2011

Global Workspace Theory, Shanahan, and LIDA

S. Franklin

International Journal of Machine Consciousness
Vol. 3, No. 2 (2011) 1−11

DOI: 10.1142/S1793843011000728

GLOBAL WORKSPACE THEORY, SHANAHAN, AND LIDA

STAN FRANKLIN

Department of Computer Science,
Institute for Intelligent Systems,
The University of Memphis,
Memphis TN 38152, USA
franklin@menphis.edu

Following a brief review of Shanahan's so many, and so important, contributions to global workspace theory, as presented in his *Embodiment and the Inner Life*, we attempt to interpret, and flesh out, Shanahan's top-down account of GWT from a bottom-up perspective guided by our LIDA model of consciousness and cognition.

1. Wittgenstein, Metaphysics and the Post-Reflective Inner View

Shanahan's *Embodiment and the Inner Life* [2010], begins with a chapter espousing that we manage the age-old metaphysical problems surrounding consciousness [Chalmers, 1996] by following the later dictates of Wittgenstein and taking what Shanahan calls the post-reflective inner view. He argues that "To put it bluntly, we must learn to do without the habit of metaphysical thinking." Put another way, we may embark on the journey of a scientific understanding of consciousness without addressing many of the traditional, seemingly intractable, concerns of philosophy, e.g., the mind−body problem. Being convinced by his arguments, I shall, post-reflectively, say no more on the subject.

2. Embodied Cognition *et al.*

A chapter on "Cognition and Embodiment" presents Shanahan's views on what, in cognitive science, is called embodied cognition [Barsalou, 1999; de Vega *et al.*, 2008; Glenberg and Robertson, 2000]. The remainder of the book is an utterly fascinating, if abbreviated, account of Shanahan's own journey into the scientific exploration of consciousness. It seems that, like so many of us these days, Shanahan feels this journey must be taken in the context of Baars' global workspace theory (GWT) [2002; Franklin *et al.*, 2005]. Subsequent chapters detail

Shanahan's views of how the brain's network anatomy and dynamics support the needs of GWT and, so, allow consciousness to play its vital role in deliberation and creative problem solving.

3. Global Workspace Theory

Global workspace theory [Baars, 1988] was originally conceived as a neuropsychological model of conscious and unconscious processes, but has been broadened into a higher-level theory of human cognitive processing, supported by empirical evidence [Baars, 2002]. GWT views the nervous system as a distributed parallel system with many different specialized processes. Coalitions of these processes enable an agent [Franklin and Graesser, 1997] to make sense of the sensory data coming from the current environmental situation. Other coalitions, incorporating the results of the processing of sensory data, compete for attention in what Baars calls a global workspace. The contents of the winning coalition are broadcast to all other processes. The contents of this broadcast are proposed to be phenomenally conscious. This conscious broadcast serves to recruit other, unconscious, processes to be used to select an action in response to the current situation. GWT is therefore a theory of how consciousness functions within cognition.

GWT postulates that the brain's multitude of relatively small, special purpose processes are almost always unconscious [Franklin and Baars, 2010]. Communication between them is rare and over a narrow bandwidth. Coalitions of processes can find their way into the global workspace (and perhaps into consciousness). This limited capacity workspace serves to broadcast the contents of the coalition to all the unconscious processors, in order to recruit other processors to join in handling the current situation, or in solving the current problem. Thus, consciousness in this theory enables the agent to deal with novel or problematic situations that cannot be dealt with efficiently, or at all, by habituated unconscious processes. In particular, consciousness provides access to appropriately useful resources, thereby solving the *relevance problem* [Franklin, 2003], constituting a major function of consciousness in cognition.

A second major function of consciousness in cognition is the enabling of learning, the encoding of knowledge about the past for use in the present. GWT supports the conscious learning hypothesis: Significant learning takes place via the interaction of consciousness with the various memory systems [Franklin *et al.*, 2005]. That is, all memory systems rely on conscious cognition for their updating, either in the course of a single cycle or over multiple cycles (see below for a discussion of cognitive cycles). Shanahan has little to say about consciousness and learning, which does not appear in the subject index.

Data flow according to GWT can be visualized as having an hourglass shape with sensory data coming in the top and flowing through the upper chamber. The bottleneck at the center represents the limited capacity global workspace acting as a relevance filter before broadcasting conscious contents throughout the brain, represented by the bottom chamber. Are brains structured so as to allow such data flow?

December 28, 2011 3:06:05pm WSPC/258-IJMC 00072 ISSN: 1793-8430

4. Shanahan's Contributions to Global Workspace Theory

Shanahan's contributions to GWT are, in my opinion, both numerous and major. They are too numerous and too diverse for me to attempt to recount them all here except as a high-level overview. Essentially, he asks what structures are required to implement GWT in brains, and do those structures exist in brains? He also talks of the dynamics of neural computations and of the role of consciousness in problem solving via making connections. We will briefly describe his views on each of these issues.

4.1. *Neural evidence for the possibility of broadcasts*

Brains are reasonably thought as networks composed of nodes (neurons) and links (axons/synapses). A brain *module* would be a cluster of neurons densely connected within but sparsely connected to the outside. What sort of network structure would allow a GWT architecture to be implemented within a brain? Shanahan argues that a "*small world*" *network*, one that is densely (sparsely) connected locally (globally) and with a short path between any two nodes, is needed. He further asserts that this small world property is needed at multiple scales of organization of modules and submodules, say from cell assembly modules (small processes in GWT) to much larger cognitive modules. This implies that the network should be "hierarchically modular". Finally, these networks should be liberally provided with *connector hubs*, through which many paths pass from one module to another. These connector hubs, together with their interconnections, constitute the *connective core* of the network. Shanahan argues that a hierarchically modular small world network structure provided with a connective core constitutes the neural network communicative infrastructure needed to implement GWT by enabling small processes to influence the global system by being broadcasted.

Do brains have the kind of network infrastructure so described as needed to support GWT? At this point Shanahan's arguments become empirical, resting heavily on an extensive neuroscience literature. After describing both the means and the findings of research into the brain's connectivity he concludes that they are indeed looking at a hierarchically modular, small world network. He then shows that "The brain embeds a network with a pronounced connective core that is capable of globally disseminating the influence of a process or coalition of processes," in such a way that "only one coalition of processes at a time can take over the connective core, to the exclusion of its rivals". Voila, the network infrastructure we need for enabling GWT.

4.2. *Dynamics of neural computations*

With a GWT compatible network infrastructure in hand, what can be said about the patterns of activity it must carry, and how do they change? That is, what are the neurodynamics like? Shanahan bases his description of these dynamics on

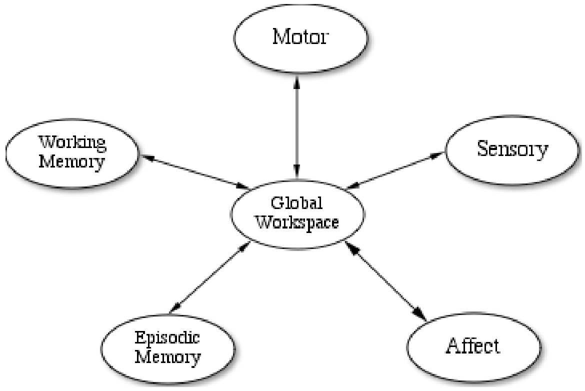

Fig. 1. Brain "super modules" à la Shanahan.

anatomically distributed coalitions of brain processes (cell assemblies), drawn from "super modules" as illustrated in Fig. 1 above. The global workspace constitutes the connective core of the brain network, with only one coalition taking over at a time resulting in a global broadcast.

Shanahan hypothesizes the locations of these super modules in brains to be roughly as follows:

> "...*sensory module* → *occipito-temporal regions*; *motor module* → *fronto-parietal regions*; *affect module* → *sub-cortical structures including the amygdala and basal ganglia*, *plus orbito-frontal cortex*; *episodic memory* → *medial regions including entorhinal cortex and the hippocampus*; *working memory* → *prefrontal cortex*."

Following the lead of Skarda and Freeman [1987], Edelman and Tononi [2000] and Kelso [1995], Shanahan describes the behavior of these coalitions in the language of nonlinear dynamics with the trajectories of their activations "wandering among the actractors" before settling into one. Call a system *segregated* to the extent that its parts choose their activity independently of one another and *integrated* to the extent that each part's activity is influenced by that of the whole. A system "is dynamically complex when this influence is not too great", when there is a balance between segregation and integration. When applied to brain coalitions as found in the modules of Fig. 1, such a dynamically complex system should

> "...*facilitate not merely swift and appropriate behavior selection*, *but also the immediate assembly of novel behaviors tailored to meet the situation at hand*, *as well as ensuring that off-line exploration of the space of affordances was open-ended*."

That is just what one would expect a GWT-based system to achieve.

But how do these various parts communicate with each other, say within coalitions, or during the global broadcast? Shanahan hypothesizes that

> "*...the selective passage of information between two groups of* [Buzsaki, 2006] *neurons is facilitated if* (1) *there is synchronous oscillation within each group, and* (2) *this synchronous activity is coherent* (*phase-locked*) *between the two groups*."

This hypothesis of communication through synchronous, phase-locked oscillators has become widely popular among cognitive neuroscientists [Buzsaki, 2006; Canolty *et al.*, 2006; Freeman, 1999].

4.3. *Importance of consciousness to making connections*

Shanahan finishes the book with a chapter applying his earlier observations and conclusion about the needed structure of communication networks for GWT, and neuroscience evidence its existence in human brains, to their consequences for our "inner life", thereby making full cycle with the books title. He says:

> "*The chapter argues that an internal sensorimotor loop operating through the global workspace can account for foresight and planning*, *as well as episodic memory and other facets of the imagination*."

All this is based on what he calls the "*simulation hypothesis*" which asserts that "...conscious thought can be accounted for in terms of simulated interactions with the world." This hypothesis is made plausible by the brain's sensory and motor processes being able to operate offline.

Pulling in the work of others, Shanahan goes on to speak of the simulation hypothesis, utilizing a GWT architecture with its inner loop that can connect essentially any process to any other, enabling "open-ended affordance", "conceptual blending", and "cognitive fluidity", making possible genuinely creative solutions to problems where any process may turn out to be relevant and important. Thus, as is implicit in GWT [Franklin, 2003], this architecture effectively solves the relevance problem.

> "*Negotiations between potential partners*, *the recruitment of new coalition members*, *the suppression of rival coalitions — all this goes on in parallel, until one coalition emerges from the melee as dominant.... The deciding factor in this battle is relevance*."

Shanahan talks of our inner life as being constructive, including episodic memories as well as planning and imagination. "Episodic memory ... is a matter of the *reassembly* of the elements of a past episode..." Shanahan takes on human language as "just another form of interaction with the world". Finally, he addresses our ubiquitous inner speech, our thinking in words, saying that it "arises through simulated interaction with the social environment".

The past several sections of this review describe Shanahan's significant contributions to our understanding of the relationship between GWT and the brain. He has detailed the network infrastructure required to implement GWT, argued persuasively for its presence in brains, described the communication medium of synchronous, phase-locked oscillators, and made a strong case for GTW in brains actually solving the relevance problem. Much of this was accomplished taking what this reviewer sees as a top-down approach to cognition. The next sections describe a more bottom-up approach, and on attempt to interpret Shanahan's work from that perspective guided by the LIDA model.

5. The LIDA Model of Consciousness and Cognition

Faithfully based on GWT, the LIDA model incorporates ideas from many other cognitive theories. Intended as a band, comprehensive, integrated, conceptual and computational cognitive model, it has been described in detail in numerous previous papers. [See for example, Baars and Franklin, 2003, 2007; Franklin *et al.*, 2005; McCall *et al.*, 2010; Wallach *et al.*, 2010]. Here, we will recap only enough of the LIDA model to allow the reader to appreciate what it can add to Shanahan's description of a dynamic GWT.

Based on the idea of a cognitive cycle roughly made of three parts: perception, sense-making and action, the LIDA model takes a bottom-up approach to cognition. Every autonomous agent [Franklin and Graesser, 1997], be it human, animal, or artificial, must frequently sample (sense) its environment and select an appropriate response (action). More sophisticated agents, such as humans, process (make sense of) the input from such sampling in order to facilitate their decision-making. Neuroscientists call this three-part process the action-perception cycle [Freeman, 2002]. The agent's "life" can be viewed as consisting of a continual cascade of these cognitive cycles. A cognitive cycle can be thought of as a moment of cognition — a cognitive "moment". Higher-level cognitive processes are composed of many of these cognitive cycles, each a cognitive "atom".

Just as atoms are composed of protons, neutrons and electrons, and some of these are composed of quarks, bosons, muons, etc., these cognitive "atoms" have a rich inner structure. We will next see what the LIDA model hypothesizes as the rich inner structure of the cognitive cycle.

6. LIDA's Cognitive Cycle

During each cognitive cycle, the LIDA agent first makes sense of its current situation as best as it can by updating its representation of its world, both external and internal. By a competitive process, as specified by global workspace theory, it then decides what portion of the represented situation is most in need of attention. Broadcasting this portion, the current contents of consciousness, enables the agent to finally choose an appropriate action and execute it. Thus, the LIDA cognitive cycle can be subdivided into three phases: the understanding phase, the consciousness phase, and the action selection phase. Figure 2 is intended to help the reader follow the upcoming descriptions of the three phases.

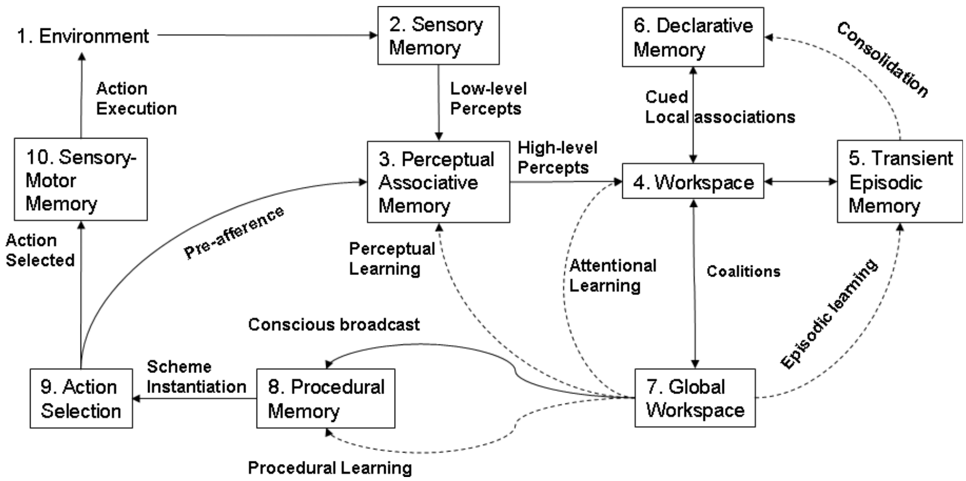

Fig. 2. The LIDA cognitive cycle.

Beginning the *understanding phase*, incoming stimuli activate low-level feature detectors in Sensory Memory. The output is sent to Perceptual Associative Memory where higher-level feature detectors feed in to more abstract entities such as objects, categories, actions, feelings, events, etc. The resulting percept is sent to the Workspace where it cues both Transient Episodic Memory and Declarative Memory producing local associations. These local associations are combined with the percept to generate a current situational model, the agent's understanding of what is going on right now.

Attention Codelets begin the *consciousness phase* by forming coalitions of selected portions of the current situational model and moving them to the global workspace. A competition in the global workspace selects the most salient coalition whose contents become the content of consciousness that are broadcast globally.

The *action selection phase* of LIDA's cognitive cycle is also a learning phase in which several learning processes operate in parallel. New entities and associations, and the reinforcement of old ones, occur as the conscious broadcast reaches Perceptual Associative Memory. Events from the conscious broadcast are encoded as new memories in Transient Episodic Memory. Possible action schemes, together with their contexts and expected results, are learned into Procedural Memory from the conscious broadcast. Older schemes are reinforced. In parallel with all this learning and using the conscious contents, possible action schemes are recruited from Procedural Memory. A copy of each is instantiated with its variables bound and sent to Action Selection, where it competes to be the behavior selected for this cognitive cycle. The selected behavior triggers Sensory-Motor Memory to produce a suitable algorithm for the execution of the behavior. Its execution completes the cognitive cycle.

The LIDA model hypothesizes that all human cognitive processing is via a continuing iteration of such cognitive cycles. These cycles occur asynchronously, with each taking roughly 300 milliseconds. The cycles cascade, that is, several cycles may have different processes running simultaneously in parallel. This cascading must, however respect the way consciousness processes information serially in order to maintain the stable, coherent image of the world with which consciousness endows us [Franklin, 2005; Merker, 2005]. It must also respect the seriality of action selection. This cascading, together with the asynchrony, allows a rate of cycling in humans of five to ten cycles per second. A cognitive "moment" is thus quite short! For instance, for cascading cycles of 300 milliseconds each offset by 100 ms, cycles occur at 10 Hz. There is considerable empirical evidence from neuroscience suggestive of and consistent with such cognitive cycling in humans [Doesburg *et al.*, 2009; Massimini *et al.*, 2005; Sigman and Dehaene, 2006; Uchida *et al.*, 2006; Willis and Todorov, 2006]. The evidence demonstrates the existence of neurophysiological rhythms that one can interpret as the manifestation of cognitive cycles.

6.1. *Shanahan's model from a LIDA perspective*

In the book, Shanahan for the most part takes a top-down view of human cognition, that is, taking a perspective motivated by high-level cognitive processes. As we have just seen, a LIDA perspective is more bottom-up, employing cognitive cycles as cognitive "atoms" or moments out of which higher-level cognitive processes are constructed. Each high-level, that is, multi-cyclic, cognitive process is comprised of at least one behavior stream from LIDA's Procedural Memory. Remember that some of LIDA's actions are executed internally, for example, writing an intention into the workspace. The behaviors in a defining behavior stream, if selected by Action Selection over multiple cycles, constitute a sequence of actions implementing the higher-level process.

The information flow, as indicated by the two-headed arrows in Shanahan's Fig. 1, goes both ways. Taking a top-down perspective, as he does, this is as it should be, with the arrows out of the global workspace carrying the conscious broadcasts of some of the multiple cycles. If we restrict that diagram to the information flow of a single cognitive cycle, the diagram in Fig. 3 results. Let us compare these two diagrams.

The first noticeable difference is that Fig. 3 carries no arrow from Global Workspace into Affect. Shanahan includes Affect as a major module, both because of its importance to the various cognitive processes, and because brain modules such as the amygdale are strongly correlated with affect. Enthusiastically concurring with Shanahan as to its importance, the LIDA model incorporates aspects of affect into each and every module displayed in Fig. 2, with the possible exception of Sensory Memory. Affect's initial impact in the LIDA model occurs as part of Perceptual Associative Memory (PAM), which would be included in Fig. 1's Sensory major module. The double arrow between Affect and Sensory in Fig. 3 denotes activation passing back and forth between feeling and emotion nodes in PAM and various other nodes such as, objects, agents, events, etc.

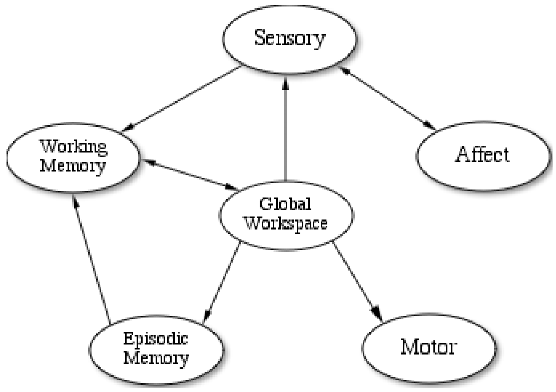

Fig. 3. LIDA cognitive cycle version of Shanahan's diagram.

Next note that in Fig. 3, the only input into the global workspace comes from Working Memory. In Fig. 1, Shanahan uses the term "Working Memory" in the standard sense of Baddeley, which includes conscious contents over multiple cycles (1992). In Fig. 3, please interpret Working Memory to mean LIDA's preconscious workspace, from whose Current Situational Model attention codelets choose structures and from coalitions that will compete for control of the global workspace.

Figure 2 shows how LIDA's workspace (the Working Memory major module in Fig. 3) receives input from Sensory Memory through PAM (the Sensory major module in Fig. 3) as well as from Transient Episodic Memory and Declarative Memory (the Episodic Memory major module in Fig. 3). Fig. 2 also illustrates how the output from the global workspace, aside from the various modes of learning, goes to Action Selection via Procedural Memory (the Motor major module in Fig. 3).

The point of all this discussion is that the bottom-up perspective provided by the inner structure of the LIDA cognitive cycle serves to flesh out Shanahan's version of cognition à la GWT.

This fleshing out is illustrated one final critically important distinction concerning conscious mediation. The learning theorists talk of implicit learning in which a skill is learned, for example to identify acceptable strings of letters, without being conscious of the rule upon which the skill is based. Such implicit learning, sometimes erroneously termed unconscious learning, is always consciously mediated, that is completely dependent on the learner having been conscious of each instance of presentation of a letter string. The conscious mediation we are concerned with here involves action selection.

The LIDA model distinguishes four methods of action selection, volitional, consciously mediated, automatized, and alarm [Wallach *et al.*, 2010], the first two of which concern us here. Following GWT, and fleshing it out, *volitional action selection* in LIDA is based on James' ideomotor theory [Baars, 1988]. The action is selected consciously over multiple cognitive cycles with various possibilities "popping into mind", that is, becoming conscious. *Consciously mediated action selection* is the name given in LIDA to the action selection process that takes place during each single cognitive cycle. This method is implicit in the sense that selection mechanism, although crucially based on the conscious broadcast, is entirely unconscious, in fact, never conscious [Franklin and Baars, 2010]. Note that consciously mediated actions are always selected during a single cycle. Note also that Shanahan in a number of places refers to consciously mediated action selection when we would call the multi-cyclic selection he is speaking of volitional. He also sometimes refers to what we would call consciously mediated action selection as automatized. These action selections have certainly been well learned, but we reserve "automatized" to indicate that one action calls the next with no conscious mediation, as steps sometimes do when walking down an empty hallway.

7. Conclusion

Shanahan's Embodiment and the Inner *Life* makes significant and important contributions to our understanding of how consciousness, as formulated by GWT, can operate in brains. However, the most appreciated contribution by this reviewer were his arguments for the post-reflective view, which allow me to ignore some of the old philosophical problems of consciousness without a feeling of guilt.

References

Baars, B. J. [1988] *A Cognitive Theory of Consciousness* (Cambridge University Press, Cambridge).

Baars, B. J. [2002] "The conscious access hypothesis: Origins and recent evidence," *Trends in Cognitive Science* **6**, 47–52.

Baars, B. J. and Franklin, S. [2003] "How conscious experience and working memory interact," *Trends in Cognitive Science* **7**, 166–172.

Baars, B. J. and Franklin, S. [2007] "An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA," *Neural Networks* **20**, 955–961.

Barsalou, L. W. [1999] "Perceptual symbol systems," *Behavioral and Brain Sciences* **22**, 577–609.

Buzsaki, G. [2006] *Rhythms of the Brain* (Oxford University Press, Oxford).

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E. *et al.* [2006] "High gamma power is phase-locked to theta oscillations in human neocortex," *Science* **313**, 1626–1628.

Chalmers, D. J. [1996] *The Conscious Mind* (Oxford University Press, Oxford).

de Vega, M., Glenberg, A. and Graesser, A. (eds.) [2008] *Symbols and Embodiment: Debates on Meaning and Cognition* (Oxford University Press, Oxford).

Doesburg, S., Green, J., McDonald, J. and Ward, L. [2009] "Rhythms of consciousness: Binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception," *PLoS ONE* **4**(7), e6142.

Edelman, G. M. and Tononi, G. [2000] *A Universe of Consciousness* (Basic Books, New York).

Franklin, S. [2003] "An autonomous software agent for navy personnel work: A case study. in *Human Interaction with Autonomous Systems in Complex Environments: Papers from 2003*

AAAI Spring Symposium, (eds.) D. Kortenkamp and M. Freed, (AAAI) Palo Alto, pp. 60−65.

Franklin, S. [2005] "Evolutionary pressures and a stable world for animals and robots: A commentary on merker," *Consciousness and Cognition* **14**, 115−118.

Franklin, S. and Baars, B. [2010] "Two varieties of unconscious processes," in *New Horizons in the Neuroscience of Consciousness*, (eds.) E. Perry, D. Collerton, H. Ashton and F. LeBeau, (Amsterdam: John Benjamin), pp. 91−102.

Franklin, S., Baars, B. J., Ramamurthy, U. and Ventura, M. [2005] "The role of consciousness in memory," *Brains, Minds and Media* **1**, 1−38.

Franklin, S. and Graesser, A. C. [1997] "Is it an agent, or just a program? A taxonomy for autonomous agents," *Intelligent Agents III*, (Springer Verlag, Berlin), pp. 21−35.

Freeman, W. J. [1999] "Consciousness, intentionality and causality". in *Reclaiming Cognition* (eds.) R. Núñez and W. J. O. R. o. E. e. Freeman, (Imprint Academic, Exeter), pp. 143−172.

Freeman, W. J. [2002] "The limbic action-perception cycle controlling goal-directed animal behavior," *Neural Networks* **3**, 2249−2254.

Glenberg, A. M. and Robertson, D. A. [2000] "Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning," *Journal of Memory and Language* **43**, 379−401.

Kelso, J. A. S. [1995] *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MA: MIT Press, Cambridge).

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H. and Tononi, G. [2005] "Breakdown of cortical effective connectivity during sleep," *Science* **309**, 2228−2232.

McCall, R., Franklin, S. and Friedlander, D. [2010] *Grounded Event-Based and Modal Representations for Objects, Relations, Beliefs, Etc.* Paper presented at the FLAIRS-23, Daytona Beach, FL.

Merker, B. [2005] "The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution," *Consciousness and Cognition* **14**, 89−114.

Shanahan, M. [2010] *Embodiment and the Inner Life* (Oxford University Press, Oxford).

Sigman, M. and Dehaene, S. [2006] "Dynamics of the central bottleneck: Dual-task and task uncertainty," *PLoS Biol.*, **4**.

Skarda, C. and Freeman, W. J. [1987] "How brains make chaos in order to make sense of the world," *Behavioral and Brain Sciences* **10**, 161−195.

Uchida, N., Kepecs, A. and Mainen, Z. F. [2006] "Seeing at a glance, smelling in a whiff: Rapid forms of perceptual decision making," *Nature Reviews Neuroscience* **7**, 485−491.

Wallach, W., Franklin, S. and Allen, C. [2010] "A conceptual and computational model of moral decision making in human and artificial agents". in *Topics in Cognitive Science, Special Issue on Cognitive Based Theories of Moral Decision Making* (eds.), W. Wallach and S. Franklin, (Cognitive Science Society), pp. 454−485.

Willis, J. and Todorov, A. [2006] "First impressions: Making up your mind after a 100-Ms exposure to a face," *Psychological Science* **17**, 592−599.