Geometry of task representations in human frontal cortical neurons is predictive of task switch costs.

Daniel Deng (hdeng3@caltech.edu)
Division of Biology and Biological Engineering, California Institute of Technology
1200 East California Blvd, Pasadena, CA 91125, USA

Hristos Courellis (hristos@caltech.edu)
Division of Biology and Biological Engineering, California Institute of Technology
1200 East California Blvd, Pasadena, CA 91125, USA

Ivan Skelin (ivan.skelin@uhn.ca)
Krembil Research Institute, University Health Network, University of Toronto
60 Leonard Ave, Toronto, ON M5T 0S8, Canada

Juri Minxha, PhD (jminxha@caltech.com)
Division of Biology and Biological Engineering, California Institute of Technology
1200 East California Blvd, Pasadena, CA 91125, USA

Taufik Valiante, MD PhD (taufik.valiante@uhn.ca)
Neurosurgery Department, Toronto Western Hospital
399 Bathurst St, Toronto, ON M5T 2S8, Canada

Adam Mamelak, MD (adam.mamelak@cshs.edu)
Neurosurgery Department, Cedars-Sinai Medical Center
8700 Beverly Blvd, Los Angeles, CA 90048, USA

Ueli Rutishauser, PhD (urut@caltech.edu)
Division of Biology and Biological Engineering, California Institute of Technology
1200 East California Blvd, Pasadena, CA 91125, USA

Abstract: The neurophysiological mechanisms underlying task switch costs in humans remain elusive, particularly the irreducible cost that persists when given sufficient time to prepare following instructions. Two competing theories, reconfiguration and task-set inertia, provide differing accounts for the generation of switch costs, but without support from single-unit recordings. Here, we analyze the activity of large populations of single-neurons in the medial frontal cortex while neurosurgical patients are engaged in instructed task-switching. We demonstrate that task representations undergo reconfiguration on switch trials, and that inertia in the baseline representations of task context are predictive of upcoming switch costs, providing support for both theories.

Keywords: task switch costs, neural representational geometry, medial frontal cortex, single neuron

Introduction

The process of switching between tasks occurs countless times throughout the day for an individual. Every instance of switching is accompanied by a cost, a decrease in task accuracy and/or speed immediately after switching that rapidly fades away (Monsell, 2015). Though this switch cost is reducible when preparatory time is given after instructions, an irreducible switch cost is always present the first time one engages in a task when switching from a different task. The presence of switch costs in animals is debated, being absent from some species entirely, but is a prominent aspect of human cognition (Caselli & Chelazzi, 2011; O'Donoghue & Wasserman, 2021; Stoet & Snyder, 2003). The neural mechanisms that generate switch costs remains unknow and are hotly debated. Theories center around two possible causes: reconfiguration and lingering activity (inertia) related to the prior task (Monsell, 2015). Some evidence from intracranial recordings exists supporting these proposed explanations (Minxha et al., 2020; Weber et al., 2023), which indicate a key role of the medial frontal cortex (MFC). However, the neurophysiological basis of switch costs remains elusive.

To arbitrate between different theories of switch costs, we recorded the activity of large populations of single neurons in the MFC of neurosurgical patients performing a task with frequent instructed switching. We find that the task context representations immediately following and far from a switch exist in orthogonal subspaces composed of non-overlapping populations of neurons. The task representation in the latter subspace persistently encoding the previous task is predictive of switch costs.

Methods

Task Subjects alternated between two possible tasks: categorization (e.g. "Is this an image of X**?**", where X is the target category), and memory (e.g. "Have you seen this image before?") (Fig. 1a). Each experiment consisted of 48 blocks of 8 trials. Task instructions were given once at the start of each block, and needed to be remembered for the ensuing 8 trials (Fig. 1b). All questions were yes/no questions, with subjects answering as quickly as possible. We refer to the

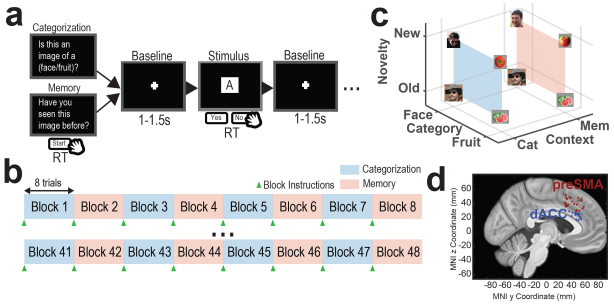

Figure 1. (a) Task trial structure and **(b)** block structure. **(c)** Task condition structure during the stimulus period. 8 unique conditions lead to 35 balanced dichotomies. Cat = categorization, Mem = memory **(d)** Implant locations.

question being answered as the context for that block, either Categorization (Cat) or Memory (Mem). Images belonged to one of two categories (fruits, faces), with some repeated ("old") and some shown the first time, resulting in 8 total possible conditions (Fig. 1c). A balanced number of trials of each condition were present in every block and at every trial number across blocks. Switch costs were operationalized as the excess time taken to complete the first trial after switching tasks. For each block, patients control when to proceed from the instruction screen to the first trial (Fig. 1a), such that they are sufficiently prepared and the behavioral cost present during Trial 1 after a switch is the irreducible switch cost.

Recording and Processing Patients with pharmacologically intractable epilepsy were implanted with Behnke-Fried electrodes (Fried et al., 1999) that allowed for recording of single-unit activity from medial frontal cortical (MFC) structures including the dorsal anterior cingulate (dACC) and pre-supplementary motor area (preSMA) (Fig. 1d). Unit activity from these regions was isolated using standard spike sorting techniques (Rutishauser et al., 2006). Spikes were counted during two time periods: baseline (-1 to 0 s prior to stimulus onset) and stimulus (0.2 to 1.2s after stimulus onset). "Trial 1" baseline spikes are recorded after a patient has read the instructions and pressed a button initiating a block, but has not yet performed the task instructed for that block.

Quantifying Representational Geometry We used two metrics to quantify the content and format of neural representations: shattering dimensionality (SD) and cross-condition generalization performance (CCGP). Discussion of these metrics and their uses are available in prior work (Bernardi et al., 2020; Courellis et al., 2023). In brief, metrics operate over balanced dichotomies of task conditions (Fig. 1B), which are formed by splitting the 8 unique conditions into two equal groups of 4 conditions (e.g. 4 points in category vs 4 points in memory is the context dichotomy). Each metric is computed independently for the 35 possible unique dichotomies. SD is the average decoding accuracy over all balanced dichotomies, and is an index of the expressiveness of a representation**.** CCGP is an index of abstraction, with high CCGP dichotomies indicating that those variables are disentangled from other decodable variables in the representation because performance

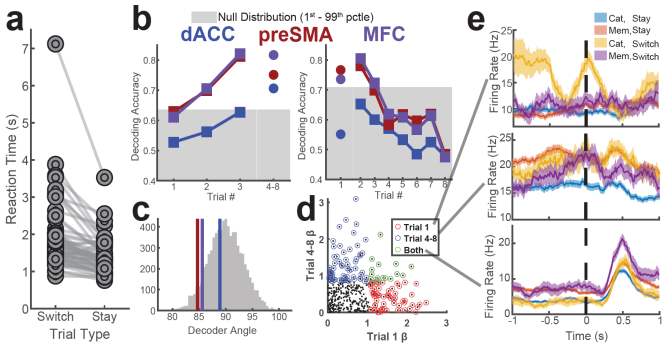

Figure 2. Baseline period context representations. (a) Mean reaction time during Switch (1) and Stay (4-8) trials. Every pair of points corresponds to a single session. **(b)** Baseline context decoder trained on Trials 4-8 (right) and on Trial 1 (left) after task switching. Circles indicate cross-validated training performance and squares indicate generalization performance to held-out trials. 95th pctle of shuffle null distribution shown in gray. **(c)** Angle between context coding vectors computed from Trial 1 and Trial 4-8 decoders. Gray histogram indicates shuffle null. **(d)** Scatter plot of single-neuron importance index (β) for Trial 1 and Trial 4-8 decoders. Each black dot corresponds to one neuron. Neurons in the top 20% for Trial 1 decoder (red), Trial 4-8 decoder (blue), or both (green) are circled. **(e)** Example PSTHs of neurons contributing to each of the context decoders. **(f)** Correlation of baseline context representations and Trial 1 reaction time (switch cost) for the preSMA context decoder trained on Trial 1 (left) and the dACC context decoder trained on Trials 4-8 (right).

generalizes to completely held-out conditions. These geometric measures are computed in neural state spaces constructed from the activity of all recorded neurons.

Results

Context representation following instructions predict switch costs. Data recorded over 56 sessions (n = 35 patients) yielded 757 well isolated neurons. Switching costs were robust for both tasks (Fig. 2a, each line is a session), with Trial 1 after an instruction screen on average 40% slower than the average block RT. We decoded task context from spikes counted during the baseline period and found context to be robustly decodable from activity of MFC neurons during Trials 4-8 after a switch (Fig. 2b, left, decoder trained on Trials 4-8). However, this decoder (henceforth steady-state subspace) did not generalize to decode activity in Trial 1. Yet, context was decodable from Trial 1 when training and testing a decoder during Trial 1 only (Fig. 2b, right, red). Conversely, the Trial 1 decoder failed to generalize to Trials 4-8, with context decodability in the subspace identified by this decoder (henceforth switch subspace) falling to chance after Trial 3 post-switch. These two context coding subspaces were orthogonal (Fig. 2c) by virtue of being largely non-overlapping populations of neurons (Fig. 2d,e). Greater context decodability in both subspaces predicted faster RT (lower switch cost) on the upcoming trial (Fig. 2f). On slow trials, the context of the previous block was decodable from dACC as indicated by below-chance decoding (Fig. 2f, right).

Stimulus representations reconfigure during switch trials. Representational geometry was quantified during the stimulus period by performing SD and CCGP analysis on Trials 4-8 (Stay) and Trial 1 (Switch). All three stimulus properties (context, novelty, category) were decodable on Stay trials in both dACC and preSMA (Fig. 3a). However, dACC alone exhibited a significant decrease in SD (Fig. 3a, black line) and CCGP for context (Fig. 3b, red) on switch trials. The mis-configuration of the dACC representation on Switch trials is visualized in Fig. 3c,d by performing multi-dimensional scaling (MDS) on condition-averaged neural activity from dACC alone. The systematically structured Stay

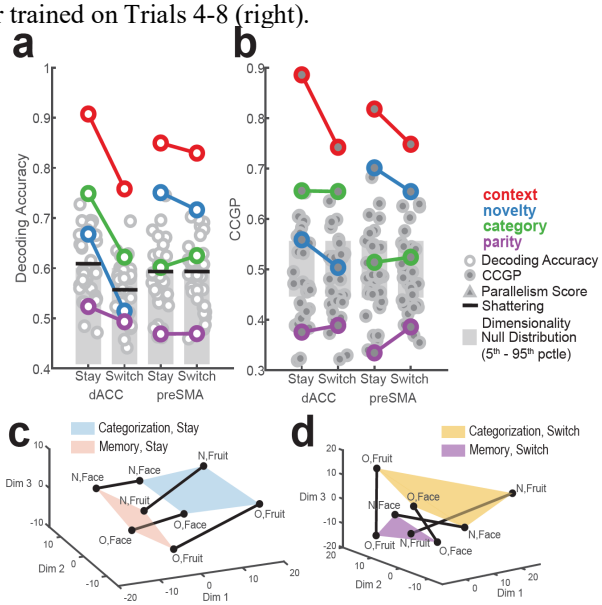

Figure 3. Stimulus period context and stimulus representations. (a) Reduction in decodability of task-relevant dichotomies and reduction of shattering dimensionality on switch trials compared to stay trials in the dACC (left). This effect is absent from the preSMA (right). **(b)** CCGP of context representation significantly reduced on switch trials in dACC. Dimensionality reduction of dACC neural responses using MDS during stay **(c)** and switch **(d)** trials.

trial representation (Fig. 3c) is contrasted with the relatively disorganized Switch trial representation (Fig. 3d).

Discussion

Both the task-set inertia and reconfiguration theories are consistent with aspects of our data. Baseline and stimulus period task representations in the MFC undergo reconfiguration following switch trials, and previous-context decodability is correlated with higher switch costs (inertia). Further analysis is needed to explore switch cost prediction during the stimulus period, switch-trial response conflicts, and to clarify the effect of practice, which can reduce switch costs.

Acknowledgments

We thank the members of the Adolphs and Rutishauser labs for discussion, C. M. Reed, J. M. Chung, and the staff of the Cedars-Sinai and Toronto Western Epilepsy Monitoring Units for their support with recording, and we thank the patients and their families for their boundless generosity.

Funding: NIH U01NS117839, Simons Foundation Collaboration on the Global Brain (542941), and the Caltech NIMH Conte Center P50MH094258).

References

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell*, *183*(4), 954-967.e21. https://doi.org/10.1016/j.cell.2020.09.031

Caselli, L., & Chelazzi, L. (2011). Does the Macaque Monkey Provide a Good Model for Studying Human Executive Control? A Comparative Behavioral Study of Task Switching. *PLoS ONE*, *6*(6), e21489. https://doi.org/10.1371/journal.pone.0021489

Courellis, H. S., Mixha, J., Cardenas, A. R., Kimmel, D., Reed, C. M., Valiante, T. A., Salzman, C. D., Mamelak, A. N., Adolphs, R., Fusi, S., & Rutishauser, U. (2023). *Abstract representations emerge in human hippocampal neurons during inference behavior* (p. 2023.11.10.566490). bioRxiv. https://doi.org/10.1101/2023.11.10.566490

Fried, I., Wilson, C. L., Maidment, N. T., Engel, J., Behnke, E., Fields, T. A., MacDonald, K. A., Morrow, J. W., & Ackerson, L. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients. Technical note. *Journal of Neurosurgery*, *91*(4), 697–705. https://doi.org/10.3171/jns.1999.91.4.0697

Minxha, J., Adolphs, R., Fusi, S., Mamelak, A. N., & Rutishauser, U. (2020). Flexible recruitment of memory-based choice representations by human medial-frontal cortex. *Science (New York, N.Y.)*, *368*(6498), eaba3313. https://doi.org/10.1126/science.aba3313

Monsell, S. (2015). Task-set control and task switching. In *The handbook of attention* (pp. 139–172). Boston Review. https://doi.org/10.1093/med:psych/9780198528883.003.0002

O'Donoghue, E., & Wasserman, E. A. (2021). Pigeons proficiently switch among four tasks without cost. *Journal of Experimental Psychology: Animal Learning and Cognition*, *47*(2), 150–162. https://doi.org/10.1037/xan0000287

Rutishauser, U., Schuman, E. M., & Mamelak, A. N. (2006). Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. *Journal of Neuroscience Methods*, *154*(1–2), 204–224. https://doi.org/10.1016/j.jneumeth.2005.12.033

Stoet, G., & Snyder, L. H. (2003). Executive control and task-switching in monkeys. *Neuropsychologia*, *41*(10), 1357–1364. https://doi.org/10.1016/S0028-3932(03)00048-4

Weber, J., Iwama, G., Solbakk, A.-K., Blenkmann, A. O., Larsson, P. G., Ivanovic, J., Knight, R. T., Endestad, T., & Helfrich, R. (2023). Subspace partitioning in the human prefrontal cortex resolves cognitive interference. *Proceedings of the National Academy of Sciences*, *120*(28), e2220523120. https://doi.org/10.1073/pnas.2220523120