. On DIRECT trials, associations between the previous two choices and their respective rewards best explained current decision-making behaviour. Conversely, on NBACK trials, the strongest influence on current behaviour originated from previous choices being associated with rewards that had been observed on the preceding trials.

4 Imaging analysis

To examine how different contingencies might be supported neuronally, we set up a GLM to explain BOLD activity in the outcome phase of each trial. We used separate regressors for different trial types.

Critical for our analysis were those trials on which subjects had switched options between the previous and the current trial, as illustrated below.

The choice that follows a switch reveals which rule is being used

In this example, a subject chose B on the first trial (which was rewarded or unrewarded). They then switched to A on the second trial, which led to a reward. What choice should be made on the third trial?

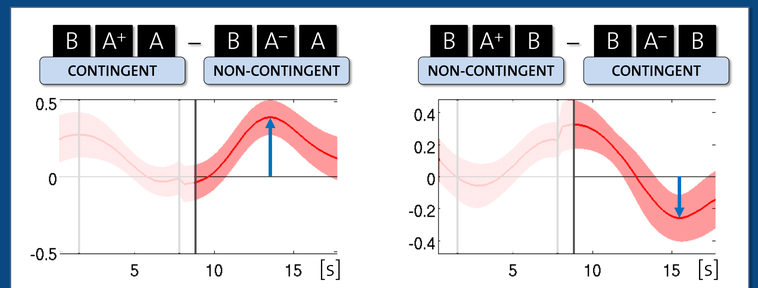
DIRECT block: the reward after choosing A should be associated with the *current* choice, i.e., A should be chosen again.

NBACK block: the reward after A should be associated with the *previous* choice, i.e., B should be chosen next.

Examining the choice on the third trial allowed us to distinguish between those rewards that had correctly influenced behaviour and those that had not. This enabled us to compare contingent and non-contingent LOFC activity.

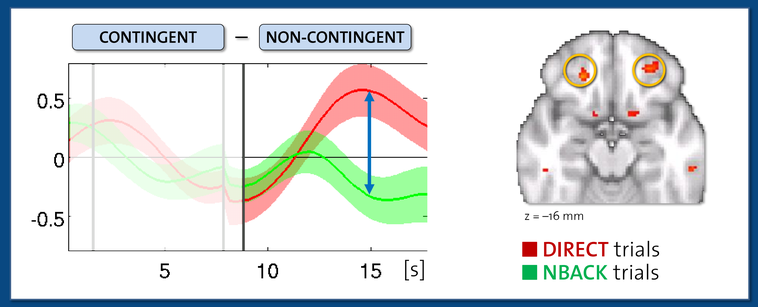
5 Imaging results

The two contrasts below show that LOFC activity neither simply coded for rewards nor for reward prediction errors.



Temporal evolution of GLM contrasts over intra-trial time (parameter estimates +/– contrast variance).

Instead, LOFC activity increased specifically whenever the correct contingency was being used, i.e., when an outcome was associated with the correct choice, whether in the DIRECT or the NBACK condition.



6 Conclusions

- As shown previously, a lesion to the LOFC impairs association learning in monkeys. We set out to expain this effect using fMRI in humans.
- We found that LOFC is neither simply driven by rewards and losses, nor by reward prediction errors, nor by switches and stays.
- Rather, activity in LOFC (as well as VS and right anterior PFC) specifically encodes whether the correct contingency is being applied. Contingencies are indicated irrespectively of whether reinforcement should be distributed to the *current* choice (DIRECT condition) or to the *previous* choice (NBACK condition).

References

1. Rushworth, M.F.S. & Behrens, T.E., 2008. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, 11(4). 389-97.
2. Schultz, W., 2006. Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87-115.
3. Walton, M.E. et al., 2010. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*, 65(6), pp.927-939.