intentional action. However, an additional signal strength analysis revealed that both processes interact with one another. In particular, neural activity within the paramedial frontal cortex was increased for internal as compared to external timing, but only when action selection was specified externally. Combined, these observations challenge the idea of a unitary control system and underscore the existence of distinct, albeit interdependent, decisional processes that together shape voluntary action (Brass and Haggard, 2008). Of note is that this dependency is in accordance with the argument that aspects from the component processes have to be considered in order for an action and its consequences to be evaluated.

By introducing an innovative experimental paradigm, the work of Krieghoff et al. (2009) has provided a significant step into the understanding of intentional control by specifying dissociable brain regions that deal with component functions. Further research into the dynamics of voluntary action is necessary in order to detail the functional architecture of intentional behavior and its neural correlates. Future studies may elaborate on the inter-regional influences and coordination of the brain areas that make up the network activity, and the means by which context-related factors modify intentional processing within the circuitry.

REFERENCES

Brass, M., and Haggard, P. (2007). To do or not to do: the neural signature of self-control. *J. Neurosci.* 27, 9141–9145.

Brass, M., and Haggard, P. (2008). The what, when, whether of intentional action. *Neuroscientist* 14, 319–325.

Cunnington, R., Windischberger, C., Deecke, L., and Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: a study of event-related fMRI. *Neuroimage* 15, 373–385.

Cunnington, R., Windischberger, C., Robinson, S., and Moser, E. (2006). The selection of intended actions and the observation of others' actions: a time-resolved fMRI study. *Neuroimage* 29, 1294–1302.

Deiber, M. P., Honda, M., Ibañez, V., Sadato, N., and Hallett, M. (1999). Mesial motor areas in self-initiated versus externally triggered movements examined with fMRI: effect of movement type and rate. *J. Neurophysiol.* 81, 3065–3077.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nat. Rev. Neurosci.* 9, 934–946.

Haggard, P., and Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Conscious Cogn.* 12, 695–707.

Krieghoff, V., Brass, M., Prinz, W., and Waszak, F. (2009). Dissociating what and when of intentional actions. *Front. Hum. Neurosci.* 3, 1–10. doi: 10.3389/neuro.09.003.2009.

Lau, H., Rogers, R. D., and Passingham, R. E. (2006). Dissociating response selection and conflict in the medial frontal surface. *Neuroimage* 29, 446–451.

Lau, H. C., Rogers, R. D., Ramnani, N., and Passingham, R. E. (2004). Willed action and attention to the selection of action. *Neuroimage* 21, 1407–1415.

Mueller, V. A., Brass, M., Waszak, F., and Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *Neuroimage* 37, 1354–1361.

Nachev, P., Rees, G., Parton, A., Kennard, C., and Husain, M. (2005). Volition and conflict in human medial frontal cortex. *Curr. Biol.* 15, 122–128.

Rushworth, M. F. (2008). Intention, choice, and the medial frontal cortex. *Ann. N. Y. Acad. Sci.* 1124, 181–207.

Forget before you remember: dynamic mechanism of memory decay and retrieval

Tatiana Engel* and David Andrieux**

Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, CT, USA
*Correspondence: *tatiana.engel@yale.edu; **david.andrieux@yale.edu

A commentary on

Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model
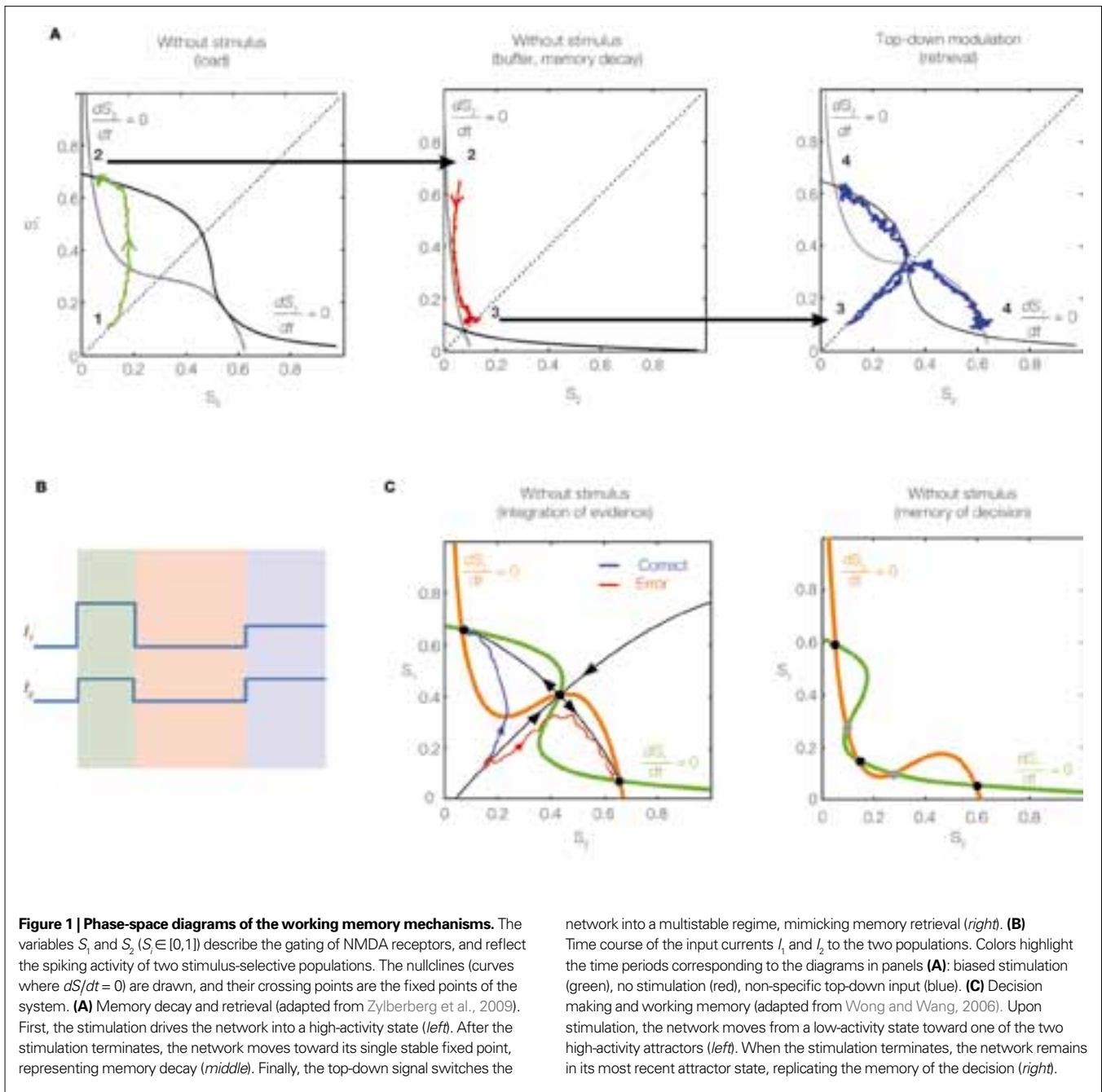
by Ariel Zylberberg, Stanislas Dehaene, Gabriel B. Mindlin and Mariano Sigman

Sensory systems are confronted with a continuous stream of inputs, but only a small fraction of these sensory stimuli reaches our awareness, is consciously perceived and can be remembered. Perception is never driven solely by the bottom-up stimulation, but crucially depends on the top-down modulations. Top-down signals convey behavioral context, such as attention, expectation and perceptual task, and are reflected in the context-specific response modulation in single neurons (Miller and Cohen, 2001; Corbetta and Shulman, 2002). Top-down interactions can be of many different kinds: augmenting or multiplying responses, sharpening tuning curves, controlling contextual influences, or acting as a modulator of plasticity (Desimone and Duncan, 1995; Maunsell and Treue, 2006). Although a lot of empirical knowledge has been accumulated on how top-down interactions modulate neural responses, only a few theoretical attempts have been made so far to explain the underlying biophysical mechanisms and to bridge the gap between the behavioral and single-cell data (Buia and Tiesinga, 2006; Deco and Rolls, 2006; Ardid et al., 2007).

The recent study by Zylberberg et al. (2009) published in Frontiers in Computational Neuroscience aims to uncover these biophysical mechanisms in a particular setting of top-down memory retrieval. The authors try to answer several general and long-standing questions: How do the bottom-up and top-down signals interact to produce a perception? What are the neural mechanisms of effortless (iconic) vs voluntary (working) memory?

Figure 1 | Phase-space diagrams of the working memory mechanisms. The variables S_1 and S_2 ($S_i \in [0,1]$) describe the gating of NMDA receptors, and reflect the spiking activity of two stimulus-selective populations. The nullclines (curves where $dS/dt = 0$) are drawn, and their crossing points are the fixed points of the system. **(A)** Memory decay and retrieval (adapted from Zylberberg et al., 2009). First, the stimulation drives the network into a high-activity state (*left*). After the stimulation terminates, the network moves toward its single stable fixed point, representing memory decay (*middle*). Finally, the top-down signal switches the network into a multistable regime, mimicking memory retrieval (*right*). **(B)** Time course of the input currents I_1 and I_2 to the two populations. Colors highlight the time periods corresponding to the diagrams in panels **(A)**: biased stimulation (green), no stimulation (red), non-specific top-down input (blue). **(C)** Decision making and working memory (adapted from Wong and Wang, 2006). Upon stimulation, the network moves from a low-activity state toward one of the two high-activity attractors (*left*). When the stimulation terminates, the network remains in its most recent attractor state, replicating the memory of the decision (*right*).

How do multiple sensory stimuli compete for representation in the working memory? These questions have been addressed in behavioral experiments, which established that the information is temporally stored in a sensory buffer and decays exponentially (Sperling, 1960; Jolicoeur, 1999; Graziano and Sigman, 2008). In parallel, neurophysiological studies revealed a two-wave activation pattern of single neurons in visual cortical areas: The first wave of activity is largely driven by stimulus prop-

erties, and the second wave is determined by behavioral relevance and contextual aspects (Lamme and Roelfsema, 2000). The work by Zylberberg et al. (2009) attempts to reconcile these experimental observations within a computational circuit model, and to provide a dynamical mechanism for exponential sensory decay and memory retrieval.

The authors propose an interesting two-stage mechanism (**Figures 1A,B**), whereby sensory stimulation biases the

initial state of the network (load), before a non-specific top-down modulation turns the network into the multistable attractor regime (retrieval). During the time-gap (buffer) between the load and retrieval, the network passively moves toward a single stable fixed point, representing a memory decay, i.e., forgetting of the stimulation bias. Joint influences from the stimulation bias, exponential memory decay and noise determine the final state of the network during the retrieval phase, and so the success of memory retrieval.

This mechanism of broadcasting sensory stimuli to the working memory is implemented using a circuit model previously developed for decision making and working memory (Wang, 2002; Wong and Wang, 2006), but in a different operational mode (different set of parameters). The power of this approach is that diverse complex behaviors are captured within a single network model, suggesting that a common circuitry may underlie different cognitive functions. The model originally introduced in References (Wang, 2002; Wong and Wang, 2006) possesses multiple attractor states, and stimulation induces a transition from one attractor state to another, mimicking decision formation and its memory (**Figure 1C**). In contrast, the model of Zylberberg et al. (2009) has a single (low-activity) stable state in the absence of external currents, which leads to the exponential memory decay during the buffering. The top-down modulation then switches the network to a regime with multiple (high-activity) stable states and allows memory retrieval (**Figure 1A**). The idea of switching between single- and multistable regimes by the top-down modulation is interesting in the light of the ongoing debate about the nature of cortical circuits underpinning working memory. While some models propose that persistent activity is subserved by a multistable attractor network (Wang, 2002), others suggest that it reflects slow transients in a network with a single stable state (Goldman, 2009). Dynamical switching between single- and multistable regimes by the top-down modulation may bridge these two alternative hypotheses. Moreover, recovery to the single stable state – when the top-down modulation is turned off – resolves the apparent problem of "switching off" the persistent activity in the multistable attractor network. This important issue is, however, not fully resolved in the paper, since the gating top-down signals are turned on and off "by hand", and no neural mechanism is provided for the dynamics of top-down modulations.

In the model of Zylberberg et al. (2009), both buffering and memory retrieval occur within the sensory areas responsive to the stimulus, without the need for any additional specific "buffer" area. The characteristic time of the exponential memory decay appears insensitive to the stimulus amplitude, but can be modulated by the amount of top-down projections. In the present model, this decay time crucially depends on the slow (~100 ms) N-methyl-D-aspartic acid (NMDA) receptor time course, as well as on the amount of recurrent excitation in the network. Accordingly, the model predicts that the decay time should decrease with the attention deployed during the task, and that blocking NMDA receptors should impair memory retrieval. Pursuing this logic, and under the hypothesis that larger receptive fields arise from stronger local recurrent excitation, the model predicts that the decay time should increase as one proceeds along the visual hierarchy. Importantly, these predictions can be directly tested in experiments.

The model by Zylberberg et al. (2009) not only makes testable experimental predictions, but also opens new avenues for the theoretical investigation of top-down interactions. First of all, as conscious perception depends on both sensory stimulation and top-down expectations (Carpenter and Grossberg, 2003; Gilbert and Sigman, 2007), it would be interesting to explore how specific top-down expectations (as opposed to a non-specific modulatory input) bias the competition of sensory stimuli and their representation in the working memory. In the same vein, it remains to be explored how top-down modulation would interact with the synaptic memory mechanism, whereby the recurrent excitation at the basis of working memory is enhanced by short-term plasticity (Hempel et al., 2000; Mongillo et al., 2008). Finally, behavioral data reveal mechanisms acting in parallel with, but going beyond, the passive decay of information (Jolicoeur, 1999); mechanisms that are not accounted for in the present model. Building on the ideas of Zylberberg et al. (2009), it is important to develop a complete model circuitry that would provide a basis for a more direct comparison with behavioral and neurophysiological data, yielding a critical test of the proposed mechanism. The complete model would also help to identify the relative contributions of local recurrent connections versus feedback projections to activity patterns in visual cortical areas, and in this way greatly contribute to our understanding of the dynamical substrates of cortical computations.

REFERENCES

Ardid, S., Wang, X.-J., and Compte, A. (2007). An integrated microcircuit model of attentional processing in the neocortex. *J. Neurosci.* 27, 8486–8495.

Buia, C. I., and Tiesinga, P. H. E. (2006). Attentional modulation of firing rate and synchrony in a model cortical network. *J. Comput. Neurosci.* 20, 247–264.

Carpenter, G. A., and Grossberg, S. (2003). Adaptive resonance theory. In The Handbook of Brain Theory and Neural Networks, Second Edition, Michael A. Arbib, ed. (Cambridge, MA, MIT Press), pp. 87–90.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Deco, G., and Rolls, E. (2006). Attention, short-term memory, and action selection: a unifying theory. *Prog. Neurobiol.* 76, 236–256.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696.

Goldman, M. S. (2009). Memory without feedback in a neural network. *Neuron* 61, 621–634.

Graziano, M., and Sigman, M. (2008). The dynamics of sensory buffers: geometric, spatial, and experience-dependent shaping of iconic memory. *J. Vis.* 8, 1–13.

Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J. Neurophysiol.* 83, 3031–3041.

Jolicoeur, P. (1999). Concurrent response-selection demands modulate the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1097–1113.

Lamme, V. A. F., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579.

Maunsell, J. H. R., and Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322.

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546.

Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–30.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36, 955–968.

Wong, K.-F., and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* 26, 1314–1328.

Zylberberg, A., Dehaene, S., Mindlin, G. B., and Sigman M. (2009). Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model. *Front. Comput. Neurosci.* 3, 1–16. doi:10.3389/neuro.10.004.2009