

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Constraining Theories of Working Memory with Biophysical Modeling

Nicholas Myers and George Wallis

Department of Experimental Psychology and Oxford Centre for Human Brain Activity, Oxford University, Oxford, OX3 7JX, United Kingdom
Review of Wei et al.

A debate about the basic unit of representation in working memory (WM) has recently been prominent in the field. Some argue that WM consists of a limited number of “slots.” These have been equated with object representations (Luck and Vogel, 1997), but could also arise if the neural coding scheme for WM was fundamentally quantized (e.g., by distinct phases of slow oscillations; cf. Jensen and Lisman, 1998). In either case, when all memory slots are occupied, no new information can enter WM. Others have argued that there is a continuous neural resource available for WM, which is flexibly shared between the items being represented (Bays and Husain, 2008). In this case, the amount of neural resource dedicated to one item decreases as the number of stored items increases, leading to a decrease in the precision of information about each item. Behavioral evidence has been presented in favor of both theories, and while individual studies tend to stand strongly in favor of one position or the other, there is still no consensus in the field. Neuronal modeling might be able to provide a richer account of slot-like and resource-like behavioral effects.

In a recent paper in *The Journal of Neuroscience*, Wei et al. (2012) describe a

model with which they aim to explain both slot-like and distributed resource-like phenomena as emerging from a single model of the neural substrate of working memory. The authors used a spiking network model intended to capture the cortical storage of continuous visual features (such as orientation or spatial location; Compte et al., 2000). The model is built around a ring of 4096 glutamatergic neurons, with each neuron having a preferred orientation. Neurons are strongly connected with other neurons of similar feature tuning, the connection strength falling off with increasing feature distance. In addition, there is a fixed level of unstructured excitatory connectivity (exciting all neurons equally, regardless of their feature tuning). An unstructured inhibitory net, driven by the excitatory neurons, counterbalances the excitation. Oriented inputs are modeled as brief current injections to a small subset of neurons representing that orientation.

When the model is given a small set of evenly separated inputs, it forms local “bumps” of activity around the excited neurons that persist after the offset of current injection (Wei et al., 2012, their Fig. 2A). These bumps of activity are maintained by local excitation, largely mediated by slow NMDA receptor dynamics (Compte et al., 2000). When bumps form in close proximity, they sometimes merge together, in which case the items' individual orientations are replaced with a representation of the average orientation. Additionally, when there are many bumps

present, the excitatory neural pool representing each bump is smaller than the summed inhibitory activity, and bumps can fade out. This can be interpreted as forgetting. (Wei et al., 2012, their Fig. 2A). The dynamics of bump-merging and bump fade-out give rise to all the capacity limitations described by Wei et al. (2012).

In continuous resource models, there is often an implicit assumption that when a representation is allocated less of the WM resource, it is stored with less precision (“the proportion of resource allocated to each item determines the precision with which it is remembered”; Bays and Husain, 2008). It is worth noting that this is not the case in the model by Wei et al. (2012). For example, in their Figure 3C, for uniform arrays below the capacity of ~6 (i.e., in the absence of bump-merging), the precision of storage is unchanged between 1 and 6 items (Wei et al., 2012, their Fig. 3C).

The veracity of the model's memory is tested by simulating many trials and testing how tightly (and how often) bumps are centered on the presented feature after a delay of 1–10 s. Memory precision declines whenever bumps begin to merge or fade out. Because merging depends on the proximity of bumps, precision dips at lower loads when stimulus arrays are randomly spaced (such that occasionally two stimuli are similar). To echo the behavioral terminology, each bump is like a slot of fixed precision. Rather than some items simply not being allocated a slot when capacity is exceeded (fade-out), neighboring

Received Oct. 8, 2012; revised Nov. 1, 2012; accepted Nov. 2, 2012.

We thank Anna C. Nobre, Mark Stokes, and Christopher Summerfield for helpful comments on the manuscript.

Correspondence should be addressed to George Wallis at the above address. E-mail: george.wallis@psy.ox.ac.uk.

DOI:10.1523/JNEUROSCI.4758-12.2013

Copyright © 2013 the authors 0270-6474/13/330385-02\$15.00/0

discrete representations can merge together. There is no load-dependent gradation of precision within bumps. However, if the arrays are randomly spaced, then as load increases, the fidelity of bumps degrades, simply because of a higher likelihood of bump-merging (Wei et al., 2012, their Fig. 2). For randomly spaced arrays (the norm in human behavioral experiments), the model can therefore give rise to the typical graded fall-off in precision (Bays and Husain, 2008), even below capacity (cf. Zhang and Luck, 2008).

The model presented by Wei et al. (2012) makes a strong prediction that there is no way to independently store the same or very similar points on a feature dimension (they will merge). The authors interpret a recent paper examining the influence of feature similarity on memory performance (Lin and Luck, 2011) as evidence for their merging account, proposing that where two very similar features are stored, the representation of these features will be roughly their average value. This prediction is testable in other behavioral experiments. The model would be falsified by any experiment that found that two similar features could be stored independently without bias (i.e., it could be demonstrated that their representations were unmerged). A caveat is that such experiments must ensure that interactions do not result from crowding effects at encoding (Rosenholtz et al., 2012). This could be mitigated by sequential presentation.

The present model is impressive in its ability to create both slot and resource-like phenomena in a single neural substrate. However, we still have to ask whether it is a plausible neural model for visual memory. Working memory requires gating and readout mechanisms, and mechanisms for binding of features and the coordination of object-like representations. The model presented here does not account for such functions. We suggest that the network presented in this paper is therefore best interpreted as a feature buffer that would be embedded in a larger functional and neural framework *in vivo*. In critiquing the model, we can ask how plausibly it could be incorporated into a wider network, without the need for major changes that would invalidate the specific dynamics described in this paper.

Sensory and parietal cortex are known

to exhibit spatiotopic tuning (Swisher et al., 2007). Adding a spatial dimension to the model by Wei et al. (2012) would render the model more neurally plausible, if it were equated with a posterior cortical feature buffer, as we suggest here. Such a 2D model would also better separate items with similar (or identical) feature values in neural space, and thus might prevent bump-merging. In this case, the empirical prediction is that effects of feature and spatial proximity should interact in memory. A recent study in macaques found load effects on visual WM performance for items presented in one hemifield, but not across hemifields (Buschman et al., 2011). There is some evidence for similar effects in humans (Umamoto et al., 2010). Importantly, however, the reduced probability of bump-merging would reduce the model's ability to explain resource-like precision changes.

Could the network model be modified to support the binding of stored features to, or into, object representations? Some models of feature binding in visual short term memory (VSTM) propose that the representation of an item across several cortical areas is maintained by synchronization in the gamma frequency band (Jensen and Lisman, 1998). A test of the viability of the current model could include establishing whether it can maintain persistent stimulus-specific activity that is locked to an ongoing gamma oscillation (and possibly synchronized with persistent activity in other brain areas). Some evidence from a very similar model (Compte et al., 2000) indicates that changing the relative influence of AMPA- and NMDA-mediated signaling could induce oscillations. It would be interesting to test how a synchrony mechanism might interact with the merging of bumps (or possibly prevent merging if populations storing similar items fire out of phase).

As we stress above, the model's explanation of hybrid slot/resource-like behavior depends on bump-merging. This is not the only way in which items could interact in working memory. As Bays and Husain (2008) have suggested, the precision with which a memory is stored might correspond to the proportion of neural resource it has received. Indeed a key tenet of this "implicit" model is that the resource can be flexibly allocated. The Wei et al. (2012) model demonstrates that this kind of increase in precision does not necessarily occur if more neurons with pre-specified feature tuning are recruited.

Neurons in the flexible resource pool might then also require flexible feature tuning, a property not shared by this model, and also unlikely to be found in the labeled lines of the sensory cortex. Such flexible coding may be a property of prefrontal networks (Duncan, 2001).

Much of the behavioral literature on VSTM implicates a model similar to the one presented here for feature storage. By highlighting the potential difficulties and unexpected behavior of such a model, computational studies can help advance the field, forcing us to think more explicitly about the nature of neural resources for memory.

References

- Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321:851–854. [CrossRef Medline](#)
- Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. *Proc Natl Acad Sci U S A* 108:11252–11255. [CrossRef Medline](#)
- Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10:910–923. [CrossRef Medline](#)
- Duncan J (2001) An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci* 2:820–829. [CrossRef Medline](#)
- Jensen O, Lisman JE (1998) An oscillatory short-term memory buffer model can account for data on the Sternberg task. *J Neurosci* 18:10688–10699. [Medline](#)
- Lin PH, Luck SJ (2009) The influence of similarity on visual working memory representations. *Visual Cognit* 17:356–372. [CrossRef](#)
- Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390:279–281. [CrossRef Medline](#)
- Rosenholtz R, Huang J, Ehinger KA (2012) Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision. *Front Psychol* 3:13. [Medline](#)
- Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC (2007) Visual topography of human intraparietal sulcus. *J Neurosci* 27:5326–5337. [CrossRef Medline](#)
- Umamoto A, Drew T, Ester EF, Awh E (2010) A bilateral advantage for storage in visual working memory. *Cognition* 117:69–79. [CrossRef Medline](#)
- Wei Z, Wang XJ, Wang DH (2012) From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *J Neurosci* 32:11228–11240. [CrossRef Medline](#)
- Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. *Nature* 453:233–235. [CrossRef Medline](#)