2009

Evolutionary pressures for perceptual stability and self as guides to machine consciousness

S. Franklin

S. Mello

B. Baars

U. Ramamurthy

EVOLUTIONARY PRESSURES FOR PERCEPTUAL STABILITY AND SELF AS GUIDES TO MACHINE CONSCIOUSNESS

By Stan Franklin, Sidney D'Mello, Bernard J. Baars and Uma Ramamurthy

The currently leading cognitive theory of consciousness, Global Workspace Theory (Baars 1988, 2003), postulates that the primary functions of consciousness include a global broadcast serving to recruit internal resources with which to deal with the current situation and to modulate several types of learning. In addition, conscious experiences present current conditions and problems to a "self" system, an executive interpreter that is identifiable with brain structures like the frontal lobes and precuneus (Baars, 1988).

Be it human, animal or artificial, an autonomous agent (Franklin and Graesser 1997) is said to be *functionally consciousness* if its control structure (mind) implements Global Workspace Theory and the LIDA Cognitive Cycle, which includes unconscious memory and control functions needed to integrate the conscious component of the system. We would therefore consider humans, many animals (Seth, Baars, and Edelman 2005) and even some virtual or robotic agents (Franklin 2003, Shanahan 2006) to be functionally consciousness. Such entities may approach phenomenal consciousness, as found in human and other biological brains, as additional brain-like features are added.

Here we argue that adding mechanisms to produce a stable, coherent perceptual field (Merker 2005) in a LIDA controlled mobile robot might provide a small but significant step toward phenomenal consciousness in machines (Franklin, 2005). We also propose that implementing several of the various notions of self in such a LIDA controlled robot may well prove another step toward phenomenal consciousness in machines.

1. Machine Consciousness

In the last decade there have been increasing efforts to address the question of machine consciousness. A number of computational models have been proposed and implemented, international conferences have been held, and this peer-reviewed journal has been founded.

A 2001 workshop entitled "*Can a machine be conscious*?" was the impetus for a community of researchers to embark on the serious, scientific study of the possibility of machine consciousness. This was followed by subsequent workshops in Torino, Italy (2003), Lesvos, Greece (2005) and Espoo, Finland (2008). At these meetings various projects aimed at eventually achieving machine consciousness were reported on. These include Igor Aleksander's MAGNUS (2000), Rodney Cotterill's CyberChild (2003), Owen Holland's CRONOS (2007), Pentti Haikonen's Cognitive Machine (2007), Stan Franklin's LIDA (Franklin and Patterson 2006) and others.

At essentially the same time a related effort, dubbed Artificial General Intelligence (AGI), arose. Fifty years ago artificial intelligence began with dreams of creating machines with human-level intelligence. The difficulty of the endeavor quickly led to a focus on "narrow AI" -- software programs that perform intelligently in specific, narrow domains like chess, medical diagnosis, etc. These programs lack the human ability to generalize their knowledge across different domains, and typically can neither reflect on themselves nor their own behaviors.

AGI represents a move back toward the original AI endeavor of creating human-like intelligence. Like machine consciousness, the AGI movement has been supported by workshops (e.g. the AGIRI Workshop, Washington, DC, 2006), conferences (AGI-08, Memphis, 2008) and a new journal (*Journal of Artificial General Intelligence*). Though not directly aimed at machine consciousness, producing human level intelligence in a machine may well require that it be conscious. Thus AGI is likely to be an able and useful partner in the endeavor to enable machine consciousness.

2. Functional vs. Phenomenal Consciousness

The currently leading empirical theory of consciousness, Global Workspace Theory (Baars 1988, 2003), postulates that the primary functions of consciousness include a global broadcast serving to recruit internal resources with which to deal with the current situation and to modulate several types of learning. In addition, conscious experiences present current situations and problems to a "self" system, an executive interpreter that is identifiable with brain structures like the frontal lobes and precuneus (Baars 1988).

Be it human, animal or artificial, an *agent* (Franklin and Graesser 1997) is said to be *functionally consciousness* if its control structure (mind) implements Global Workspace Theory and the LIDA Cognitive Cycle, which includes unconscious memory and control functions needed to integrate the conscious component of the system. We would consider humans, many animals (Seth, Baars, and Edelman 2005) and even some virtual or robotic agents (Franklin 2003, Shanahan 2006) to be functionally consciousness.

We must carefully distinguish functional consciousness from the biopsychological meaning of "consciousness," which assumes phenomenal experience, the subjective experience of qualia. To keep this distinction clear we will refer to consciousness in this typical usage as "*phenomenal consciousness*." The projects toward machine consciousness mentioned above are all aimed at eventually achieving artificial phenomenal consciousness. Is this even possible? We believe that an embodied, robotic version of LIDA --- which would meet a number of criteria for human consciousness --- will perhaps be the closest entity to artificial phenomenal consciousness to date. Phenomenal consciousness is plausibly believed to exist in biological entities that have a sizable set of known features, for example mammals (Seth et al, 2005; Baars, 1988). All mammals share the core brain anatomy of the thalamus and neocortex; waking and sleeping EEG look dramatically different in this core brain, and many scientists have suggested that waking activity in the thalamocortical core reflects the basic physiological substrate of consciousness. Biologically speaking there are many close homologies between human and other mammalian brains, particularly in the enlarged thalamus and cortex. But if all mammals are subjectively conscious, as these strong homologies suggest, why should consciousness not extend just as much to artificial systems that process information in much the same way?

As functionally conscious computational entities achieve more and more of the computational features of biological consciousness it may become difficult to distinguish robots from animals as far as consciousness is concerned. (Please see http://consc.net/mindpapers/6.1d for a current list of articles on machine consciousness.)

3. The IDA Software Agent

IDA (Intelligent Distribution Agent) is an intelligent software agent (Franklin and Graesser 1997) developed for the US Navy (Franklin et al. 1998). Although IDA was developed for a specific set of human tasks, it reflects broader principles of human cognition. In the initial IDA implementation, its aim was to simulate human "detailers," whose job it is to assign US Navy sailors to suitable jobs. At the end of each sailor's tour of duty, he or she is assigned to a new billet (e.g., job). This complex assignment process is called distribution (hence the name IDA). The Navy employs some 300 trained people, called detailers, full time to effect these new assignments. IDA facilitates this process by completely automating the role of the human detailer (Franklin 2001). Communicating with sailors by email in unstructured English, IDA negotiates with them about new jobs, employing constraint satisfaction, deliberation and volition. IDA eventually assigns each sailor a job constrained by both human and organizational requirements. The IDA software agent is currently up and running, and in testing by the Navy has matched the performance of the Navy's human detailers (personal communication).

IDA is quite a complex software agent (Franklin & Graesser 1999) that models a broad swath of human cognition including "consciousness" in the sense of implementing Global Workspace Theory. IDA exhibits both external and internal voluntary action selection (Franklin 2000), as well as consciously mediated action selection[1] of both the internal and external variety. Implementing an earlier version of the LIDA cognitive cycle (Baars and Franklin 2003, Franklin et al 2005), it uses its "consciousness" module to handle routine problems with novel content. The consciousness module also allows IDA to watch for unexpected events – both dangers and opportunities. All this together makes a strong case, in our view, for functional consciousness as defined above.

But, is IDA phenomenally consciousness? We have argued earlier that there are "no convincing arguments for such a claim" (Franklin 2003) and currently see no reason to change that view. It seems that IDA implements part, but not all, of consciousness. What needs to be added to an IDA based software agent to achieve phenomenal consciousness? We have no definitive answer to this question. However, we do have conjectures as to at least some parts of the answer, as we will go on to describe below.

4. Merker's Evolutionary Pressure for Phenomenal Consciousness

The neurobiologist Bjorn Merker has suggested one plausible selection pressure that may have served to increase the evolutionary fitness of phenomenal consciousness in humans and other conscious animals. He points out that phenomenal consciousness produces a stable, coherent perceptual world for animals by distinguishing real motion in the world from apparent motion produced by the movement of sensory receptors (Merker 2005). One can experience the loss of this stable, coherent sensory world by a simple experiment. Close one eye and press gently with an index finger on the upper eyelid of the open eye. The movement of the eyeball produces an apparent motion of whatever is present in the experimenter's perceptual field. This external intervention therefore defeats the normal compensatory mechanisms that keep our subjective perceptual world stable. But when the constant movements of eyes, the head, and the body are endogenously controlled, no such movement of the world is perceivable. Thus brain mechanisms underlying conscious perception must act to keep the world stable in spite of a vast and complex variety of movements in which we normally engage.

Merker does not claim that phenomenal consciousness is the only process capable of producing such a stable, coherent perceptual world. Nor does he claim that this process of distinguishing and suppressing apparent motion provides the only evolutionary selection pressure. He simply suggests that providing perceptual stability and coherency is one fitness benefit of phenomenal consciousness. But, what has all this to do with consciousness in machines?

5. A Perceptually Stable and Coherent LIDA Controlled Robot

In a commentary on Merker's article, Franklin suggested that producing a robot provided with a stable, coherent perceptual world might be a step toward a phenomenally conscious machine (2005). Let us call a sense organ *spatially sensitive* if movement of the organ produces apparent motion at its surface independent of what is happening in the environment. Any autonomous, mobile robot will likely require spatially sensitive sensory mechanisms (e.g. vision) for moving appropriately in its world. Thus, the problem of distinguishing real motion from self-produced, apparent motion will be ubiquitous among such robots. One solution would be to build in mechanisms to shield the robot's action selection from apparent motion that is self-produced by the movement of its sense organs. Such shielding mechanisms might conceivably be based on any of several different principles. One such principle would have the robot construct its own individual, coherent and stable world, suppressing self-produced apparent motion, as Merker argues that consciousness does for some animals. Such a stable, coherent perceptual world would prevent self-induced apparent motion from interfering with the robot's action selection.

Here we propose a LIDA controlled autonomous mobile robot with such a built-in shielding mechanism producing a coherent, stable, perceptual worldview. LIDA (Learning IDA) is a conceptual, and

---

[1] Consciously mediated action selection uses the current contents of consciousness to aid in the unconscious selection of an appropriate action. For example, in walking through a room, a person uses conscious information as a basis for typically unconscious action selection to avoid bumping into the furniture.

partially computational, cognitive architecture (Franklin and Patterson 2006, Ramamurthy, D'Mello, and Franklin 2006), derived from IDA primarily by adding several modes of learning. The most accessible description of the LIDA architecture can be found on the web at http://ccrg.cs.memphis.edu/tutorial/index.html.

Can such a shielding mechanism be designed? That is an empirical question for robot designers. Our experience with designing IDA and LIDA suggests that essentially any human cognitive process, including deliberation and volitional decision-making (Franklin 2000), can be effectively simulated in a software agent. Why not in a robot?

On superficially attractive way of creating a stable visual, for example, field for a robot would be to interrupt incoming visual sensing during motion of the visual receptors (camera). (This is what happens in humans during saccades of the eye (Castet and Masson 2000). Such motion might result from movement of any or all of the robot's entire body, its head if it's humanoid, or just of its camera. Even such a simple process should work. It can be refined a little by tying the interruption of incoming visual sensory data specifically to any movement of the camera, whether it's a part of some body or head movement or not. These same arguments should apply to any spatially sensitive sensory receptors.

Will all these interruptions lead to a stable, apparently continuous, sensory field? If the analogy with human vision is to be believed, interruptions of sensory input during receptor movement would have to be sufficiently infrequent and of sufficiently short duration. A human viewing a motion picture film will see it as continuous motion if the frame rate is greater than 17 per second (17hz). Theater projection systems typically use 23 or 24hz. According to our LIDA model of cognition the frame rate of conscious contents is roughly 5 to 9 cognitive cycles per second (Franklin et al 2005), but we experience a stable, continuous visual field because at a lower frame rate since each of our frames (cognitive cycles) typically contains motion. These considerations should provide some guidance to developers of a LIDA controlled mobile robot provided with stable sensory fields for its spatially sensitive senses.

6. The Various Human and Animal Self Systems

Philosophers, psychologists and neuroscientists have postulated various form of a "self" in humans and in other animals. All of these selves seem to assume some form of consciousness. Presumably, a phenomenally conscious machine should possess some or all of these selves. Here we propose adding such selves to a LIDA controlled robot as another possible step toward phenomenal consciousness. Figure 1 offers one organizational chart of some of these various selves. In this section we'll provide a brief description of each of them.

Neuroscientist Antonio Damasio has conceived of the *proto-self* as a short-term collection of neural patterns of activity that represent the current state of the organism (1999). His proto-self receives neural and hormonal signals from visceral changes. The proto-self is not conscious, but it constitutes the biological precedent of the "self."

The *minimal,* or *core, self* is attributed to essentially all animals by biologists, philosophers and neuroscientists (Bekoff and Sherman 2004, Damasio 1999, Gallagher 2000, Strawson 1999). It is produced whenever the processing of an object modifies the proto-self. For example, when an object is being visually sensed, the organism's lens and pupil must be adjusted, changing the proto-self. According to Damasio, this change produces a core conscious experience of the object. The core self represents the lowest level that can be regarded as conscious, in a kind of immediate, unreflecting consciousness presumably possessed by most animals, not just by human beings. It is limited to that which is immediately present. This core consciousness is continually regenerated in a series of pulses (frames, LIDA cognitive cycles), which blend together to give rise to a continuous stream of consciousness. The minimal or core self is sometimes partitioned into the self-as-agent (the acting self), the self-as-experiencer (the experiencing self) and the self-as-subject (the self that can be acted upon by other entities in the environment).

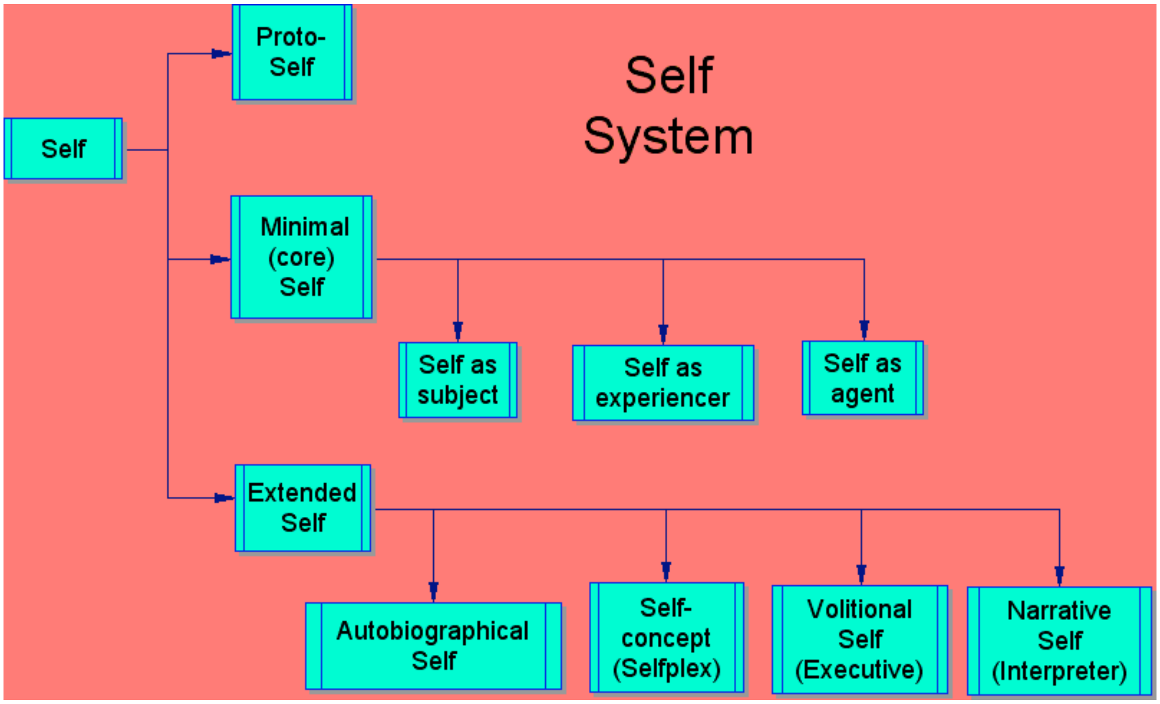

Figure 1. The various self systems

The *extended self*, consisting of the autobiographical self, the self concept, the volitional or executive self and the narrative self is attributed to human and, perhaps, higher animals. The *autobiographical self* derives directly from episodic memory, the memory of event, the what, the where, and the when (Baddeley, Conway and Aggleton 2001, Franklin et al 2005). The *self concept*, also referred to as the *context of self* (Baars 1988) or the *self-plex* (Blackmore 1999) is spoken of as consisting of enduring self beliefs and intentions, particularly those having to do with personal identity. The *volitional self* provides executive function (Baars 1988). Finally, the *narrative self* is able to report, sometimes equivocally, contradictorily or self-deceptively, on actions, intentions, etc. (Gazzaniga 1998).

7. A LIDA Controlled Robot with a Self

While the LIDA model implements much of Global Workspace Theory, it does not, as yet, include a self. GW theory postulates a mostly unconscious, many-layered self system that, through context, influences conscious events. This motivates Baars to propose that "… attempts at machine consciousness must incorporate such many-layered self systems." (Baars 1988 Chapter 9) Here we'll explore possible ways of implementing the various selves in a LIDA controlled mobile robot. Again, our idea is that a LIDA controlled robot provided with a self-system might be more likely to be phenomenally conscious than one without. This assumption is made because we attribute phenomenal consciousness to humans, and they have selves that seem to be involved in their consciousness (Baars 1988 Chapter 9, Baars, Ramsoy and Laureys 2003).

Implementing a minimal self: How might a minimal, or core, self be implemented in the LIDA model of cognition? We suggest that all three aspects of the minimal self be implemented as sets of entities in the LIDA ontology (Franklin and Ferkin 2006), that is, computationally as collections of nodes in LIDA's perceptual associative memory (PAM). PAM is implemented as a semantic net with activation called the slipnet, a la Hofstadter and Mitchell's Copycat architecture (1994). Nodes of the slipnet constitute the agent's perceptual symbols (Barsalou 1999), representing individuals, categories, and perhaps higher-level ideas and concepts. PAM provides an integrated perceptual system for LIDA, allowing the system to recognize, categorize and understand.

Let's implement *self-as-agent* as the set of action nodes in PAM, that is nodes representing actions by the agent such as lie-down, stand, roll-over, walk, glance-left, etc. Having such action nodes in the slipnet would allow actions to 1) be part of structure building in the workspace, 2) to be included in cues to episodic memories, 3) to come to consciousness, 4) to be written to episodic memory as part of events, 5) to be available for the creation of new schemes by the procedural learning mechanism. This kind of implementation would give such actions first-class status among the ontological entities of the LIDA model. Self-as-agent would then be realized as the set of all action nodes in the slipnet that represent actions to be taken by the LIDA agent, together with their "halo" in PAM, that is, with the nodes to which they are directly linked.

The actions discussed here are used in the LIDA action *selection* process, and must be carefully distinguished from sensory-motor automatisms (SMA's) that implement action *execution*. From a neurobiological perspective applied to vision, these actions are part of the ventral stream, while the SMA's receive unconscious sensory input from the dorsal stream (Goodale and Milner 2004). Recall that perception (the ventral stream) serves to make sense of the current environmental situation utilizing scene base frames of reference and relative metrics, all in the service of action selection. In contrast, SMA's (the dorsal stream) employ egocentric frames of reference and absolute metrics to effect the carrying out of these actions.

We may also want to include as part of self-as-agent LIDA' expectation codelets[2]. A particular kind of attention codelet spawned with each action selected, an expectation codelet attempts to bring to consciousness any items in the workspace that bear on how well the action achieved its expected result. Expectation codelets are already part of the LIDA model.

Similarly, let's implement *self-as-subject* as the set of acted-upon nodes in PAM, that is nodes representing actions by other entities upon the agent such as being pushed, stroked, hugged, slapped, yelled-at, fallen-upon, etc. We must also include in self-as-subject LIDA's attention codelet that wants to bring to consciousness the fringe consciousness feeling (Mangan 2001) of having been acted upon.

The self-as-experiencer might be thought of as being comprised of all of the rest of PAM. Thus the minimal self seem to be relatively easy to implement in a LIDA controlled mobile robot.

Implementing an extended self: Recall that the extended self consists of the autobiographical self, the self-concept, the volitional self and the narrative self. In contrast to the several aspects of the minimal self, those of the extended self will require different kinds of implementations in a LIDA controlled mobile robot. The autobiographical self can already be found in the LIDA model as its autobiographical memory, one part of its declarative memory. Similarly, the beliefs about the self that are contained in LIDA's semantic memory, the other part of declarative memory, seem to already constitute a self concept.

Following Global Workspace Theory, LIDA's volutional decision making process, its volitional or executive self, is implemented using William James's ideomotor theory (James 1890, Baars 1988, Franklin 2000). It was already computationally implemented in the original IDA software agent, and is currently an integral part of the LIDA conceptual model.

The narrative self requires the most effort to implement in the LIDA model due to the need for higher-level cognitive processes. Gazzaniga's narrative self responds to queries about perceptions, actions and motivations (1998). Understanding the query will require adding appropriate nodes to LIDA's PAM, needed structure-building codelets to operate in her workspace building new perceptual structures with which to understand the query, and attention codelets to bring such queries to consciousness. Motivation for responding will result from a feeling node in PAM as the source of pleasure at producing a reply. Finally, we must add behavior streams[3] sufficient to generate understandable replies. Thus implementing a

[2] Codelets are small pieces of code running as independent threads, each of which is specialized for some relatively simple task.
[3] Behavior streams are sets of behaviors, partially ordered in an and/or fashion (Franklin et al 2005).

narrative self in a LIDA controlled robot would be far and away the most difficult undertaking of all the implementations of self. Except for the narrative self, the other extended selves are already included in the LIDA model, or seem relatively straightforward to implement.

But what is the connection between self and consciousness? Why might the addition of such a self-system to a robot make it more likely to be phenomenally conscious? If one damages the self-system of a human, than conscious contents may also disappear. Recall that in people with split brains, deep dissociation like dissociative identity disorder, fugue states, some cases of deep hypnosis (see Ernest Hilgard's book, Divided Consciousness), the dissociated executive loses access to the conscious contents of the other executive (Baars 1988 Chapter 9, Baars, Ramsoy and Laureys 2003).

8. Is Phenomenal Consciousness Possible in such a Robot?

We humans attribute phenomenal consciousness to other humans because we experience it in ourselves, and perceive others as being similar to us. Most of us don't take seriously the possibility of a zombie, in the philosophical sense (Chalmers 1995), because there is no evidence that such a being could exist. Attribution of phenomenal consciousness to animals often results, as with attributions to humans, from the similarity of their nervous systems to ours (Seth, Baars and Edelman 2005), or from the similarity of their behaviors to ours (Mather 2008). But, why might one attribute phenomenal consciousness to a robot? Certainly not because of any similarity of nervous systems. Perhaps because of similarity in behavior. We would have no problem attributing phenomenal consciousness to a robot such as Star Trek the Next Generation's Commander Data, were he real rather than fictional. Recent experimental evidence suggests the likelihood of such attribution to artificial entities that behave like humans (Krach et al 2008). Another possibility is attribution because of the similarity in the control architecture (mind) of the agent, be it human or robot.

Might a LIDA controlled robot that produces a stable, coherent perceptual world, as described above, be subjectively conscious? What about one that, in addition, had all these selves? It would seem at least possible for several reasons. Such a robot would be functionally conscious. Based on the LIDA architecture, which is both psychologically and neuro-scientifically grounded, its control structure would be quite similar to that of a human. In addition, it would satisfy Merker's coherent, stable perceptual world condition, and would possess all the various selves. But, might not other, additional, and as yet unknown, processes be needed in order to enable phenomenal consciousness in a robot? Indeed, they might. Note how Merker's work, and the implementation of selves, gives direction to robot designers attempting to produce conscious robots. We claim that building a robot as described above might well prove to be a significant step towards producing a phenomenally conscious robot. Further, we suspect that humans may eventually l build conscious software agents and/or conscious robots, so intelligent, so sophisticated, and so communicative that people will simply assume that they are sentient beings with phenomenal consciousness.

Irrespective of whether we succeed in building phenomenally conscious machines, there is the important issue of developing methodologies to experimentally verify that a machine claiming phenomenal consciousness, is in fact conscious. Accurate report (AR) or verifiable report is perhaps the simplest methodology by which claims of phenomenal consciousness in humans can be empirically validated (Baars, 1997). According to an AR inspired methodology, phenomenal consciousness can be attributed to an entity that consciously reports on some event that the experimenter can independently verify. For example, phenomenal consciousness can be attributed to a participant that can report seeing a blue circle appear on a computer screen, if the experimenter can confirm that he or she did in fact display a blue circle (not red, green, or black circle) on the screen.

Although AR is a simple and useful tool for studies involving human participants, it is unlikely to suffice as the singular condition for machine consciousness. For example, it is simple to design a machine that monitors a computer screen and reports on the presence of blue circles versus green squares. An experimenter can easily verify that the machine is accurate and that all requirements for AR have been met.

Nevertheless, it would be unreasonable to argue that such a machine is phenomenally conscious. So why the difference between the utility of AR in humans versus machines? The answer lies in some of the assumptions that are made well before the AR experiment commences. The AR paradigms work for humans because the attribution of phenomenal consciousness in humans is made prior to the actual experiment. The experiment is merely confirming this attribution. It is of course reasonable to attribute phenomenal consciousness to humans because we perceive it in ourselves. But the story is more complicated for conscious machines. Their mechanisms are so different from our biological and cognitive underpinnings that it is difficult to reasonably defend an attribution of phenomenal consciousness in machines.

But the future of the field of machine consciousness hinges on the ability to verify or refute allegedly conscious machines. Therefore, methodologies to empirically verify claims of phenomenal consciousness must parallel the development of phenomenally conscious machines. Similar to the criteria for consciousness in non-human mammals (Seth, Baars, and Edelman's 2005), a criteria for consciousness in machines must emerge.

References

1. Aleksander, I. 2000. *How to Build a Mind*. London: Weidenfeld and Nicolson.
2. Baars, Bernard J. 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
3. Baars, B.J. (1997). *In the Theater of Consciousness*: The Workspace of the Mind. NY: Oxford University Press.
4. Baars, B J. 2003. How brain reveals mind: Neural studies support the fundamental role of conscious experience. *Journal of Consciousness Studies* 10: 100–114.
5. Baars, Bernard J and Stan Franklin. 2003. How conscious experience and working memory interact. *Trends in Cognitive Science* 7: 166–172.
6. Baars, B J, T Ramsoy, and Baars, B J, T Ramsoy, and S Laureys. 2003S Laureys. 2003. Brain, conscious experience and the observing self. *Trends Neurosci.* 26: 671–675.
7. Baddeley, Alan, Martin Conway, and John Aggleton. 2001. *Episodic memory*. Oxford: Oxford University Press.
8. Barsalou, L W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22: 577–609.Bekoff, Marc and Paul W Sherman. 2004. Reflections on animal selves. *TRENDS in Ecology and Evolution* 19: 176–180.
9. Blackmore, Susan. 1999. *The meme machine*. Oxford: Oxford University Press.
10. Castet, Eric and Guillaume S Masson. 2000. Motion perception during saccadic eye movements. *Nature Neuroscience* 3: 177 - 183.
11. Chalmers, David. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2, no. 3: 200-219.
12. Cotterill, R M J. 2003. Cyberchild: A simulation test-bed for consciousness studies. *Journal of Consciousness Studies* 10: 31–45.
13. Damasio, Antonio R. 1999. *The feeling of what happens*. New York: Harcourt Brace.
14. Franklin, Stan. 2000. Deliberation and voluntary action in 'conscious' software agents. *Neural Network World* 10: 505–521
15. Franklin, Stan. 2001. Automating human information agents. In *Practical applications of intelligent agents*, ed. Z Chen and L C Jain: 27–58 Berlin: Springer-Verlag.
16. Franklin, Stan. 2003. IDA: A conscious artifact? *Journal of Consciousness Studies* 10: 47–66.
17. Franklin, Stan. 2005. Evolutionary pressures and a stable world for animals and robots: A commentary on merker. *Consciousness and Cognition* 14: 115–118.
18. Franklin, S, B J Baars, U Ramamurthy, and Matthew Ventura. 2005. The role of consciousness in memory. *Brains, Minds and Media* 1: 1–38, pdf.
19. Franklin, Stan and A C Graesser. 1997. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent agents iii*:21–35. Berlin: Springer Verlag.
20. Franklin, Stan and Michael H Ferkin. 2006. An ontology for comparative cognition: A functional approach. *Comparative Cognition & Behavior Reviews* 1: 36–52.
21. Franklin, S, A. Kelemen, and L. McCauley. 1998. IDA: A Cognitive Agent Architecture. In IEEE Conf on Systems, Man and Cybernetics: IEEE Press.
22. Franklin, Stan and F G Jr Patterson. 2006. The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent. In *IDPT-2006 Proceedings Integrated Design and Process Technology)*: Society for Design and Process Science.

23. Gallagher, Shaun. 2000. Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Science* 4: 14–21.
24. Gazzaniga, Michael S. 1998. *The mind's past*. Berkeley: University of California Press.
25. Goodale, M. A., and D. Milner. 2004. Sight Unseen. Oxford: Oxford University Press.
26. Haikonen, Pentti O. 2007. *Robot brains; circuits and systems for conscious machines*. UK: Wiley and Sons.
27. Holland, Owen. 2007. A strongly embodied approach to machine consciousness. *Journal of Consciousness Studies*. Special Issue on Machine Consciousness.
28. James, William. 1890. *The principles of psychology*. Cambridge, MA: Harvard University Press.
29. Krach, S, F Hegel, B Wrede, G Sagerer, and Binkofski F. 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3, no. 7: e2597.
30. Mather, Jennifer A. 2008. Cephalopod consciousness: Behavioral evidence. *Consciousness and Cognition 17, no. 1: 37-48.*
31. Merker, Bjorn. 2005. The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition* 14: 89–114.
32. Mangan, Bruce. 2001. Sensation's ghost: The non-sensory "Fringe" Of consciousness. *Psyche* 7: http://psyche.cs.monash.edu.au/v7/psyche–7–18–mangan.html.
33. Ramamurthy, Uma, Sidney K D'Mello, and Stan Franklin. 2006. LIDA: A computational model of global workspace theory and developmental learning. In *BICS 2006: Brain Inspired Cognitive Systems*.
34. Seth, A K, B J Baars, and D B Edelman. 2005. Criteria for consciousness in humans and other mammals. *Consciousness and Cognition* 14: 119–139.
35. Shanahan, M P. 2006. A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition* 15: 433-449.
36. Strawson, G. 1999. The self and the sesmet. In *In models of the self*, ed. Shaun Gallagher and J Shear:483–518. Charlottesville, VA: Imprint Academic.