Linking neural representations for decision-making between monkey and human cortex

Paula Kaanders (paula.kaanders@wadham.ox.ac.uk)
Dept. of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road
Oxford, OX1 3QX, United Kingdom

Hamed Nili (hamed.nili@psy.ox.ac.uk)
Dept. of Experimental Psychology, Walton Street
Oxford, OX2 6BW, United Kingdom

Tim Behrens (behrens@fmrib.ox.ac.uk)
Wellcome Centre for Integrative Neuroimaging, FMRIB, John Radcliffe Hospital
Oxford, OX3 9DU, United Kingdom

Laurence Hunt (laurence.hunt@psych.ox.ac.uk)
Oxford Centre for Human Brain Activity, Warneford Hospital
Oxford, OX3 7JX, United Kingdom

Abstract

Neuroscience has struggled to link animal models to human brain functioning. Hunt et al. (2017) identified a triple functional dissociation in prefrontal cortex (PFC) using single-cell recordings from macaque during an information gathering task. We attempt to find a similar dissociation in the human brain using functional magnetic resonance imaging (fMRI) during a similar task using two approaches: mass univariate analysis and representational similarity analysis (RSA). In a mass univariate analysis, we find evidence of a belief confirmation signal in ACC, consistent with that identified in single cell recordings in the same task. Using RSA, we successfully produced clear representational geometries in primary visual cortex and the fusiform gyrus for spatial location and cue type (face/house) respectively. However, we found no clear relationships between RSA matrices in anatomical regions of interest for dlPFC, OFC or ACC, in contrast to what was found in the macaque data. These findings do not completely rule out RSA as a means of mapping animal and human data to a common space in PFC, as there is still much space for further exploration of the data.

Keywords: decision-making; value-guided choice; attention; prefrontal cortex; representational similarity analysis

Introduction

Animal models have great potential in helping us understand brain function. The monkey brain has proven particularly useful to model those functions unique to primates, such as complex cognitive processes like problem-solving and decision-making. There is much evidence that PFC is similarly structured in primate and human brains (Price, 2010). Neubert et al. (2015) were able to link the majority of a selection of human PFC subregions involved in decision-making to homologous regions in macaque using resting-state fMRI and diffusion-weighted MRI. Single-cell recordings in macaque brains have been invaluable in starting to understand the role of subregions of PFC in value-based choice (Padoa-Schioppa and Assad, 2006; Kennerley et al., 2009).

In a recent paper by Hunt et al. (2017), a triple functional dissociation is established in an information gathering task using single-cell recordings from prefrontal areas in macaque. However, how to generalize such electrophysiological recordings from animals to be able to draw conclusions about human brains is unclear.

Hunt et al. found specific neural signatures in three prefrontal subregions in macaque: orbitofrontal (OFC), anterior cingulate (ACC), and dorsolateral prefrontal cortex (dlPFC; Fig. 1).

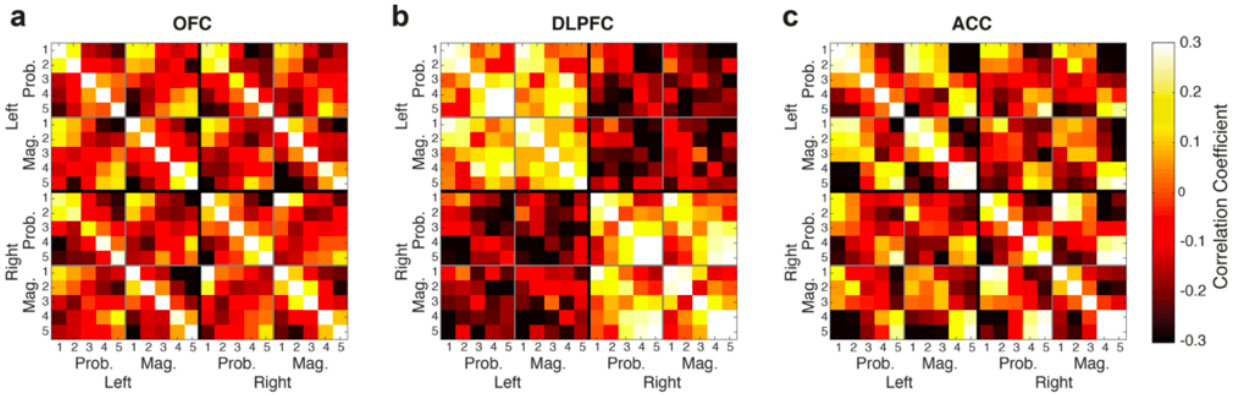

Figure 1: (A-C) Representational similarity matrices of the 20 conditions in an information gathering task. The conditions are ten stimuli presented either on the left or right side of the screen. The color of each pixel represents the correlation between a population of neurons' firing rates in that specific brain region in response to the two conditions being compared. Reproduced from Hunt et al. (2017).

Activity in OFC was found to be related to the value of the stimulus when only one stimulus had been attended to. When the values of more than one stimulus within a trial became known to the participant, OFC activity started reflecting an attention-guided value comparison between the available options instead. dlPFC activity was shown to contain a representation of the current locus of spatial attention, as it was mostly modulated by the location on the screen of the cue the participant was attending to. Activity in ACC encoded confirmation or disconfirmation of the belief established by the first presented cue when subsequent cues were attended.

In this paper, we try to replicate Hunt et al.'s findings in human functional magnetic resonance imaging (fMRI) data using a similar paradigm. To this end we use two approaches: mass univariate analysis and representational similarity analysis (RSA).

In a mass univariate analysis, we find that activity in the anterior rostral cingulate zone (RCZ) of the ACC, as defined by Neubert et al. (2015), correlates negatively with belief confirmation similarly to that found in the macaque paper. Our initial RSA did not produce similar representations to those seen in Figure 1.

Methods

14 healthy human participants (aged 18-50) attended two study sessions: one behavioral training session (1hr) and one fMRI session (2hr 15min). In the behavioral session, participants learned the meanings of ten stimuli: five faces and five houses.

Five of these stimuli represented a reward probability (10%, 30%, 50%, 70%, or 90%) and the other five a reward magnitude (10, 30, 50, 70, or 90 points).

In the main task performed in the scanner, participants are asked to choose between two pairs of cues (Fig. 2). Each pair consists of a probability and a magnitude cue. The reward associated with that option is the number of points represented by

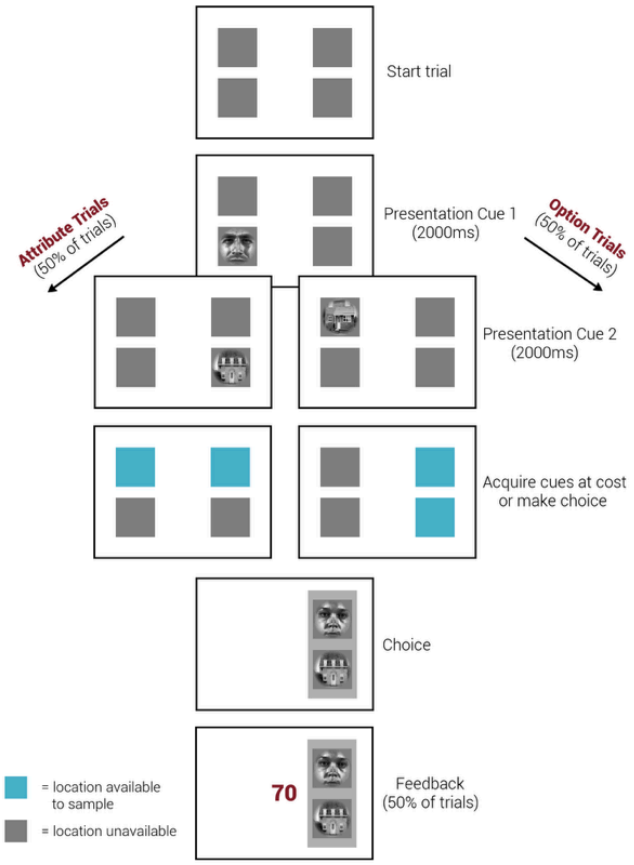

Figure 2: Main Task Paradigm.

the magnitude cue awarded probabilistically in accordance with the probability cue. 'Optimal' behavior in the task (maximizing long-run expected reward) would be to choose the side with the higher expected value (reward probability multiplied by reward magnitude). However, a trial starts with the four cues being hidden, where the pair of cues on the left is one option and the pair on the right another. Participants are initially shown two different cues sequentially. These can either be two cues from the same option ('option trials', 50% of trials) or two cues from different options representing the same attribute (probability or magnitude; 'attribute trials', 50% of trials). After this, participants get the opportunity to view the remaining two hidden cues at a cost by pressing the corresponding buttons. When participants decided to stop sampling the additional cues, they made a choice between the two options.

Whole-brain fMRI measurements were made using a Siemens Prisma 3T scanner with a 2 x 2 x 2mm voxel size, repetition time (TR) = 1.235s, echo time (TE) = 20ms, flip angle = 65° with an axial orientation angled to AC-PC using a 64-channel head coil. The sequence used was MB3 PAT2. The average number of volumes collected per participant was 2587. T1-weighted structural images were obtained using an MPRAGE sequence with 1 x 1 x 1mm voxel size, on a 174x192x192 grid, TE = 3.97ms, TR = 1.9s.

For RSA we used the human neuroanatomical homologues of the regions studied in the macaque paper. We first estimated a whole-brain GLM with regressors describing cue onset as well cue identity. After projecting the results from all runs into MNI space, we then extracted the parameter estimates for all voxels within each mask for each area of interest (containing the areas on both hemispheres). This produces a [voxels*conditions] matrix of parameter estimates for each mask, which can then be used to compute the representational similarity between conditions.

For mass-univariate analyses, we used a similar set of regressors to those used in Hunt et al. (2017), At the time of cue 1 and cue 2 presentation, this included several regressors relating to the value of currently and previously attended options. Perhaps most importantly, however, it also included a regressor that is orthogonal to value, reflecting whether the second attended cue *confirms* or *disconfirms* the belief set up at the first cue for which option should be chosen (see Hunt et al., 2017, for full details).

Results

We found deactivation of an area inside dorsal ACC when the second cue *confirms the belief formed by the first cue* (Fig. 3; area of peak deactivation: max Z = 4.5, MNI x = 8, y = 30, z = 32 mm). In Neubert's atlas for cingulate and orbitofrontal cortex, this area is classified as part of the anterior rostral cingulate zone of ACC (Neubert et al., 2015). Importantly, in option trials the second cue is considered confirmatory if it is of a similar value to the first cue. In contrast, in attribute trials the second cue is considered confirmatory if it is of a very different value from the first cue.

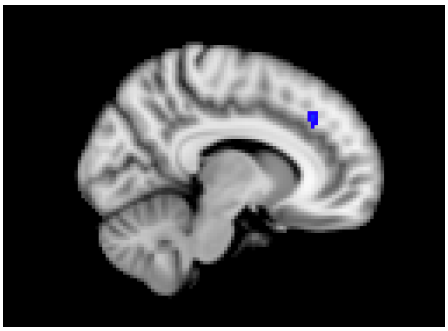

Figure 3: ACC deactivation in response to *belief confirmation* in option trials (thresholded at Z > 3.5 for display purposes). MNI x = 10mm.

In Figure 4 it can be seen that RSA successfully discriminated between basic properties of sensory processing – for example, a mask placed over the calcacarine sulcus clearly distinguishes left vs. right attended items, and a mask placed over the fusiform gyrus clearly distinguishes currently attended faces vs. houses. However, when we examined similar matrices in predefined anatomical ROIs of OFC, DLPFC and ACC, no clear pattern could be distinguished that closely matched what we had observed in RSA on monkey single cell data in these regions.

Discussion

The ACC deactivation we observe in response to cues confirming the belief formed in response to the first cue is in line with many recent accounts describing ACC in terms of value comparison between a current best 'default' option, versus exploring other alternatives in the environment (Daw et al. 2006; Boorman et al., 2013). Whereas most studies have considered this belief updating *across* trials we here see a signal consistent with belief updating *within* a trial, as a decision is being formed. We observe deactivation in response to confirmation of the belief formed from the first cue presentation in all trial types. This suggests ACC cannot simply be encoding value comparison here,

as in the option trials the observed cues belong to the same option and so there is nothing yet to compare against. As such, the decision may not really be framed in terms of direct comparison between the two alternatives, but instead in terms of whether to take or leave the current 'default' option.

So far we have not obtained clarity in the RSA similar to that in the macaque study for our three PFC regions of interest: OFC, dlPFC and dACC. The OFC RSA was weakly significantly correlated with attended value, even though visually this relationship could not be distinguished in the matrix. We did not find any meaningful correlations between the dlPFC and ACC RSA matrices and any of the templates used. There are several possible reasons for this. One possibility is that the regions studied in the macaque brain are not anatomically homologous to the regions of interest in the human brain; we next intend to use searchlight RSA to explore direct mapping between regions that are homologous to those identified in the macaque study. An alternative explanation to practical concerns is that the scale of information encoding in PFC can inherently not be probed by RSA. For example, it remains debated whether RSA really probes fine-grained spatial information that is closer to that identified by single neuron data, or whether it is primarily driven by more macroscopic signals that can be isolated in mass univariate analyses. It may be that the intermixed positive and negative coding that supported the successful RSA in Hunt et al. (2017) is not observable at the voxel level in human fMRI.

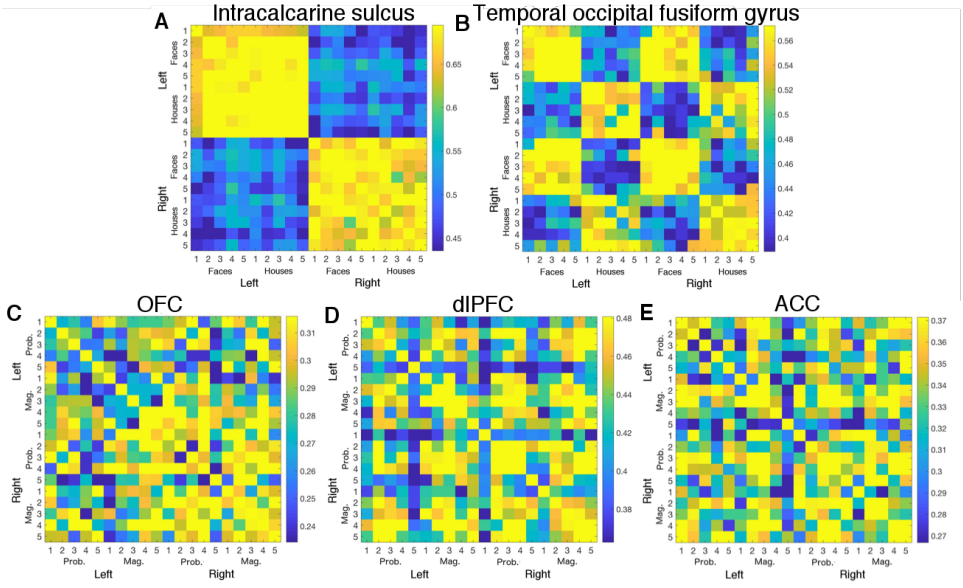

Figure 4: Representational similarity analysis in (A) intracalcarine sulcus, (B) temporal occipital fusiform gyrus, (C) OFC, (D) dlPFC, and (E) ACC between the 10 possible stimulus identities and which side they were presented on. The color of each pixel in the matrix refers to the correlation r between activity in the two conditions across all voxels in the mask.

References

Boorman, E.D., Rushworth, M.F., & Behrens, T.E. (2013). Ventromedial Prefrontal and Anterior Cingulate Cortex Adopt Choice and Default Reference Frames during Sequential Multi-Alternative Choice. *J Neurosci, 33*(6), 2242-2253.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., & Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876-9.

Hunt, L.T., Malalasekera, W.M.N., de Berker, A.O., Miranda, B., Farmer, S.F., Behrens, T.E.J., & Kennerley, S.W. (2017). Triple dissociation of attention and decision computations across prefrontal cortex. *bioRxiv preprint.* https://doi.org/10.1101/171173

Kennerley, S.W., Dahmubed, A.F., Lara, A.H., & Wallis, J.D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *J Cogn Neurosci, 21*(6), 1162-1178.

Neubert, F., Mars, R.B., Sallet, J., & Rushworth, M.F.S. (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision-making in human and monkey frontal cortex. *PNAS, 112*(2), e2695-2704.

Padoa-Schioppa, C., & Assad, J.A. (2006). Neurons in orbitofrontal cortex encode economic value. *Nature, 441*, 223-226.

Price, J.L. (2010). Architectonic structure of the orbital and medial prefrontal cortex. *The Orbitofrontal Cortex,* 1-17.