Representational drift across the macaque ventral stream does not affect all timepoints and stimuli alike: first evidence for a sequence of three different, yet comparatively stable clusters in V4.

Daniel Anthes (danthes@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrück
Wachsbleiche 27, 49090 Osnabrück, Germany

Sushrut Thorat (sthorat@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrück
Wachsbleiche 27, 49090 Osnabrück, Germany

Matthew Ainsworth (matthew.ainsworth@psy.ox.ac.uk)
Dept. Experimental Psychology, University of Oxford
Tinsley Building, 13 Mansfield Road, Oxford,OX1 3SR, UK

Andrew H Bell (andrew.h.bell@kcl.ac.uk)
Institute of Psychiatry, Psychology, and Neuroscience (IoPPN), King's College London
10 Cutcombe Road London, SE5 9RJ, UK

Mark J Buckley (mark.buckley@psy.ox.ac.uk)
Dept. Experimental Psychology, University of Oxford
Tinsley Building, 13 Mansfield Road, Oxford,OX1 3SR, UK

Marieke Mur (mmur@uwo.ca)
Western University
1151 Richmond Street, London, ON N6A 3K7, Canada

Nikolaus Kriegeskorte (nk2765@columbia.edu)
Zuckerman Mind Brain Behavior Institute, Columbia University
3227 Broadway, New York, NY 10027, USA

Peter König* (pkoenig@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrück
Wachsbleiche 27, 49090 Osnabrück, Germany

Tim C Kietzmann* (tkietzma@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrück
Wachsbleiche 27, 49090 Osnabrück, Germany

* shared last author

Neural representations supporting stable behavior have been shown to drift on the timescale of days to weeks. With some exceptions, most studies investigating this phenomenon analyze populations of single neurons in mice. Here, we add to the growing evidence for representational drift by showing that the phenomenon is also found in LFPs recorded from the ventral stream of a macaque monkey during a passive object viewing task. Additionally, we show that as representations evolve over the trial time course, the main axes of the representational geometry are relatively stable, suggesting that drift is not uniform over time and stimuli.

Keywords: representational drift; representations; local field potentials; decoding; ventral stream; electrophysiology

Introduction

Understanding the neural representations that underlie stable perception and behavior is a longstanding effort in neuroscience. Recently, it has been shown that even when an animal's behavior is constant, neural representations relevant to these behaviors continue to change on the timescale of days to weeks (Driscoll, Pettit, Minderer, Chettih, & Harvey, 2017). This phenomenon is termed representational drift, and is most commonly studied at the level of populations of single neurons in mice (Driscoll et al., 2017; Marks & Goard, 2021; Schoonover, Ohashi, Axel, & Fink, 2021) (with notable exceptions (Pinotsis & Miller, 2022; Roth & Merriam, 2023)).

Adding to this, we present a case study of representational drift in local field potentials (LFP) from two regions of the macaque ventral stream (V4, TEp) during a passive object viewing task. We find drift in both studied recording sites.

Additionally, we differentiate neural responses over the trial time course, showing that, consistent with prior work (Kietzmann et al., 2019) the geometry of neural responses changes over the trial time course. At times where neural responses form distinct geometries, subsets of stimuli that cluster together are more stable, while responses to the same stimuli drift substantially at other points in their trajectory.

Results

Stimulus identity can be decoded in all sessions. Linear Support Vector Classifiers (SVC) trained on individual session data perform similarly well across sessions (Fig 1A, V4: $\mu = 0.216, p > 0.05$, TEp: $\mu = 0.288, p > 0.05$), indicating that stimulus identity information is consistently present.

Decoding accuracy decreases with time between training and test session. Testing classifiers on other sessions leads to decreased performance as a function of days between sessions (Fig 1B, $p < 0.001$ in both areas, permutation tests). This suggests that neural responses to stimuli systematically 'drift' over sessions, resulting in lower cross-decoding performance compared to within-session decoding.

The correlation of neural responses to the same stimulus decreases with distance between sessions. The representational drift index (RDI) (Marks & Goard, 2021) can be computed separately for each stimulus and provides a complementary perspective to cross-decoding. It quantifies representational drift for a stimulus as the drop in correlation of responses over sessions, normalized by within-session noise. In agreement with our decoding results this measure also indicates drift in both V4 ($p < 0.001$, permutation test) and TEp ($p < 0.01$, permutation test) (Fig 1C).

Neural responses are reliable at multiple time points with distinct signal geometries. Within session reliability peaks at multiple times in V4 (80ms, 115ms, 175ms) and TEp (130ms), indicating stimulus-relevant information is available at these moments. However, decoding performance for stimulus identity shows only a single peak. Representational dissimilarity matrices (RDMs) (Kriegeskorte, Mur, & Bandettini, 2008) computed at peak reliability times reveal qualitative differences, suggesting distinct representational stages (Kietzmann et al., 2019), that provide varying information about stimulus identity (Fig 1D). Hierarchical clustering of the RDMs at different times reveals clusters with different stimuli and structures.

The main clusters in RDMs at different times are relatively stable. We assess the stability of representational geometry by computing the standard deviation over session-wise RDMs at three time points. Distances between stimuli within and between the main clusters are relatively constant over sessions (low standard deviation), while distances to other stimuli are more variable. Stimuli in the main clusters show little drift across sessions (low RDI, Fig 1F, top), and high reliability within sessions (Fig 1F, bottom). However, these stimuli exhibit varying reliability and drift at other times in the neural trajectory.

These results suggest that neural responses do not drift uniformly, but that different features are prominent at distinct times during the neural trajectory. Responses aligned with these features remain relatively stable when they are most salient.

Conclusion

We find drift in LFPs in macaque V4 and TEp during a natural object viewing task. The extent of this drift differs over both stimuli and the neural time course, potentially providing insight into the stability of different neural processes that unfold over time.

Method

Experiment & Data In each experimental session a macaque repeatedly viewed a set of 92 objects on gray backgrounds (Kriegeskorte et al., 2008). Stimuli were presented in an RSVP paradigm for 250ms each while the monkey was tasked to fixate in the center of the image. The experiment was repeated for 9 sessions over approximately 100 days. Data was recorded from two Utah arrays in V4 (45 channels) and TEp (64 channels). Recordings are filtered between 0.1Hz and 60Hz.

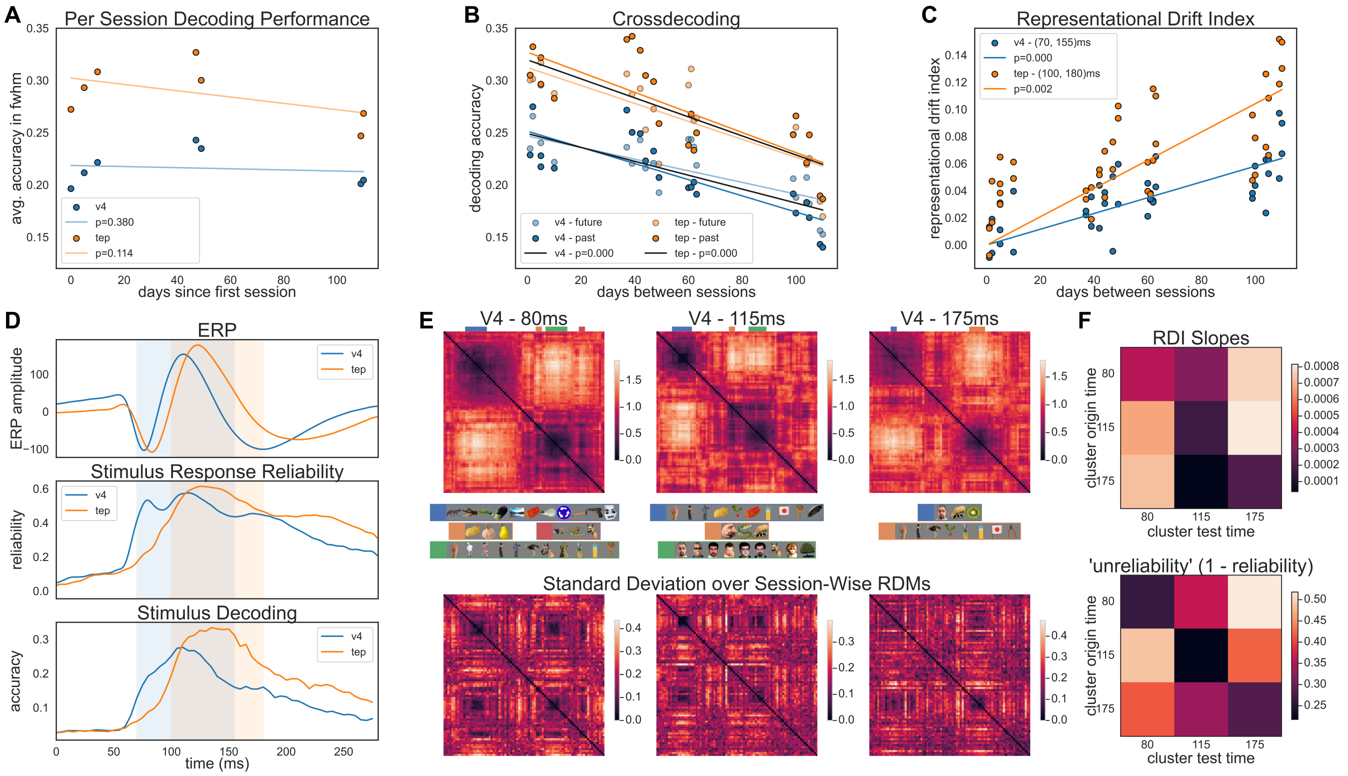

Figure 1: **A:** Average (82-way) session-wise decoding performance in the fwhm interval. **B:** Cross-decoding performance averaged over the decoding fwhm interval. **C:** RDI per recording site. Computed every 5ms and averaged over the decoding fwhm interval. **D:** grand average ERP per recording site (top), reliability averaged over stimuli and sessions (middle) and decoding time courses averaged over sessions. Shaded areas indicate fwhm intervals. (bottom). **E:** Average session-wise RDMs for time points in V4 sorted with hierarchical clustering. (top), and standard deviation over sessions (bottom). **F:** Average RDI slope (top) and within-session reliability (bottom) for stimuli that form clusters at different time points.

We follow the method described in (Kietzmann et al., 2019) for pre-processing: Noisy channels are rejected if variance in the baseline interval deviates more than three standard deviations from the median over all channels in the same recording site. Noisy trials are removed with Autoreject (Jas, Engemann, Bekhti, Raimondo, & Gramfort, 2017), without interpolation. All pre-processing is performed separately for each recording session and recording site and channels that are rejected in any session are excluded from all sessions. Trials that are identified as outliers in one recording site are rejected in both arrays.

Finally, all stimuli with less than three trials in any session after pre-processing are excluded from the analyses. The final dataset contains 82 unique stimuli. All presented analyses are conducted on the residual signals after subtracting the session-wise ERP for each channel.

Decoding Decoding analyses use SVCs with linear kernels trained separately for each session and time bin (steps of 5ms). Regularization parameter C is chosen by stratified cross validation (3 folds). Results in Fig 1A and B are averages over the full width at half max (fwhm) interval for each array (shaded areas, Fig 1D). Sessions where the mean decoding performance in the fwhm interval is more than 1.5 standard deviations from the median are excluded from all analyses (seven sessions remain). For cross-decoding, we refit each SVC on all data from a single session with C chosen by cross validation.

Representational Drift Index We quantify the per stimulus change in neural response with the representational drift index (RDI), introduced by (Marks & Goard, 2021). Due to the low number of repeats for each stimulus in a session we perform repeated sampling instead of split half estimates.

Reliability Reliability is computed as the oracle correlation: For each stimulus and session we repeatedly correlate each trial with the average over the remaining n-1 trials.

Stimulus Response Geometry RDMs are computed per session, repeatedly sampling two trials per stimulus that are averaged to compute one sample RDM. All RDMs are computed using correlation distance. Hierarchical clustering is performed based on average distance. Threshold distance for forming flat clusters is 0.4. Singleton clusters are discarded.

Acknowledgments

The project was financed by the funds of the research training group "Computational Cognition" (GRK2340) provided by the Deutsche Forschungsgemeinschaft (DFG), Germany and the European Union (ERC, TIME, Project 101039524). Compute resources used for this project are additionally supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 456666331. Data was recorded as part of the MRC CBSU projects MC_UU_00005/13 and MC_UU_00005/6.

References

Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N., & Harvey, C. D. (2017). Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, *170*(5), 986–999.

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for meg and eeg data. *NeuroImage*, *159*, 417–429.

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

Marks, T. D., & Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. *Nature communications*, *12*(1), 5169.

Pinotsis, D. A., & Miller, E. K. (2022). Beyond dimension reduction: Stable electric fields emerge from and allow representational drift. *NeuroImage*, *253*, 119058.

Roth, Z. N., & Merriam, E. P. (2023). Representations in human primary visual cortex drift over time. *Nature Communications*, *14*(1), 4422.

Schoonover, C. E., Ohashi, S. N., Axel, R., & Fink, A. J. (2021). Representational drift in primary olfactory cortex. *Nature*, *594*(7864), 541–546.