

Commentary on Webb

Abstract: 64 words

Main Text: 399 words

References: 364 words

Total Text: 809 words

Models as Implementations of a Theory, rather than Simulations: Dancing to a Different Drummer

Stan Franklin

Institute for Intelligent Systems

University of Memphis

Memphis, TN

38152

USA

franklin@memphis.edu

<http://www.msci.memphis.edu/~franklin>

Abstract

Robots, as well as software agents, can be of use in biology as implementations of a theory rather than as simulations of specific real world target systems. Such implementations generate hypotheses rather than representing them. Their behavior is not predicted, but rather observed, and is not expected to duplicate that of a target system. Scientific knowledge is gained through the testing of generated hypotheses.

Robots, or software agents, can be of value to biology as generators of hypotheses as well as the in the more traditional ways so well described by Webb. Let me illustrate with a software agent example. Essentially the same architecture could control a robot.

IDA (Franklin, 2000b; Franklin, Kelemen, & McCauley, 1998) implements global workspace theory, a psychological and neuropsychological theory of consciousness and cognition (Baars, 1988, 1997; Franklin & Graesser, A., 1999), as well as parts of other theories (Barsalou, 1999; Glenberg, 1997; Kintsch, 1998; Sloman, 1999). Supported by the US Navy, IDA is intended to replace a human personnel agent in the real world task of job distribution (Franklin, 2001).

Models such as IDA have the potential to play a synergistic role in modeling a scientific theory. The theory constrains the design of the agent or robot that implements (models) that theory. While a theory, including many from cognitive neuroscience, may be abstract and only broadly sketch an architecture, an implemented computational design provides a fully articulated architecture and a complete set of at least computational mechanisms. This architecture and set of mechanisms provides a richer, more concrete and more decisive theory, as well as both a conceptual and a computational model. Moreover, every design decision taken during an implementation translates into a hypothesis that serves to flesh out the theory. These hypotheses may motivate experiments to test them thus providing direction for biological research. Conversely, the results of such experiments motivate corresponding modifications of the architecture and mechanisms of the agent or robot and the cycle starts again. These ideas have been discussed more fully elsewhere (Franklin, 1997).

IDA also implements William James' theory of voluntary action (Franklin, 2000a; James, 1890; Kondadadi & Franklin, 2001). Experimental work of neuroscientist Benjamin Libet lends support to this implementation of voluntary action as mirroring what happens in humans (1983, 1999). The IDA implementation offers a new interpretation of Libet's work and suggests possible experimental tests (Franklin & Graesser, A., 2001).

It is also likely that observing the behavior of such software agent or robotic implementations will lead to additional hypotheses that can then be tested. As of this writing a full version of IDA is being tested, and will soon be running continuously so that such observations can be made.

References

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

- Baars, B. J. (1997). *In the Theater of Consciousness*. Oxford: Oxford University Press.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.
- Franklin, S. (1997). Autonomous Agents as Embodied AI. *Cybernetics and Systems*, 28, 499-520.
- Franklin, S. (2000a). Deliberation and Voluntary Action in 'Conscious' Software Agents. *Neural Network World*, 10, 505-521.
- Franklin, S. (2000b). Modeling Consciousness and Cognition in Software Agents. In N. Taatgen (Ed.), *Proceedings of the International Conference on Cognitive Modeling*, Groenigen, NL, March 2000 (Series Eds. 100-109). : .
- Franklin, S. (2001). Automating Human Information Agents. In Z. Chen & L. C. Jain (Eds.), *Practical Applications of Intelligent Agents*. Berlin: Springer-Verlag.
- Franklin, S., & Graesser, A. (2001). Modeling Cognition with Software Agents. In J. D. Moore & K. Stenning (Eds.), *CogSci2001. Proceedings of the 23rd Annual Conference of the Cognitive Science Society (August 1-4, 2001, Series Eds. 301-306)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Franklin, S., & Graesser, A. (1999). A Software Agent Model of Consciousness. *Consciousness and Cognition*, 8, 285-305.
- Franklin, S., Kelemen, A., & McCauley, L. (1998). IDA: A Cognitive Agent Architecture. In *IEEE Conf on Systems, Man and Cybernetics (Series Eds. 2646-2651)*. : IEEE Press.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1-19.
- James, W. (1890). *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
- Kintsch, W. (1998). *Comprehension*. Cambridge: Cambridge University Press.
- Kondadadi, R., & Franklin, S. (2001). A Framework of Deliberative Decision Making in "Conscious" software Agents. In *Proceedings Of Sixth International Symposium on Artificial Life and Robotics (AROB-01) (Series Eds. 280-283)*.
- Libet, B. (1999). Do We Have Free Will? *Journal of Consciousness Studies*, 6.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential: The Unconscious Initiation of a Freely Voluntary Act. *Brain*, 106, 623-642.
- Sloman, A. (1999). What Sort of Architecture is Required for a Human-like Agent? In M. Wooldridge & A. Rao (Eds.), *Foundations of Rational Agency*. Dordrecht, Netherlands: Kluwer Academic Publishers.