Available online at www.sciencedirect.com

SCIENCE DIRECT®

Artificial Intelligence 169 (2005) 181–183

Artificial Intelligence

www.elsevier.com/locate/artint

Book review

Jeff Hawkins and Sandra Blakeslee, On Intelligence, Times Books, 2004.

On intelligence as memory

Jerome A. Feldman

International Computer Science Institute, Berkeley, CA 94704-1198, USA

Available online 3 November 2005

On Intelligence by Jeff Hawkins with Sandra Blakeslee has been inspirational for non-scientists as well as some of our most distinguished biologists, as can be seen from the web site (http://www.onintelligence.org). The book is engagingly written as a first person memoir of one computer engineer's search for enlightenment on how human intelligence is computed by our brains. The central insight is important—much of our intelligence comes from the ability to recognize complex situations and to predict their possible outcomes.

There is something fundamental about the brain and neural computation that makes us intelligent and AI should be studying it. Hawkins actually understates the power of human associative memory. Because of the massive parallelism and connectivity, the brain essentially reconfigures itself to be constantly sensitive to the current context and goals [1]. For example, when you are planning to buy some kind of car, you start noticing them. The book is surely right that better AI systems would follow if we could develop programs that were more like human memory. For whatever reason, memory as such is no longer studied much in AI—the Russell and Norvig [3] text has one index item for memory and that refers to semantics.

From a scientific AI/Cognitive Science perspective, the book fails to tackle most of the questions of interest. It is certainly true that vision, motor control, language, planning, learning, etc. involve recognizing new situations as similar to known ones, but memory alone does not address any of the core issues in these areas. Scientists (whether in AI or biology) study particular phenomena such as color, grammar, feedback, evidence combination, and neural development. The book should be read as suggesting how one essential component of intelligence might be realized by the human brain as we understand it.

E-mail address: feldman@icsi.berkeley.edu (J.A. Feldman).

doi:10.1016/j.artint.2005.10.010

The technical content of the book is concentrated in Chapter 6, modestly entitled "How the Cortex Works". Vision is described in terms of a four-level abstraction hierarchy—V1, V2, V4 and IT. There is a good discussion of how hierarchical top down feedback in such a system can be viewed as prediction and this architecture is extrapolated to other senses. Abstraction is nicely portrayed as the activity of a group of neurons constituting a "name" that covers a variety of equivalent lower level firings. Given these names, invariant sequences can be learned and then used in prediction.

The nearest thing in the book to an explicit computational model is a detailed discussion of the structure of the six cortical layers. Hawkins starts with some rather detailed facts about cell types and connectivity, but then adds speculative features even in the face of massive contrary evidence. For example, a great deal is known about motor control and skill memory and it bears no resemblance to Hawkins' model. And vision itself is known to involve dozens of cortical areas with complex internal structure and interactions.

But big picture thinkers do this kind of thing. Using his simplified model of the brain, Hawkins discusses how hierarchical feedback memory might be implemented. Any feedback system that abstracts needs some way to direct the top-down feedback to appropriate instances. The following quote (p. 143) conveys Hawkins' level of precision in addressing such issues:

> "Similarly, when I hear the next note in a melody, my brain has to take a generic interval such as a fifth and convert it to the correct specific note, such as C or G. The horizontal flow of activity across layer 1 provides the mechanism for doing this. For high level invariant predictions to propagate down the cortex and become specific predictions, we must have a mechanism that allows the flow of patterns to branch at each level. Layer 1 fits the bill. We could predict the need for it, even if we didn't know it existed."

The most concrete algorithm is a Hebbian learning story (p. 148) that seems to depend on magical reciprocal connections from layers 2, 3, and 5 to layer 1. Strengthening synapses INTO layer 1 somehow increases activation FROM layer 1 cells and thus provides prediction. Using this magic, Hawkins goes on to speculate on how a cortical column might learn to activate a particular name if and only if an appropriate corresponding sequence of lower level names was active. The algorithm is computationally reasonable and appears to be a standard structured connectionist circuit, depending on top down feedback. No explicit learning rules are suggested.

Any memory-based brain theory must deal with the Hippocampus. Hawkins cleanly refutes the old idea that memories are first stored in the Hippocampus and then somehow transferred to cortex. But then, inexplicably, he goes on to suggest that novel situations are treated exactly like that—they get transferred from the Hippocampus to the cortex. There are several well-developed computational models of Hippocampal function that do deal with novel experience and memory [2]. Bizarrely, the book has no references to the several excellent books on human memory such as those by Schacter [4] and Squire [5].

This leaves us with an intriguing question—why are these distinguished scientists rhapsodizing over the book? In my experience, most neuroscientists don't think at all computationally (just as most AI researchers don't think at all biologically). The memory-based story is easy to understand and is consistent with some of our intuitions. It must be comforting for biologists to have a computer expert say that we don't need to deal with computation in order to understand how brain creates the mind.

Perhaps this simplistic intelligence-as-memory story will give rise to new fruitful experiments, which would be great. There are some experimental predictions in the Appendix. Most of these concern properties of neurons (in various cortical layers) that would be consistent with the theory. The most general prediction is that "Representations move down the hierarchy with training". I can't think of an interpretation of this that is plausible, but who knows.

Human memory is a crucial part of our intelligence and still vastly outperforms any computer system. A biologically and computationally testable model of human associative memory would not solve the AI problem, but it would be a major advance. If *On Intelligence* helps generate effort on this important topic, it will have earned its plaudits.

References

[1] J. Feldman, From Molecule to Metaphor: A Neural Theory of Language, MIT Press, Cambridge, MA, 2006.
[2] L. Nadel, M. Moscovitch, The Hippocampal complex and long-term memory revisited, Trends in Cognitive Sciences 5 (2001) 228–230.
[3] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Pearson, Upper Saddle River, NJ, 2003.
[4] D.L. Schacter, Searching for Memory: The Brain, the Mind, and the Past, Basic Books, New York, 1996.
[5] L.R. Squire, E.R. Kandel, Memory: From Mind to Molecules, W.H. Freeman, New York, 2000.