

## Journal Club

**Editor's Note:** These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at [www.jneurosci.org](http://www.jneurosci.org). For more information on the format, review process, and purpose of Journal Club articles, please see <http://jneurosci.org/content/preparing-manuscript#journalclub>.

## Working Memories Are Maintained in a Stable Code

 Joao Barbosa

Theoretical Neurobiology, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain  
Review of Spaak et al.

Working memory is the ability to maintain, in a ready-to-manipulate state, information that is no longer available for sensory processing. A classic way of studying visual working memory is to briefly show a visual cue at one of several possible locations and, after a short delay, to ask the subject to report where the cue was presented. Early neurophysiological experiments in primates found that during the delay a fraction of the neurons in the prefrontal cortex (PFC) carries information about the previously shown stimulus as persistent activity (Fuster and Alexander, 1971). Such neurons have subsequently been found in many other areas of the monkey brain (Leavitt et al., 2017) and human brain (Kamiński et al., 2017), supporting the hypothesis that items are maintained in working memory through a persistent code (Zylberberg et al., 2017).

Despite the fact that some neurons are persistently active during the maintenance period in working memory tasks, most recorded neurons are not (Leavitt et al., 2017). This has led to the emergence of alternative hypotheses concerning the neural bases of working memory. For example, instead of relying on a stable code accomplished by persistent activity in single neurons, information could be stored

“silently” in enhanced synaptic strengths through short-term plasticity (the synaptic hypothesis; Mongillo et al., 2008). Alternatively, a stable representation might be achieved by combining transient activity of a large population of neurons, each of which is active at different epochs of the delay period (the dynamic code hypothesis; Goldman, 2009; Druckmann and Chklovskii, 2012). It is important to stress that in both hypotheses, single neurons are not persistently active and their selectivity is not stable during the delay. Only models based on the synaptic hypothesis rely on synaptic plasticity mechanisms, however.

Spaak et al. (2017) provide additional insights about the neural mechanisms of working memory by reanalyzing three previously published data sets (Watanabe and Funahashi, 2007, 2014) of single-unit recordings of PFC neurons from four monkeys trained to do variations of a working memory task. To assess whether the same neural code is used during different trial periods, the authors trained a decoder on single-unit activity during one period of a task and used this decoder to extract stimulus information from activity recorded during other periods (here called “cross-temporal analysis,” as in the study by Stokes, 2015). If the decoder could extract similar amounts of stimulus information from different trial periods, the working memory code was said to be generalizable across those periods. In addition, the authors derived a measure of stability, called the “dynamic index”

(DI), with low DI values indicating that neuronal activity was stable during a given period in the trial. In all data sets, Spaak et al. (2017) found evidence for both dynamic and stable codes, depending on the task period. Specifically, neural activity during most of the maintenance period showed a key signature of a stable code: single neurons did not change their selectivity, which led to robust code generalization, and had low DI values, therefore rejecting the presence of a dynamic code during memory maintenance. But during stimulus presentation, early delay (up to ~500 ms after stimulus removal), and response periods, single neurons had high DI values and often changed their selective location, which is in line with the dynamic code hypothesis. These dynamic periods were also characterized by a less generalizable code, as revealed by cross-temporal decoding analysis. These results are consistent with those of two other recent studies that used similar methods to analyze single units recorded from the medial-temporal lobe in humans performing a multi-item visual working memory task involving complex figures (Kamiński et al., 2017) and from the PFC in monkeys performing a vibrotactile working memory task (Murray et al., 2017). Those studies also found transient dynamics during encoding, followed by a stable mnemonic code during maintenance, hence supporting these dynamics as a general mechanism for working memory.

One possible explanation for the transient activity during stimulus presentation

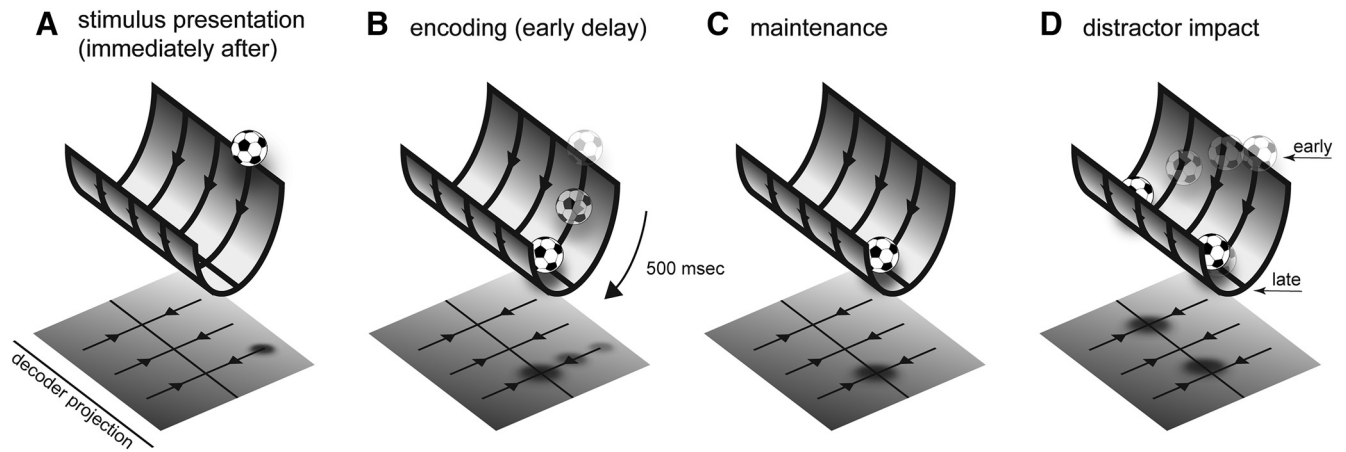
Received May 31, 2017; revised July 21, 2017; accepted July 25, 2017.

I thank Albert Compte and Tatiana Pasternak for discussing and reviewing the text, and Adriano Martins for helping with the figure.

Correspondence should be addressed to Joao Barbosa, Theoretical Neurobiology, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain. E-mail: [palerma@gmail.com](mailto:palerma@gmail.com).

DOI:10.1523/JNEUROSCI.1547-17.2017

Copyright © 2017 the authors 0270-6474/17/378309-03\$15.00/0



**Figure 1.** Dynamic encoding of working memories is compatible with working memory through line attractors. The attractor states of the system form a one-dimensional valley, along which the system converges, in the high-dimensional space of network activity. **A**, Stimulus presentation puts the system (represented here as a football) close to the attractor that represents that stimulus. **B**, Upon stimulus removal, the system moves toward a stable state. **C**, This results in some dynamic activity while the system moves toward the attractor, where it settles for the whole delay period. During encoding, because the system is not yet settled in a stable state, the system is still susceptible to distractors that might push the system closer to another, unrelated stable state. **D**, The same distractor late in the delay period has a negligible impact. Figure adapted from Seung (1996).

and early delay is that a subpopulation of neurons active during stimulus presentation undergoes synaptic enhancement through short-term synaptic plasticity (STP). Upon stimulus removal, synaptic enhancement decays and the system takes some time to settle into a stable state (i.e., an attractor state). These dynamics during stimulus encoding are nevertheless also consistent with working memory using a stable code and without taking STP into account. As implemented by Murray et al. (2017) in a theoretical model of neurons without synaptic plasticity, an external stimulus may induce a network activity pattern that does not match the one that will eventually be stabilized in the delay period. From a dynamic systems perspective, external stimulation puts the system close to the attractor that represents the corresponding cue location, but not exactly there (Fig. 1A). This results in some dynamic activity patterns while the system moves toward the attractor (Fig. 1B), until it reaches it and settles for the remainder of the delay period (Fig. 1C).

Regardless of the underlying mechanism, one behavioral prediction is that during a working memory task subjects should be sensitive to distractors during the dynamic, unstable period, but not during the delay period when the system is settled into a stable state. In fact, a study in monkeys (Pasternak and Zaksas, 2003) and another in humans (Vogel et al., 2006) found that masking impaired working memory performance during the early delay period (~200 ms), but had no effect later on (after ~500 ms). In more recent human studies (van Lamsweerde and Johnson, 2017), it was found that applying transcranial

magnetic stimulation over the occipital cortex decreased working memory precision only during the early delay period (~200 ms), again having no influence later on. Spaak et al. (2017) now provide a neural correlate of these earlier, intriguing behavioral findings.

A recent fMRI study by Sprague et al. (2016) seems, however, to contradict the idea of maintenance through a stable code. In that study, a retro cue, introduced during the delay of a two-item working memory task, selectively increased information about the uncued feature of the cued item. This was interpreted as information recovery from a hidden synaptic trace. However, Watanabe and Funahashi (2014), in a data set reanalyzed by Spaak et al. (2017), found a similar effect that supports a different interpretation. After a dual attention-and-working memory task was reduced to a single working memory task, spiking activity of prefrontal neurons increased and with it the information about the memorized stimulus. The fact that spiking activity did not stop carrying stimulus information in either study is critical to render synaptic working memory unnecessary. Indeed, in a task similar to that used in the study by Sprague et al. (2016), Cisek and Kalaska (2005) recorded single units from the monkey premotor cortex and also found that information about a memorized location increased after retrocuing it. In line with the findings of Sprague et al. (2016), neurons keeping information about the then irrelevant location decreased their activity, suggesting that neural populations holding different locations in working memory might inhibit each other. In fact, Cisek (2006) later

modeled this information increase through a decrease in mutual inhibition between two neural population without taking any plasticity mechanism into account. While reactivation from a synaptic trace is still a possible cause of the information increase reported in the studies by Spaak et al. (2017) and Sprague et al. (2016), there is an alternative explanation under the framework of competing neural populations (Cisek, 2006).

The findings of Spaak et al. (2017) also call for caution when interpreting the failure to detect stimulus information in noninvasive human recording signals. Two independent studies (Rose et al., 2016; Wolff et al., 2017), reported that if an item previously stored in working memory was outside the focus of attention during the delay, modern decoding methods failed to detect stimulus category information in BOLD or EEG signals. When the memory was brought back to the focus of attention, a significant amount of information was again detectable using the same methods. These findings were interpreted as evidence in favor of the synaptic hypothesis. Critically, one limitation of these studies is the lack of direct access to single-unit activity; instead, they rely on signals that are an average of large neural populations (Dubois et al., 2015), orders of magnitude larger than the selectivity clusters found in cortical areas associated with working memory (e.g., PPC and PFC; Masse et al., 2017). By analyzing single-unit activity in a conceptually similar task, Spaak et al. (2017) overcame this limitation and provided evidence in favor of an alternative interpretation. Specifically, to solve a dual task, monkeys had to memorize a cued location while attending

somewhere else. Much like in studies by Rose et al. (2016) and Wolff et al. (2017), when the memory was outside the focus of attention, neurons exhibited a decrease in stimulus selectivity. Nevertheless, even under the most difficult attention conditions, persistent activity was still significantly above baseline through the delay and carried stimulus information. This point can be strengthened by results of the cross-temporal decoding analysis on spiking activity in Spaak et al. (2017): as in the other two data sets, the mnemonic code was generalizable and remained stable throughout the delay period when monkeys were actively attending elsewhere. Because information was never absent from spiking activity in PFC, synaptic working memory mechanisms once again appear unnecessary.

In conclusion, Spaak et al. (2017) offer new evidence that is relevant for the ongoing debate on the stability of working memory code. Their findings suggest reinterpretation of recent influential findings (Rose et al., 2016; Sprague et al., 2016; Wolff et al., 2017). Furthermore, the work supports the hypothesis that a stable code maintains items in working memory. Nevertheless, the early delay period—when memories are still being encoded and susceptible to distractors—is characterized by transient dynamics, which are characteristic of a dynamic code. These findings highlight the need for the development of theoretical models that integrate these two hypotheses, which are often regarded as being mutually exclusive.

## References

- Cisek P (2006) Integrated neural processes for defining potential actions and deciding between them: a computational model. *J Neurosci* 26:9761–9770. [CrossRef Medline](#)
- Cisek P, Kalaska JF (2005) Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45:801–814. [CrossRef Medline](#)
- Druckmann S, Chklovskii DB (2012) Neuronal circuits underlying persistent representations despite time varying activity. *Curr Biol* 22:2095–2103. [CrossRef Medline](#)
- Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* 35:2791–2802. [CrossRef Medline](#)
- Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. *Science* 173:652–654. [CrossRef Medline](#)
- Goldman MS (2009) Memory without feedback in a neural network. *Neuron* 61:621–634. [CrossRef Medline](#)
- Kamiński J, Sullivan S, Chung JM, Ross IB, Mamelak AN, Rutishauser U (2017) Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat Neurosci* 20:590–601. [CrossRef Medline](#)
- Leavitt ML, Mendoza-Halliday D, Martinez-Trujillo JC (2017) Sustained activity encoding working memories: not fully distributed. *Trends Neurosci* 40:328–346. [CrossRef Medline](#)
- Masse NY, Hodnefield JM, Freedman DJ (2017) Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *J Neurosci* 37:6098–6112. [CrossRef Medline](#)
- Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. *Science* 319:1543–1546. [CrossRef Medline](#)
- Murray JD, Bernacchia A, Roy NA, Constantiniadis C, Romo R, Wang XJ (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci U S A* 114:394–399. [CrossRef Medline](#)
- Pasternak T, Zaksas D (2003) Stimulus specificity and temporal dynamics of working memory for visual motion. *J Neurophysiol* 90:2757–2762. [CrossRef Medline](#)
- Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, Postle BR (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354:1136–1139. [CrossRef Medline](#)
- Seung HS (1996) How the brain keeps the eyes still. *Proc Natl Acad Sci U S A* 93:13339–13344. [CrossRef](#)
- Spaak E, Watanabe K, Funahashi S, Stokes MG (2017) Stable and dynamic coding for working memory in primate prefrontal cortex. *J Neurosci* 37:6503–6516. [CrossRef Medline](#)
- Sprague TC, Ester EF, Serences JT (2016) Restoring latent visual working memory representations in human cortex. *Neuron* 91:694–707. [CrossRef Medline](#)
- Stokes MG (2015) “Activity-silent” working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci* 19:394–405. [CrossRef Medline](#)
- van Lamsweerde AE, Johnson JS (2017) Assessing the effect of early visual cortex transcranial magnetic stimulation on working memory consolidation. *J Cogn Neurosci* 29:1226–1238. [CrossRef Medline](#)
- Vogel EK, Woodman GF, Luck SJ (2006) The time course of consolidation in visual working memory. *J Exp Psychol Hum Percept Perform* 32:1436–1451. [CrossRef Medline](#)
- Watanabe K, Funahashi S (2007) Prefrontal delay-period activity reflects the decision process of a saccade direction during a free-choice ODR task. *Cereb Cortex* 17 [Suppl 1]:i88–i100. [CrossRef Medline](#)
- Watanabe K, Funahashi S (2014) Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat Neurosci* 17:601–611. [CrossRef Medline](#)
- Wolff MJ, Jochim J, Akyürek EG, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci* 20:864–871. [CrossRef Medline](#)
- Zylberberg J, Strowbridge BW (2017) Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. *Annu Rev Neurosci* 40:603–627.