

# UC San Diego

## UC San Diego Previously Published Works

### Title

Dynamics Are the Only Constant in Working Memory

### Permalink

<https://escholarship.org/uc/item/2dk2x478>

### Journal

Journal of Cognitive Neuroscience, 35(1)

### ISSN

0898-929X

### Authors

Adam, Kirsten CS  
Rademaker, Rosanne L  
Serences, John T

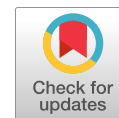
### Publication Date

2022-12-01

### DOI

10.1162/jocn\_a\_01941

Peer reviewed



# Dynamics Are the Only Constant in Working Memory

Kirsten C. S. Adam<sup>1</sup>, Rosanne L. Rademaker<sup>2</sup>, and John T. Serences<sup>1</sup>

## Abstract

■ In this short perspective, we reflect upon our tendency to use oversimplified and idiosyncratic tasks in a quest to discover general mechanisms of working memory. We discuss how the work of Mark Stokes and collaborators has looked beyond localized, temporally persistent neural activity and shifted focus toward the importance of distributed, dynamic neural codes for working

memory. A critical lesson from this work is that using simplified tasks does not automatically simplify the neural computations supporting behavior (even if we wish it were so). Moreover, Stokes' insights about multidimensional dynamics highlight the flexibility of the neural codes underlying cognition and have pushed the field to look beyond static measures of working memory. ■

The central goal of working memory research is to understand how we temporarily hold information in mind while moving through the world to achieve our behavioral goals. Working memory is a critical cognitive function that allows us to link together our experiences into a coherent narrative. Not surprisingly then, in the talks and classroom lectures that we as scientists give on working memory, we all love to begin with engaging real-world examples. We might show a crowded supermarket aisle filled with colorful vegetables and highlight how you use working memory to hold your grocery list in mind while searching for the perfect strawberries. Or, we might show a busy city street and explain how you can find a friend in the crowd by visualizing their face or signature pink beret. There is no shortage of vivid examples for opening talk slides. However, around Slide 5, we invariably pivot to the following: a few discrete gray boxes representing a computer screen, set sequentially on a timeline. In one of the first boxes, there is a swipe of stripes or splash of color, which people are asked to remember. Then, there is a gray box, meant to indicate the working memory delay. And in the last box, there are again some stripes or colors used as a test.

Those of us engaged in working memory research are so familiar with this pivot from real world to laboratory that we scarcely notice it. First-time listeners, however, might struggle to see the connection. Of course, there is a reason we all love our artificial tasks, and there is great value in them. As psychologists, we learn that a task is a means for exerting experimental control at the cost of naturalism. As neuroscientists, we find that artificial tasks are particularly useful for shepherding people's brains into approximately the same state over and over again so that we can

extract the signal from the noise. By contrast, if you find yourself wandering through the supermarket aisle one fine morning as per the example on Slide 1, it is difficult (currently, near impossible) for a neuroscientist to glean information from your mind.

Because they are simple, we like to think of our typical working memory tasks as predictable and interchangeable. Like following a recipe in the kitchen, you can predict behavior based on the time allotted for encoding and retention. Like parts from IKEA shelves, you can mix and match task components to get the desired effect. However, ongoing work has shown how even the simplest task components are not so formulaic. For example, one long-standing question that contributed to the split of cognitive psychology from behaviorism is how one stimulus can map onto many different behaviors. Drawing an ace from the deck is sometimes the best card and sometimes the worst—it all depends on which card game you are currently playing. In technical terms, the same stimulus triggers different mental operations and behaviors in different contexts.

Understanding how one stimulus can be flexibly mapped to different behaviors is a particularly challenging problem when viewed from the perspective of individual neurons. In visual neuroscience, it is often fruitful to characterize neurons' tuning preferences. It is easy to imagine mental representations arising from stably tuned neurons—if you want to represent a “vertical” item, in theory, you could achieve this by having vertical-preferring neurons persistently fire to bridge a delay. Yet, a scheme like this cannot fully account for working memory's flexibility: Sometimes, “vertical” could mean “press button A,” and other times, it could mean “look to your right.” To rapidly link arbitrary pieces of information together requires flexible shifts in the representation of information. A key insight from Stokes et al. (2013) is that a multidimensional landscape emerges when an individual neurons' activity is

<sup>1</sup>University of California, San Diego, <sup>2</sup>Ernst Strüngmann Institute for Neuroscience in cooperation with the Max Planck Society, Frankfurt, Germany

viewed in relation to the activity of all other neurons. In this landscape, each neuron traverses a single dimension over time, and all neurons together traverse a highly dynamic trajectory that can settle into stable states during various epochs of the working memory task. This dynamic and multidimensional state space can be considered across all neurons, but can also be condensed back into fewer dimensions by looking only at those components that explain most of the variance in a given task (using a dimensionality reduction technique like PCA). From Stokes et al. (2013), we learn that a lower-dimensional stable activation state can be observed during working memory maintenance, which reflects the temporarily configured network state that is dynamically tuned according to task goals. For example, a stable state can map how a memorized stimulus relates to an appropriate decision required during response. Critically, by taking into account the multidimensional nature of neural codes, many flexible behaviors can suddenly fit rather effortlessly into our theories about working memory.

Work on multidimensional codes in the context of mapping one stimulus to multiple behaviors has shown how low dimensional states can be flexibly assembled and reassembled to adapt to moment-to-moment behavioral demands. Even more remarkably, subsequent work has revealed that activity across large populations of neurons can remain highly dynamic even when the stimulus and task demands are held constant (Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020; Murray et al., 2017; Spaak, Watanabe, Funahashi, & Stokes, 2017). In a visual working memory task, observers were asked to remember a simple stimulus (like an oriented grating or spatial location). The overall pattern of neural activity during the memory delay is volleyed through a series of changes over time. Despite these rapid temporal dynamics in the population at large, the coding-scheme, or the low-dimensional subspace that represents the simple stimulus, remained remarkably stable, exhibiting only small drifts over time (Wolff et al., 2020; Murray et al., 2017). From our conscious perspective, memory of a simple stimulus such as an orientation is like a statue held “fixed” in our mind’s eye. From a neural information processing perspective, it is like a river finding its way down the different grooves in a landscape—all the while keeping the memories afloat on a stable boat (Panichello & Buschman, 2021; Panichello, DePasquale, Pillow, & Buschman, 2019).

Understanding working memory codes as highly dynamic and evolving across time was a transformative idea from Stokes and colleagues (2013), and we are only slowly beginning to understand more about the ways in which memories are maintained from this novel perspective. For example, recent work has adopted the dynamic coding framework developed by Stokes to address one of the classic questions in philosophy, cognitive psychology, and neuroscience: When you hold a memory in mind, how do you know that it is a memory, and not a representation of incoming sensory information? Put

another way, how does your brain attenuate interference between internal thoughts and sensory information? By examining the state space of multi-unit recordings, Libby and Buschman (2021) demonstrated that the sensory tuning of some neurons is stable during the maintenance of information in memory, whereas the tuning of other neurons is inverted with respect to sensory tuning. The net result is a rotation of the state-space representation of the memory code with respect to the sensory code, providing a mechanism to separate memory representations from sensory representations. Although this finding suggests a means of mitigating interference between memories and sensory inputs, these dynamics complicate the process of decoding the remembered information to guide behavior. How can a specific remembered feature be “read-out” when that feature is no longer in its original sensory-like format? Several studies—inspired again by Stokes’ approach to dynamic codes—have shown that neural response patterns can be highly dynamic over time, all while preserving the structural relationship between the stimuli being remembered, so that they remain separable in a stable subspace (e.g., Wolff et al., 2020; Bouchacourt & Buschman, 2019; Murray et al., 2017; Spaak et al., 2017).

The inspiration sparked by the idea of dynamic codes (although employing different analysis approaches) has steadily trickled into the neuroscientific thinking on an equally long-standing question: Where is the cortical locus of mnemonic representations? The classic story is that sustained spiking activity in pFC is the key mechanism supporting the stable memory representations guiding behavior. However, Stokes’ demonstration of dynamic codes forced the field to reconsider. With the notion of multidimensional and dynamic subspaces as a starting point, should we even expect any single neural locus (e.g., pFC), or a single mechanism (e.g., sustained spiking), to be the seat of working memory? Instead, for any given working memory task—be it remembering colored squares or remembering your grocery list—there should be some distributed and temporally evolving pattern of neural activity that flexibly recruits the brain areas and neural mechanisms needed to get the job done (Courtney, 2022; Iamshchinina, Christophel, Gayet, & Rademaker, 2021; Lorenc & Sreenivasan, 2021; Christophel, Klink, Spitzer, Roelfsema, & Haynes, 2017; Sreenivasan, Vytlačil, & D’Esposito, 2014). Indeed, recent work has shown that information about visual stimuli can be recoded into motor representations if a response is known in advance (Henderson, Rademaker, & Serences, 2022; Boettcher, Gresch, Nobre, & van Ede, 2021), or can be recoded into an abstracted mnemonic format (Kwak & Curtis, 2022; Rademaker, Chunharas, & Serences, 2019). In summary, Stokes’ work inspires the idea that there is not one place or one mechanism that is a constant during working memory. Instead, the only constant is flexibility and the temporal dynamics that connect sensory inputs to context-specific behavioral goals.

One implication of the dynamic coding framework is that there is not a general solution to the “problem” of working memory. To understand working memory, we need to reckon directly with its immense flexibility. To do so will require both devising new tasks, as well as carefully considering how changing mental states and behavioral contexts impact processing in even the simplest of tasks. This does not mean that some principles of working memory will not generalize—some “solutions” might be more or less similar given the relationship between different contexts. However, a better understanding of the dynamics of working memory—as revealed by Stokes and others in the past decade—should motivate more consideration about the design and relevance of our tasks for everyday life, and help to kick us out of the attractor-like state of thinking that there is only one way to implement working memory in the brain.

Reprint requests should be sent to Kirsten C. S. Adam, Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109, or via e-mail: kadam@ucsd.edu.

### Author Contributions

Kirsten C. S. Adam: Conceptualization; Writing—Original draft. Rosanne L. Rademaker: Conceptualization; Writing—Original draft. John T. Serences: Conceptualization; Writing—Original draft.

### Funding Information

Kirsten C. S. Adam, National Science Foundation (<https://dx.doi.org/10.13039/1000000001>), grant number: 2104630. John T. Serences, National Eye Institute (<https://dx.doi.org/10.13039/1000000053>), grant number: R01 EY025872.

### REFERENCES

- Boettcher, S. E. P., Gresch, D., Nobre, A. C., & van Ede, F. (2021). Output planning at the input stage in visual working memory. *Science Advances*, *7*, eabe8212. <https://doi.org/10.1126/sciadv.abe8212>, PubMed: 33762341
- Bouchacourt, F., & Buschman, T. J. (2019). A flexible model of working memory. *Neuron*, *103*, 147–160. <https://doi.org/10.1016/j.neuron.2019.04.020>, PubMed: 31103359
- Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*, 111–124. <https://doi.org/10.1016/j.tics.2016.12.007>, PubMed: 28063661
- Courtney, S. M. (2022). Working memory is a distributed dynamic process. *Cognitive Neuroscience*, *13*, 208–209. <https://doi.org/10.1080/17588928.2022.2131747>, PubMed: 36200905
- Henderson, M. M., Rademaker, R. L., & Serences, J. T. (2022). Flexible utilization of spatial- and motor-based codes for the storage of visuo-spatial information. *eLife*, *11*, e75688. <https://doi.org/10.7554/eLife.75688>, PubMed: 35522567
- Iamshchinina, P., Christophel, T. B., Gayet, S., & Rademaker, R. L. (2021). Essential considerations for exploring visual working memory storage in the human brain. *Visual Cognition*, *29*, 425–436. <https://doi.org/10.1080/13506285.2021.1915902>
- Kwak, Y., & Curtis, C. E. (2022). Unveiling the abstract format of mnemonic representations. *Neuron*, *110*, 1822–1828. <https://doi.org/10.1016/j.neuron.2022.03.016>, PubMed: 35395195
- Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, *24*, 715–726. <https://doi.org/10.1038/s41593-021-00821-9>, PubMed: 33821001
- Lorenc, E. S., & Sreenivasan, K. K. (2021). Reframing the debate: The distributed systems view of working memory. *Visual Cognition*, *29*, 416–424. <https://doi.org/10.1080/13506285.2021.1899091>
- Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, 394–399. <https://doi.org/10.1073/pnas.1619449114>, PubMed: 28028221
- Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*, 601–605. <https://doi.org/10.1038/s41586-021-03390-w>, PubMed: 33790467
- Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. J. (2019). Error-correcting dynamics in visual working memory. *Nature Communications*, *10*, 3366. <https://doi.org/10.1038/s41467-019-11298-3>, PubMed: 31358740
- Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*, 1336–1344. <https://doi.org/10.1038/s41593-019-0428-x>, PubMed: 31263205
- Spaak, E., Watanabe, K., Funahashi, S., & Stokes, M. G. (2017). Stable and dynamic coding for working memory in primate prefrontal cortex. *Journal of Neuroscience*, *37*, 6503–6516. <https://doi.org/10.1523/JNEUROSCI.3364-16.2017>, PubMed: 28559375
- Sreenivasan, K. K., Vytlačil, J., & D’Esposito, M. (2014). Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. *Journal of Cognitive Neuroscience*, *26*, 1141–1153. [https://doi.org/10.1162/jocn\\_a\\_00556](https://doi.org/10.1162/jocn_a_00556), PubMed: 24392897
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*, 364–375. <https://doi.org/10.1016/j.neuron.2013.01.039>, PubMed: 23562541
- Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., & Stokes, M. G. (2020). Drifting codes within a stable coding scheme for working memory. *PLoS Biology*, *18*, e3000625. <https://doi.org/10.1371/journal.pbio.3000625>, PubMed: 32119658