

The hippocampal formation facilitates social decision-making by transforming reference frames.

Raphael Kaplan (raphael.kaplan.09@ucl.ac.uk)

UCL Wellcome Centre for Human Neuroimaging, 12 Queen Square
London, United Kingdom WC1N 3BG

Karl Friston (k.friston@ucl.ac.uk)

UCL Wellcome Centre for Human Neuroimaging, 12 Queen Square
London, United Kingdom WC1N 3BG

Abstract:

Knowing how other people's preferences relate to our own is a central aspect of social cognition, yet how the brain performs this perspective taking is unclear. Here, we ask whether the putative role of the hippocampal formation in transforming first person and extra-personal spatial cues during navigation extends to social learning. In our experiment, subjects learn a stranger's preference for an everyday activity – relative to a personally familiar individual – and subsequently decide how the stranger's preference relates to further familiar people. We observed hippocampal responses during decisions that require precise social judgments. Investigating reference frame sensitive responses, we identified an entorhinal/subicular region responding to finely resolved decisions involving self-comparisons, but more straightforward choices otherwise. Our data highlight a potential hippocampal-entorhinal division of labor that helps assimilate newly learned information about others into our prior beliefs about the prosocial world.

Keywords: hippocampus, social cognition, learning, knowledge

Introduction

Social decision-making is facilitated by expressing personal preferences *ordinally*—whether we prefer one thing to another—and *metrically*—how much more we prefer one thing to another. Ordinal and metric coding are particularly important when learning about new people, which involves relating a new person's personal attributes to prior knowledge of other individuals, either by adapting an egocentric or extra-personal frame of reference. On one hand, progress has been made in linking the hippocampus with maintaining an ordinal sequence or 'hierarchy' of social attributes (Tavares et al., 2015, Kumaran et al., 2016). Yet, the neural representation of metrically coded social knowledge remains elusive, even though metric coding affords the

transformation of knowledge learned via egocentric and extra-personal frames of reference.

Clues about neural computations underlying the transformation of knowledge may come from research on the role of the hippocampal formation in path integration—the process of calculating one's position by estimating the direction and distance one has travelled from a known point. During path integration, specific sub-regions of the hippocampal formation are associated with integrating environmental and first person representations of space in order to reach a desired location (McNaughton et al., 2006). In particular, grid cells in entorhinal/subicular areas are selectively active at multiple spatial scales when an animal enters a set of periodic triangular locations covering the entire environment (Hafting et al., 2005;), while hippocampal place cells code specific locations in an environment (O'Keefe & Dostrovsky, 1971). Working together with boundary vector cells in entorhinal/subicular areas that code the presence of an environmental boundary at a particular direction and distance (O'Keefe & Burgess, 1996), these spatially-modulated neurons in the hippocampal formation are thought to collectively serve as a cognitive map of the environment (O'Keefe & Nadel, 1978). Notably, recent findings have extended the idea of map-like coding in the entorhinal cortex and subiculum to humans. Human entorhinal/subicular regions respond to both the distance of goal locations (Chadwick et al., 2015) and discrete abstract relations (Garvert et al., 2017), suggesting that entorhinal/subicular areas might be able to represent personal knowledge along both continuous and discrete dimensions. Taken together, these results suggest that putative neural computations in the hippocampal formation related to spatial exploration, may also help facilitate the integration of information learned in different reference frames during social decision-making (Schiller et al., 2015; Epstein et al., 2017; Kaplan et al., 2017a).

We investigated whether specific brain regions, including sub-regions of the hippocampal formation, like the entorhinal/subicular area, facilitate switching between egocentric and extra-personal reference frames during social decision-making.

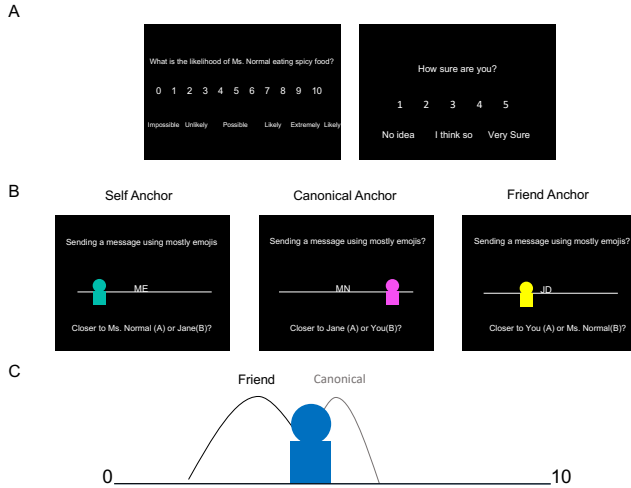


Figure 1 – *Experiment.* A. Prior to fMRI scanning, subjects were instructed to choose a friend with a different personality of the same gender. Subsequently, subjects rated from 1-9, on a 0-10 scale, how likely they (self), a close friend (friend), and the typical person (canonical) were to partake in a variety of everyday scenarios (e.g., eat spicy food, read a book, cycle to work). B. fMRI paradigm. During a forced-choice social decision-making task, subjects made a decision on the relative proximity of a stranger’s rating for an everyday scenario to their ratings of the self, friend, and canonical individuals. On each self-paced trial (max. allowed response time 9s), subjects viewed a personal preference for a new stranger that was presented on a 0-10 number line relative to one of the known individuals (anchor) and subjects had to decide whether the stranger was closer to one the two remaining familiar individuals. Crucially, the anchor was always placed in the middle of the scale, to ensure that subjects used prior social knowledge to infer the stranger’s absolute preference. C. Illustration of the behavioral model. We quantified the difficulty of discriminating a particular choice by fitting a formal model, based on the relative distance between the two choice individuals on the scale and how confident subjects were their ratings (e.g., comparing the stranger rating, represented by the blue avatar, with their rating for their friend and the canonical individual). Subjective confidence was represented by the standard deviation for each rating, where lower confidence entails higher standard deviations.

Methods

To test this hypothesis, we developed a novel experimental task, where healthy volunteers were first asked to rate on a scale of 1-9, how likely they (self), a close friend (friend), and the typical person (canonical) were to partake in a variety of everyday scenarios (e.g., eat spicy food, read a book, cycle to work; Fig. 1). Subsequently, subjects performed forced-choice fMRI task (Fig. 1), where they made a decision on the proximity of a stranger’s rating relative to their ratings of the self, friend, and canonical individuals for a given everyday scenario (Fig. 1). Specifically, on each self-paced trial, subjects viewed a personal preference for a new individual that was presented on a 0-10 scale relative to one of the individuals (*anchor individual*) and subjects had to decide whether the stranger was closer to whichever of the two remaining personally familiar individuals weren’t the anchor. Crucially, the anchor was always placed in the middle of the scale, to ensure that subjects used prior social knowledge in order to infer the novel individual’s absolute preference and form an approximate mental number line of the different individuals’ personal preferences (i.e., remembering the relative distance of the different preferences on the number line; see Fig. 1C). In other words, subjects had to infer the stranger’s true position in relation to the anchor’s rating for that scenario and the closest boundary (0 or 10). Note that this is a non-trivial task because the preferences of the *novel* individual were not conserved over attributes.

In sum, training subjects to think of personal preferences in the form of a mental number line allowed us to probe different scenarios at various levels of choice discriminability. Choice discriminability was determined by the relative distance between the individuals on the scale and how confident subjects were for a particular preference. Relating our paradigm to coordinate transforms, we tested whether a stranger’s personal preferences are represented differently in the brain in two ways. First, relative to which person the stranger’s personal preference is initially learned, and second, to whom the stranger’s preference is being compared.

We quantified the difficulty of discriminating a particular choice by fitting a formal model based on the relative distance between the two choice individuals on the scale and how confident subjects were for each rating. Specifically, we characterized choice discriminability using the entropy of choice probabilities based on a softmax function of likelihood and confidence ratings (Fig. 1C). High entropy corresponds to lower choice discriminability induced by similar ratings. To estimate the requisite softmax (sensitivity or precision) parameter, we modelled performance in

terms of entropy (H) over trials (and subjects) using a simple linear regression model. The ensuing behavioral model provided an estimate of the precision parameter (B) and associated measure of trial-specific choice discriminability for each subject. This implicit computational model of choice provided trial-specific measures of choice uncertainty (H) that enabled us to identify its fMRI correlates. We used the mean precision parameter (B), over subjects to compute trial-specific choice entropies as a predictor for our fMRI data. As expected, mean choice entropy/discriminability values correlated with performance over trials ($t(23)=9.01; P<.001$).

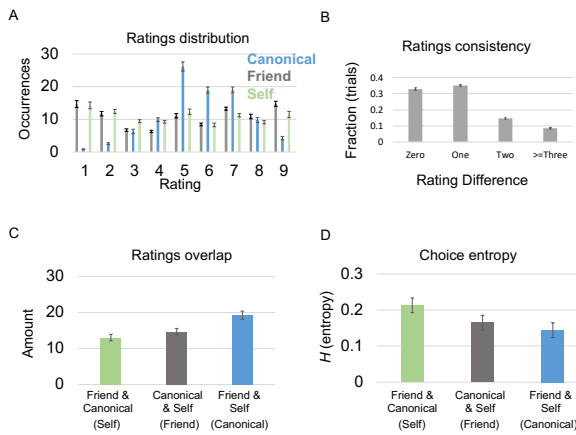


Figure 2 – Behavioral Results *A.* Mean ratings across subjects for each familiar individual and every scenario *B.* Rating Consistency. Difference in rating pre- and post-fMRI scanning for *friend* and *canonical* exemplar ratings. *C.* Ratings and Overlap. Significant main effect of condition for ratings overlap between the two individuals ($p<0.001$). Individuals being compared are listed below each bar with the corresponding anchor/condition name listed in parentheses. *D.* Choice entropy: Significant main effect of condition for choice entropy ($p<0.001$). Individuals being compared listed below each bar, with corresponding anchor (condition name) listed in parentheses.

Results

Using fMRI, behavioral modelling (Fig. 2, and a novel social decision-making paradigm, we examined how different brain regions integrate social knowledge learned via first person and extra-personal reference frames (Fig. 1). We observed hippocampal and retrosplenial cortex (RSc) signals related to successfully disambiguating fine-grained choices involving any frame of reference (Fig. 3). Highlighting reference frame sensitive responses specific to the hippocampal formation, we isolated an

entorhinal/subicular region responding to fine-grained choices involving self-comparisons, but more discretized choices otherwise. Notably, both subiculum/entorhinal and RSc signal increases were higher prior to accurate choices. In parallel, superior parietal lobule activity increased when the anchor required mental shifting from the middle of the scale towards the periphery. Lastly, we find that striatal responses also precede accurate choices, which were partially driven by decisions involving which individual is highest or lowest, instead of proximity.

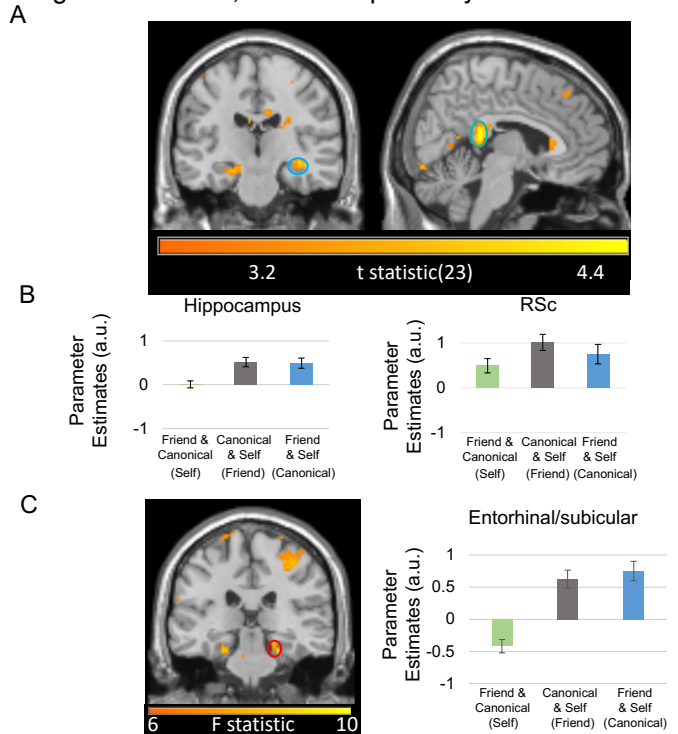


Figure 3 – Choice Discrimination Effects. *A.* Regions significantly responding to high choice entropy. Left: Coronal image showing right hippocampus circled in light blue. A portion of the left subicular cluster is also visible. Right: Sagittal image showing retrosplenial cortex (RSc) circled in turquoise. *B.* Effect size for a 10-mm sphere around the right hippocampal and RSc peaks (mean \pm SEM). A positive effect size indicates a positive BOLD correlation with choice entropy (i.e., more ambiguous choices), whereas a negative effect size indicates a negative BOLD correlation with choice entropy (i.e., straightforward choices). Individuals being compared listed below each bar, with corresponding anchor individual (condition name) provided in parentheses. *C.* Coronal image of right entorhinal/subicular region exhibiting main effect of choice entropy by condition circled in red. Portion of left entorhinal/subicular region showing same effect is also visible. *D.* Effect size for a 10-mm sphere around right entorhinal/subicular region (mean \pm SEM). All

highlighted regions survived FWE correction for multiple comparisons at $p < 0.05$ and are displayed at an uncorrected statistical threshold of $p < .005$ for display purposes.

Discussion

We observed entorhinal/subicular responses to more fine-grained choices, if self-comparisons were involved, but discretized choices involving their friend and the canonical exemplar. This result is partially a consequence of subjects' lower confidence about the friend and canonical exemplar ratings, which induced significantly higher choice entropy for that condition (Fig. 2). Despite the difference between conditions in mean choice entropy, a similar distinction was not observed in the hippocampal body, which only related to ambiguous choices that involved comparing self versus canonical/friend preferences. Notably, in rodents, hippocampal place representations are typically self-referenced and continuous. In contrast, entorhinal/subicular grid and boundary vector representations use multiple reference frames, and can either be continuously or discretely coded (Hartley et al., 2013). Our data extend this functional dissociation to social cognition, where the hippocampus related to fine-grained, self-referenced knowledge, while the entorhinal/subicular region related to knowledge at multiple levels of detail for both self and extra-personal reference frames.

Conclusion

Metric coding of decision variables informs social decisions by providing coordinates and boundaries that can be translated between different frames of reference. We provide evidence that neural computations related to integrating first person and global coordinates during spatial navigation, also extend to relating others' personal attributes to our own. Consequently, these data provide important clues about how hippocampal-entorhinal map-like coding may facilitate everyday decision-making in a domain general manner.

Acknowledgments

RK is funded by a Sir Henry Wellcome Fellowship (Ref: 101261/Z/13/Z).

References

- Chadwick, M.J., Jolly, A.E., Amos, D.P., Hassabis, D., Spiers, H.J. (2015). A goal direction signal in the human entorhinal/subicular region. *Curr Biol*, 25:87-92.
- Epstein, R.A., Patai, E.Z., Julian, J.B., Spiers, H.J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature Neurosci*, 20:1504-13.
- Garvert, M.M., Dolan, R.J., Behrens, T.E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *Elife*, 6.
- Hafting, T., Fyhn, M., Molden, S., Moser, M.B., Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436:801-6.
- Hartley, T., Lever, C., Burgess, N., O'Keefe, J. (2013). Space in the brain: how the hippocampal formation supports spatial cognition. *Philos Trans R Soc Lond B*, 369:20120510.
- Kaplan, R., Schuck N.W., Doeller, C.F. (2017a). The Role of Mental Maps in Decision-Making. *Trends Neurosci*, 40:256-9.
- Kaplan, R., King, J., Koster, R., Penny, W.D., Burgess, N., Friston, K.J. (2017b). The Neural Representation of Prospective Choice during Spatial Planning and Decisions. *PLoS Biology*, 15: e1002588.
- Kumaran, D., Banino, A., Blundell, C., Hassabis, D., Dayan, P. (2016). Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and Representing Self-Relevant Information. *Neuron*, 92:1135-47.
- McNaughton, B.L., Battaglia, F.P., Jensen, O., Moser, E.I., Moser, M.B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci*, 7:663-78.
- O'Keefe, J., Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Res*, 34:171-5.
- O'Keefe, J., Nadel, L. *The Hippocampus as a Cognitive Map*. 1978:114-52 (Oxford Univ Press)
- O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381:425-28.
- Schiller, D., Eichenbaum, H., Buffalo, E.A., Davachi, L., Foster, D.J., Leutgeb, S., Ranganath, C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map. *J Neurosci*, 35:13904-11.
- Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., Schiller, D., (2015). A Map for Social Navigation in the Human Brain. *Neuron*, 87:231-43.