Dynamic population coding and its relationship to working memory

Ethan M. Meyers
Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, Massachusetts, and School of Cognitive Science, Hampshire College, Amherst, Massachusetts

Submitted 3 April 2018; accepted in final form 4 August 2018

Meyers EM. Dynamic population coding and its relationship to working memory. J Neurophysiol 120: 2260–2268, 2018. First published September 12, 2018; doi:10.1152/jn.00225.2018.—For over 45 years, neuroscientists have conducted experiments aimed at understanding the neural basis of working memory. Early results examining individual neurons highlighted that information is stored in working memory in persistent sustained activity where neurons maintained elevated firing rates over extended periods of time. However, more recent work has emphasized that information is often stored in working memory in dynamic population codes, where different neurons contain information at different periods in time. In this paper, I review findings that show that both sustained activity as well as dynamic codes are present in the prefrontal cortex and other regions during memory delay periods. I also review work showing that dynamic codes are capable of supporting working memory and that such dynamic codes could easily be “readout” by downstream regions. Finally, I discuss why dynamic codes could be useful for enabling animals to solve tasks that involve working memory. Although additional work is still needed to know definitively whether dynamic coding is critical for working memory, the findings reviewed here give insight into how different codes could contribute to working memory, which should be useful for guiding future research.

INTRODUCTION

Working memory refers to the ability to maintain and manipulate a limited amount of information over a period of seconds (Aben et al. 2012; Baddeley 2012; Cowan 2008). Early work examining firing rates of individual neurons found that neurons in the prefrontal cortex (PFC) and other brain regions had “persistent sustained activity,” where neurons maintained elevated firing rates during delay periods (Fuster 1973; Fuster and Alexander 1971; Kubota and Niki 1971), and that neurons had “persistent selectivity,” where the activity was selective for stimuli that needed to be remembered (Funahashi et al. 1993). Many subsequent studies also found similar persistent sustained activity in the PFC and other brain regions (for reviews, see Leavitt et al. 2017; Riley and Constantinidis 2016). These findings led to the idea that persistent sustained activity is the neural basis of working memory (Nee and D’Esposito 2018), and much work has been devoted to understanding the mechanisms that underlie such sustained activity (Major and Tank 2004; Zylberberg and Strowbridge 2017) and to building computational models that exhibit persistent states (Brunel 2003; Constantinidis and Wang 2004; Durstewitz et al. 2000; Wang 2001; Wimmer et al. 2014).

Although sustained activity might be the neural basis of working memory, recent studies, often using population analysis methods, have also emphasized that many neurons show patterns of neural activity that are selective for only short periods of time during longer memory delay periods. Thus it is possible that other forms of neural activity, or even synaptic changes (Mongillo et al. 2008; Stokes 2015), are the fundamental mechanisms enabling working memory. In this paper, I review studies that have shown that dynamic neural activity is associated with working memory. I then discuss a few additional results that give further insight into whether dynamic codes could be sufficient for supporting working memory and how such codes could be “readout” by downstream brain regions. Finally, I discuss possible reasons why information in working memory might be stored in dynamic population codes. I focus here on studies that recorded action potentials from individual neurons because these data contain the temporal and spatial resolution to examine neural coding [see D’Esposito and Postle (2015) for a review that focuses more on functional MRI and EEG studies]. I hope the review of these results will help guide future research that will lead to a more precise
“algorithmic” explanation for how these types of codes contribute to behavior.

A REVIEW OF DYNAMIC CODING ASSOCIATED WITH WORKING MEMORY

As discussed in the INTRODUCTION, early findings on the neural basis of working memory focused on persistent sustained activity (Fuster 1973; Fuster and Alexander 1971; Kubota and Niki 1971; Miller et al. 1996). Although these studies showed that there was a range of activity profiles in different neurons, such as neurons that responded to the onset of a visual cue (Chafee and Goldman-Rakic 1998; Fuster 1973; Goldman-Rakic 1996), many of these studies highlighted the fact that neurons had tonic firing rates during delay periods (Fig. 1A). One notable exception to the description of tonic delay period activity was discussed in a study by Batuev et al. (1980) that found that most neurons in the lateral PFC of macaques were selective for only shorter periods of time during a memory delay period, leading to what they described as “successive ‘relay-race’ involvement of different neuronal populations in order to sustain a prolonged activation of a small group of ‘summator’ units”; however, this paper did not attract much attention.

Around 2000, several studies analyzed single-neuron activity in more detail and began to emphasize neural correlates of working memory that were different from tonic sustained activity (Brody et al. 2003a, 2003b; Romo et al. 1999; Shafi et al. 2007; Warden and Miller 2007; Zaksas and Pasternak 2006). For example, analyses by Zaksas and Pasternak (2006) found that most motion direction neurons in the middle temporal visual area and PFC were only transiently selective during a memory delay period, with most neurons maintaining their selectivity for around 150–300 ms out of the 1,500-ms delay period (Fig. 1A). Likewise, Brody et al. (2003a) found that whereas some PFC neurons were selective to tactile stimuli throughout the delay period, other neurons were only selective for shorter periods of time (Brody et al. 2003a, 2003b; Romo et al. 1999).

Neural population-level analyses also more clearly revealed “dynamic population coding” phenomena associated with working memory tasks that were different from persistent sustained activity (Fig. 2). For example, a study by Baeg et al. (2003) showed that it was possible to decode the direction a rat intended to move during a delay period using only neurons in the PFC that did not have continuous high levels of activity and thus suggested that working memory might be coded by sequential activation of different neurons rather than sustained activity. Additionally, “cross-temporal decoding” analyses, where a pattern classifier is trained with data from one time period and then tested with data from a different time, clearly showed that information was contained in different patterns of activity at different time periods (see Fig. 3 and Barak et al. 2010; Crowe et al. 2010; Meyers et al. 2008, 2012; Nikolić et al. 2009; Pastalkova et al. 2008). For example, Meyers et al. (2008) first used temporal cross-decoding to show there was dynamic coding of category information PFC and the inferior temporal cortex during a working memory task, and Pastalkova et al. (2008) used cross-correlation analyses to show dynamics in rat hippocampus also in a memory delay task. Likewise, Nikolić et al. (2009) and Crowe et al. (2010) used versions of temporal cross-decoding to show dynamic representations in delay periods for activity in V1 and in parietal cortex, respectively. Knowledge about dynamic population coding became more widespread after the publication of Stokes et al. (2013), who used temporal cross-correlation analyses to show dynamic population coding in the PFC during a memory delay in a paired association task, and with the publication of several review papers that discussed dynamic coding and its relationship to working memory (King and Dehaene 2014; Sreenivasan et al. 2014; Stokes 2015).

Fig. 1. Single-neuron results emphasizing sustained delay activity and dynamic selectivity. A: figure from a review paper by Goldman-Rakic (1996) characterizing different types of neural responses seen in prefrontal cortex (PFC) activity. Results show that neurons respond to cue onset (C; top plot) and respond to saccade onset (R; bottom plot) and that they often have sustained tonic activity during the delay period (D; middle plot). [Reprinted from Goldman-Rakic PS. Regional and cellular fractionation of working memory. Proc Natl Acad Sci USA 93: 13473–13480, 1996, with permission. Copyright (1996) National Academy of Sciences, U.S.A.] B: analysis of the direction selectivity of 67 neurons in the PFC showing that many neurons are selective for relatively short periods of time during a delay period. Blue regions indicate times at which activity associated with a preferred sample were reliably higher than those associated with an antipreferred sample. Red regions indicate the opposite relationship, and gray regions show time where the signal was not significantly direction selective. Shown is that many neurons only appear selective for short periods of time in the delay period. #, Number. [Results are adapted from Zaksas and Pasternak (2006) with permission.]
Beyond the analyses of neural spiking activity, cross-temporal decoding analyses of electroencephalography (EEG) and magnetoencephalography (MEG) data also revealed dynamic activity in those data, although changing patterns of EEG/MEG activity could be due to information traveling between brain regions rather than a dynamic code within a population of neurons (Carlson et al. 2011, 2013; Cichy et al. 2014; Isik et al. 2014; King and Dehaene 2014). Additionally, analyses of local field potentials in monkeys have revealed burst during memory delay periods, which appear to be associated with dynamic codes (Lundqvist et al. 2016). Finally, Harvey et al. (2012) used two-photon imaging of mouse posterior parietal neurons and observed that most neurons were selective for relatively short periods of time and showed that neurons with different windows of selectivity were intermingled throughout this brain region.

It is important to note that the difference between studies that found stationary and dynamic codes is not due to different data analysis methods, since several studies using the temporal cross-decoding analyses have found stationary codes as well (see Fig. 4 and King and Dehaene 2014). Currently, it is not clear why some studies find dynamic code and others find stationary codes, although there seems to be a general trend that stationary codes are present when simple stimulus attributes need to be remembered (Meyers et al. 2017; Murray et al. 2017; Spaak et al. 2017; Zhang et al. 2011), whereas dynamic codes are present when more complex stimulus transformations are needed to solve a task (Crowe et al. 2010; Meyers et al. 2008, 2012; Stokes et al. 2013). Additionally, it appears that there might be more persistent sustained selectivity in the anterior-dorsal region of the prefrontal cortex, whereas there is...
a more dynamic representation in the more posterior-dorsal regions (Leavitt et al. 2017; Meyers et al. 2017). If this is the case, then perhaps the stationary codes in the anterior-dorsal PFC are acting more like short-term memory and the dynamic codes in the posterior dorsolateral PFC region are more involved in manipulating information.

**IS DYNAMIC ACTIVITY SUFFICIENT FOR MAINTAINING INFORMATION IN WORKING MEMORY?**

The previous sections showed that both dynamic and stationary coding are robust phenomena that have been reported in many studies of working memory. However, definitively understanding the degree to which dynamic and stationary codes contribute to working memory is challenging. Nevertheless, it is possible to do analyses that can assess whether neurons that have dynamic coding properties are sufficient for maintaining information in working memory and whether such dynamic codes are “biologically plausible” in the sense that the information contained in dynamic codes could be utilized by downstream neurons to contribute to behavior. If it turns out that dynamic codes do not contain sufficient information for behavior or that information cannot be extracted from dynamic codes in a biologically plausible way, then clearly such coding would not be a viable candidate for the neural basis of working memory.

To assess whether dynamic neurons contain sufficient information to support working memory, Baeg et al. (2003) did an analysis where they found all selective neurons that showed continuous activity in the delay period (where continuous activity was defined as neurons that had higher average firing rates in 4 bins in the delay period compared with other time periods). They then did a decoding analysis with these continuously active neurons removed and showed that the decoding accuracy level was similar using only the remaining dynamic neurons. Similarly, Meyers et al. (2008) did a decoding analysis where they eliminated all selective neurons based on the training data at 1 time point (which I will call $t_{elim}$ here). This removed all neurons with persistent sustained selectivity from the population and removed all neurons with short dynamic periods of selectivity around $t_{elim}$. As expected, when the classifier was trained and tested with data from $t_{elim}$, the classification accuracy was around chance. However, when the classifier was trained and tested at other points in time with the same neurons from $t_{elim}$ removed, the classification accuracy was once again high because there were different neurons present at each point in time on a different test set of data (indicated on the x-axis). C: cross-temporal decoding plot where the classification accuracy is indicated by the color bar in the bottom of the gray region. In this example, a high classification accuracy is only seen when the classifier is trained and tested at the same point in time, as can be seen by the fact that there is a colorful diagonal in the plot. This means that a dynamic code is present since a classifier trained at 1 point in time does not have a good performance when tested at a different point in time. If a stationary code was present, then one would expect to see a solid yellow/red colored square indicating good classification performance regardless of the time when the classifier was trained (see Fig. 4B). Black, vertical lines correspond to sample stimulus onset, sample stimulus offset, and match stimulus onset. [Results in this plot are adapted from Meyers et al. (2008).]
integrate information over relatively long time scales of ~2 s, information from the population could still be extracted. Ad-
outside of the window of time when a neuron was selective did
hypothesis of neural coding was that including activity
sliding 150-ms bins. With the use of this approach, it was
was trained to learn classification weights using neural activity
set of weights for all time points. More specifically, a classifier
was shown during the 1st delay period in a delayed
were selective for short periods of time (Fig. 6). Meyers et al. (2009) examined a population of neurons that showed
dynamic changes might indeed be occurring (Stokes 2015), others have instead tried to assess whether it is possible
to extract information from dynamic codes using a potentially
“fixed set” of synaptic weights (Meyers et al. 2009; Murray et al. 2017). For example, Meyers et al. (2009) examined a population of neurons that showed
dynamic coding properties and tried to decode information
from this population using a classifier that had one single fixed
set of weights for all time points. More specifically, a classifier
was trained to learn classification weights using neural activity
that was averaged over the whole 2.5 s of an experimental trial.
The classifier was tested (on a different set of trials) using sliding 150-ms bins. With the use of this approach, it was
found that most of the information from the population could be extracted using a fixed set of weights even though many
neurons were only selective for short periods of time (Fig. 6B). The reason why most of the information could be extracted
using a fixed set of weights was because including activity
outside of the window of time when a neuron was selective did
not add enough noise to disrupt the information that was
contained in the rest of the population. Thus most of the
information from the population could still be extracted. Ad-
additionally, these results imply that even if downstream neurons integrate information over relatively long time scales of ~2 s,
most of the information from the upstream dynamic population is still available.

HOW COULD DYNAMIC CODES CONTRIBUTE TO WORKING MEMORY AND BEHAVIOR?

Taken together, these results show that dynamic codes have
sufficient information to support working memory and that
information contained in dynamic codes can be extracted by
downstream neurons in a biologically plausible way. However,
from a computational point of view, a question that arises is
why would the brain use a dynamic code to hold information
rather than a stationary code? In particular, it makes intuitive
sense that persistent sustained selectivity would be the neural
basis of working memory since encoding items in working
memory would merely consist of retaining the original input
activity and it would be trivial to retrieve the memory trace at
any time point (Wimmer et al. 2014), whereas having dynamic
codes seems to be more complicated. Indeed, it has been
suggested that the real basis of working memory is, in fact,
persistent activity and that these dynamic coding properties
might be there to support the maintenance of persistent activity
or purely epiphenomenal (Riley and Constantinidis 2016),
although other researchers have taken the opposite perspective
and have proposed that dynamics are the main basis of working
memory (Lundqvist et al. 2016). Although it is currently hard
to know whether either of these perspectives is correct, there
are several theories that can explain why dynamic coding is
useful. These theories can be broadly divided into two catego-
ries: namely, 1) that dynamic codes are used to encode informa-
tion about time or 2) that dynamic codes are part of a
sequence of processing steps needed to solve particular tasks.

J Neurophysiol • doi:10.1152/jn.00225.2018 • www.jn.org
The first class of theories, that dynamic coding is related to how the brain keeps track of time, has been put forward by several research groups. For example, researchers studying the hippocampus have proposed that these dynamic neurons are “time cells” that are used for recalling temporal memories (Eichenbaum 2014; Pastalkova et al. 2008). Similarly, researchers examining a task where monkeys needed to produce responses at particular time intervals found neurons in the medial frontal cortex that were tuned to the timing of the task, again suggesting that the dynamics are involved in keeping track of time (Wang et al. 2018). If this is the case, then the dynamic coding seen in working memory could merely be a result of information about time that is being multiplexed on top of memory signals containing information about other variables. Many studies have shown that PFC activity simultaneously contains information about multiple behaviorally relevant variables, and it has been suggested that the complex coding seen in PFC might be a useful feature for separating all possible stimulus configurations (Rigotti et al. 2013), which could include information about time as well.

If dynamic population codes are a mechanism for keeping track of time, then it seems that (at least in the PFC) information about time is contained relative to particular events of interest rather than on an absolute scale (Brody et al. 2003a; Wang et al. 2018). In particular, Brody et al. (2003a), analyzed dynamic activity in PFC in an experiment where they compared neural activity in a 3-s delay period to neural activity in a 6-s delay period (the experiment was done in a blocked design so that the monkey could quickly adapt its behavior to the new delay period length). Their results showed that the dynamic neural activity appears to evolve slower when the delay period was longer, with neural activity in the 6-s delay period being well-modeled by a stretching of the neural activity seen in the 3-s delay period. Thus, from a population perspective, the same dynamic codes would occur at the same relative position in the delay period regardless of the duration of the delay period, and so information about time would be relative to the start and end of the delay period.

The second class of theories, that dynamic codes are related to different steps needed to solve a task, has been been formulated in several related ways. One formulation suggests that the temporal dynamics are actually due to neurons coding different information at different points in time; however, because information about several variables are often confounded in a typical experiment, it could (falsely) appear that information about one single variable is being coded dynamically. For example, imagine a delayed match-to-sample task in which a monkey needs to remember a location of a cue stimulus and then saccade to a particular target if the second stimulus is at the same location. During the memory delay period in this task, a population of neurons could initially code information about the location of the first cue stimulus. However, then, later in the delay, the population could contain a signal related to triggering a motor response if the second stimulus is at the matching location. Since information about the first stimulus and triggering a motor response in anticipation of a second stimulus are confounded, then it might look like information about the first stimulus is stored in memory by a dynamic code, but really the dynamics are due to the fact that different information is being coded at different times in a trial. This interpretation is also consistent with the fact that more dynamic coding is seen in experiments that involve more complex transformation of stimulus attributes, which could be due to the neural population representing information in intermediate computational steps (Crowe et al. 2010; Meyers et al. 2008, 2012; Stokes et al. 2013). Such sequential steps of processing are also similar to the description of the PFC as an area that is responsible for integrating information over time (Fuster 2001).

A related formulation is the notion that neural processing is best described as a dynamic system that goes through a sequence of hidden states to solve particular tasks (Shenoy et al. 2013). In particular, dynamic representations have been seen in dorsal premotor cortex when monkeys are preparing to make motor movements, and monkeys’ reaction times can be predicted based on position and rate of change of dynamic coding activity, which shows there is a strong association between neural dynamics and behavior (Afshar et al. 2011). This dynamic system explanation is similar to the “confounding” explanation described in the previous paragraph in the sense that both explanations describe states that are hidden from the experimentalists, although the dynamic system perspective might not try to assign meaning to different internal states that could be continuously changing.

The concept that dynamic coding is related to different computational steps also makes sense from the perspective that it would be impossible to solve a particular task if the same pattern of neural activity that initially encoded the stimulus was
also used for maintaining the stimulus in working memory. For example, consider again the delayed match-to-sample task. In this task, if the same pattern of neural activity was used to represent both the first and second stimuli, then information about the first stimulus would be overwritten when the second stimulus was shown, which would make it impossible to tell whether the two stimuli matched. Conversely, if dynamic coding was used, then it would be possible to compare the same stimulus across time because different patterns of activity would represent the first and second occurrences of the stimulus. Thus a dynamic coding scheme seems necessary, to some degree, to solve particular tasks.

Finally, it should be noted that it is also possible that the dynamic coding might be epiphenomenal and is not enabling an animal to perform any additional behaviors. As described above, dynamic population activity can be readout by downstream neurons that integrate over the full delay period without much loss of information (Meyers et al. 2009), so downstream neurons might indeed be trying to extract constant information from the population in the face of these dynamics rather than having these dynamic codes as a useful feature that contributes meaningfully to particular behaviors, such as keeping track of time (Murray et al. 2017).

CONCLUSIONS

In this paper, I have reviewed literature showing that both persistent sustained activity and dynamic population coding are associated with working memory. Furthermore, I have described studies that have shown that the amount of information stored in dynamic codes appears to be sufficient to solve working memory tasks and that this information could be extracted from a neural population by downstream neurons using biologically plausible mechanisms. Finally, I explored possible explanations for why dynamic codes might be present and how they could potentially contribute to solving tasks.

Throughout this paper, I have described neural coding as being either persistent or dynamic. However, it should be pointed out that on an individual neuron level there often seems to be a continuum of time scale that neurons encode information (Harvey et al. 2012; Meyers et al. 2008, 2012; Zakas and Pasternak 2006), although there might also be systematic differences in time scales as well between different brain regions or even different animals (Leavitt et al. 2017; Mendoza-Halliday et al. 2014; Meyers et al. 2017; Murray et al. 2014; Runyan et al. 2017). Also, it should be mentioned that estimating the time scale that a neuron maintains information could be imprecise in many studies because information might be present at different points in time on different trials. Thus, when data are analyzed across trials, it could appear that information is maintained for longer periods of time than it really is on any single individual trial. By analyzing larger populations of neurons that are recorded simultaneously, it might be possible to track information content changes within a single trial, which could better help estimate how long neurons typically maintain information for.

In summary, although many questions remain about how persistent activity and dynamic coding contribute to working memory, it is clear that both types of coding are commonly seen phenomena that carry large amounts of information useful for solving tasks. Although much work remains, the facts that recording technologies continue to improve and that data analysis toolboxes exist that make it relatively easy to assess dynamics in neural spiking and EEG/MEG activity (Gramfort et al. 2013; Meyers 2013; Tadel et al. 2011) should lead to a more concrete description of the neural algorithms that underlie particular behaviors in the not too distant future.
ACKNOWLEDGMENTS

I thank Tomaso Poggio for continual support.

GRANTS

This work was supported by the Center for Brains, Minds and Machines, funded by National Science Foundation Science and Technology Center Award CCF-1212126. Additional support came from Adobe, Honda Research Institute USA, and a King Abdullah University of Science and Technology Grant to B. DeVore.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author.

AUTHOR CONTRIBUTIONS

E.M.M. prepared figures and drafted, edited, revised, and approved the final version of the manuscript.

REFERENCES


Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romeo R, Wang XJ. Stable population coding for working memory coexists with heterogeneous...


