Working Memory: Flexible but Finite

Kirsten C.S. Adam[1,2,*] and John T. Serences[1,2,3]
[1]Department of Psychology, University of California San Diego, La Jolla, CA, USA
[2]Institute for Neural Computation, University of California San Diego, La Jolla, CA, USA
[3]Neurosciences Graduate Program, University of California San Diego, La Jolla, CA, USA
*Correspondence: kadam@ucsd.edu

https://doi.org/10.1016/j.neuron.2019.06.025

There are inherent trade-offs between the flexibility and the capacity of working memory, or the ability to temporarily hold information "in mind." In a recent issue of *Neuron*, Bouchacourt and Buschman (2019) present a new model of working memory that demonstrates how coordinated activity between specialized sensory networks and flexible higher-order networks may support these competing constraints.

Working memory refers to our ability to temporarily hold relevant information "in mind." Two key features of working memory are its flexibility and its starkly limited capacity. Working memory is flexible enough to represent novel combinations of visual features but limited to representing only a few chunks of information at once. For example, if you see an unfamiliar flower on a hike, you can hold a precise image of its color and petal shape in working memory while you search for a match in your wildflower guide. But, if you come across a dozen unique flowers and want to ID them all, you cannot simultaneously hold them all in mind. Your memory for the distinct features of each flower will become less precise, and many flowers will be forgotten altogether. Thus, you might strategically and flexibly encode only a subset of the available information (e.g., just the colors or shapes of three flowers) each time you check your guide.

For decades, there has been great interest in modeling the neural codes that support working memory, particularly because working memory is disrupted in many clinical disorders (e.g., Schizophrenia, Parkinson's, depression). However, flexibility and capacity have rarely been modeled together. In a recent issue of *Neuron*, Bouchacourt and Buschman (2019) present a new model that captures both core aspects of working memory. This model, which we'll refer to as the "coordinated network model," shows how structured sensory networks and a flexible, higher-order network can work together to support and constrain working memory.

Studies of working memory's capacity often take working memory's flexibility for granted. For simplicity's sake, a common strategy is to measure working memory capacity using simple, controlled stimuli from a single feature space (e.g., colored squares). Because of this, many models have considered how competitive interactions in networks tuned for a particular feature space, such as color, will lead to behavioral capacity limits (e.g., Compte et al., 2000). Although such models can explain behavioral capacity limits when remembering stimuli in a common feature space, they have a more difficult time explaining how limits would arise when flexibly remembering items that are novel or from multiple feature spaces. The key insight raised by Bouchacourt and Buschman (2019) is that competitive interactions that lead to capacity limits may actually arise via less specialized, higher-order processing layers rather than via direct competition within sensory networks. Of course, these possibilities are not mutually exclusive, and competition occurs at multiple levels of processing. But, an important test is whether interference in a flexible, domain-general processing layer is sufficient to generate capacity limits.

To create a model that is both flexible and can account for the capacity limits observed in behavior, Bouchacourt and Buschman (2019) combined a structured, sensory-specific network (sensory layer) with a flexible, random network (random layer) in a two-layer network model. These two layers roughly map onto key characteristics of different areas of the brain. Early visual areas (e.g., V1–V4) are relatively specialized, with small receptive fields and neurons that are selective for a particular feature. By contrast, higher-order control regions like pre-frontal cortex are thought to flexibly represent arbitrary stimulus combinations, rules, and abstract ideas. In their model, the investigators implement these characteristics by linking highly structured sensory networks with higher-order networks via random connections. Each sensory network (one per encoded memory item) is structured in that an input of one color (e.g., red) leads to systematic excitation and inhibition as a function of color similarity; neurons tuned to similar colors are partially excited and those tuned to dissimilar colors are inhibited. Importantly, there is no direct competition between the sensory populations that encode each remembered item; instead, competition occurs in the random network. Because of this independence, the coordinated network model yields the prediction that competition will occur even when items are drawn from different feature spaces such as color and orientation (e.g., Fougnie et al., 2010). In contrast to connections within the sensory network, connections between the random network and the sensory networks are unstructured with respect to color. Memories are maintained by bidirectional, reciprocal connections between the random and sensory layers.

The storage of memories via random connections between structured sensory networks and higher-order networks provides a flexible mechanism for representing information of any type. However, it comes at a cost. For example, if we hold the colors of two flowers in mind rather than one, our memory for each color will be less precise. This cost can be predicted by the coordinated network model.

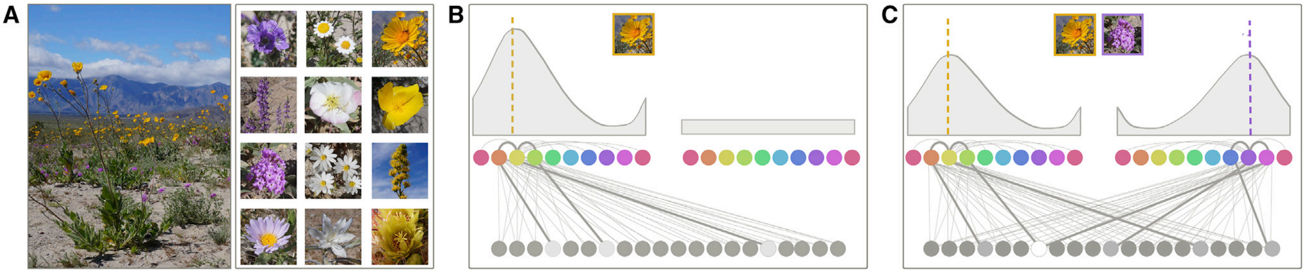

Figure 1. Holding Precise Memories in Mind

(A) Working memory is how we temporarily hold information in mind. For example, while hiking, we may spot a flower that we want to identify. When we look away from the desert scene to our flower guide, we can hold a precise representation of the flower's color in mind.

(B and C) In the coordinated network model, each item is input to a separate "sensory network." When holding just one item in mind (B), we can remember it precisely, with few errors. When storing multiple items (C), representations begin to interfere in the random network. Thick gray lines represent excitatory connections; thin gray lines represent inhibitory. Note: this figure is illustrative and does not reproduce exact model parameters.

When asked to remember one item, the modeled "neurons" selective to the color of the item in one sensory network are partially excited; these neurons in turn excite a random subset of neurons in the random network (Figure 1). A second remembered item will excite a different, but potentially overlapping, subset of neurons in the random network. Thus, on average, neurons in the random network have a greater sustained firing rate when more items are remembered, as observed in prefrontal cortex (Fuster, 1973). But, the excitation from a given neuron in the sensory network to the random network is balanced by weak inhibitory connections to all other neurons in the random network. This balanced inhibition means that the second item is *also* more likely to inhibit the neurons in the random network that are excited by the first item, thus weakening the recurrent feedback to the first item's sensory network. This has two key consequences. First, it results in divisive-normalization-like attenuation of tuning selectivity in the random and sensory networks. Consistent with empirical work, the coordinated network model predicts that the neural response to two items remembered together is a sublinear combination of the response to each item remembered on its own (Heeger, 1992). Second, this weakened recurrent feedback can lead to drift in the neural representation away from the true presented color, reducing behavioral precision.

A key ongoing debate is whether we actively maintain only a subset of items from large arrays and forget the rest (e.g., store 3 of 8 items) or if we instead maintain very imprecise representations of all items (e.g., van den Berg et al., 2014). The coordinated network model implements a "some-or-none" storage mechanism; if there is sufficient memory-related activity beyond a certain threshold, representation of the memory is sustained throughout the delay (though it may drift and become less precise). If activity is insufficient, the memory collapses to a null attractor state and is lost. This model prediction is consistent with recent behavioral work finding uniformly distributed guess responses for a subset of items from large arrays (Adam et al., 2017). The coordinated network model's some-or-none implementation of storage highlights one way that graded neural codes could yield the complete loss of items when working memory is taxed.

Prior work has shown that widely distributed brain regions participate in working memory (for review, Christophel et al., 2017), but there has been substantial debate about which of these codes is most necessary or even solely necessary for supporting working memory. Some work has emphasized that sensory areas are well suited to maintaining extremely precise representations, whereas other work has argued that sensory codes may not be useful when interrupted by new visual inputs (Bettencourt and Xu, 2016; but see Rademaker et al., 2019). Bouchacourt and Buschman's coordinated network model highlights that the interaction between multiple regions, rather than a single region or representation, may be key to understanding how different areas jointly contribute to supporting working memory.

REFERENCES

Adam, K.C.S., Vogel, E.K., and Awh, E. (2017). Clear evidence for item limits in visual working memory. Cognit. Psychol. *97*, 79–97.

Bettencourt, K.C., and Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat. Neurosci. *19*, 150–157.

Bouchacourt, F., and Buschman, T.J. (2019). A flexible model of working memory. Neuron *103*, 147–160.

Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., and Haynes, J.-D. (2017). The distributed nature of working memory. Trends Cogn. Sci. *21*, 111–124.

Compte, A., Brunel, N., Goldman-Rakic, P.S., and Wang, X.J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb. Cortex *10*, 910–923.

Fougnie, D., Asplund, C.L., and Marois, R. (2010). What are the units of storage in visual working memory? J. Vis. *10*, 27–27.

Fuster, J.M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. J. Neurophysiol. *36*, 61–78.

Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. Vis. Neurosci. *9*, 181–197.

Rademaker, R.L., Chunharas, C., and Serences, J.T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. Nat. Neurosci. https://doi.org/10.1038/s41593-019-0428-x.

van den Berg, R., Awh, E., and Ma, W.J. (2014). Factorial comparison of working memory models. Psychol. Rev. *121*, 124–149.