NUFFIELD DEPARTMENT OF
CLINICAL NEUROSCIENCES

UNIVERSITY OF
OXFORD

FMRIB, John Radcliffe Hospital, Oxford, OX3 9DU
Web: www.win.ox.ac.uk | **Tel**: +44(0)1865 611450 | **Email**: enquiries@fmrib.ox.ac.uk
Prof Tim Behrens | **Tel** +441865 610473 | **Email** behrens@fmrib.ox.ac.uk

Dear Editor,

We would like to submit the attached manuscript titled '**Generative replay for compositional visual understanding in the prefrontal-hippocampal circuit**' for consideration as a Research Article in Cell.

The manuscript

1. **Shows that replay is an important component of understanding. It makes complex configural inferences about the world online.**
   Previous proposed roles for replay have been in service of future behaviour. Here, replay makes inferences about the present. When faced with a visual understanding problem, replay **samples the possible configurations** and **converges on sequences that represent the solution**. Replay infers the configuration of the world from the current visual scene. There is no other example of replay performing online computation. We show it makes inferences by changing its content over timescales of understanding (within seconds).

2. **Provides a novel mechanism to support visual understanding.**
   Unlike modern Deep Networks, biological vision solves an "understanding problem", not simply a "recognition problem". We show that critical computations that support these forms of visual understanding are performed in the hippocampus and surrounding circuitry, and use the same computational machinery (replay and compositionality) as other "understanding" problems, such as spatial reasoning and task inferences. This will be of interest to vision scientists, and to researchers building vision in AI.

3. **Shows that the underlying representations to support visual understanding are relational and compositional.**
   This provides fundamental insight into what it means to "understand" a visual scene. The neural code is 'factorised' into building blocks and relational knowledge that can be flexibly 'conjoined' to form novel compositions. Exactly this kind of neural code has been predicted by a generative model of scene understanding that was recently published (Eslami et al., Science 2018). Furthermore, this suggests a common neural mechanism across superficially distinct functions, such as spatial navigation, model-based planning and compositional inference, as predicted by a computational model recently published in Cell (Whittington et al., Cell 2020).

We believe that these findings will be of interest in multiple disciplines including, neuroscience, artificial intelligence, psychology, psychiatry, and cognitive science, and will bridge knowledge across human and rodent studies. Thus, we believe that Cell would be an outstanding forum for this paper.

We initially submit a full paper to improve readability, but we will be happy to eventually reduce the main text to match the length restrictions of Cell Research Articles. If it would be preferable to adjust the length of the main text for the initial submission we would also be happy to do so.

We suggest the following experts as potential reviewers:

Doris Tsao (University of California, Berkeley, dortsao@berkeley.edu)
Miriam Aly (Columbia University, ma3631@columbia.edu)
Alison Preston (University of Texas, preston@austin.utexas.edu)
Talia Konkle (Harvard University, talia_konkle@harvard.edu)
Sam Gershman (Harvard University, gershman@fas.harvard.edu)
Christian Doeller (Max Planck Institute for Human Cognitive and Brain Sciences, doeller@cbs.mpg.de)

Tim Behrens & Philipp Schwartenbeck (on behalf of the co-authors)

Tim Behrens
Professor of Computational Neuroscience and Associate Director
Wellcome Centre for Integrative Neuroimaging
Nuffield Department of Clinical Neurosciences
University of Oxford

Honorary Principal Investigator
Wellcome Centre for Human Neuroimaging,
University College London.