On Intelligence

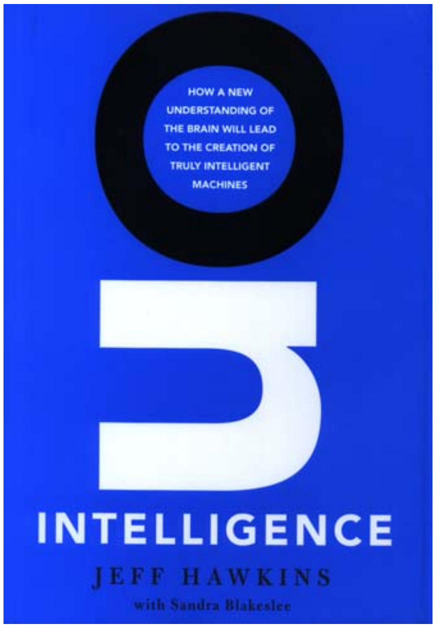
BY JEFF HAWKINS, HENRY HOLT & CO, NY, 2004. ISBN 0-8050-7456-2

REVIEWED BY MIKE HOWARD[1]

It naturally makes researchers bristle when a new guy comes along and makes grand pronouncements that everyone else has missed the big idea. Stephan Wolfram encountered some of that with his 1192 page tome, "A New Kind of Science" where he argued that cellular automata do much better at simulating complex phenomena than more sophisticated models. It helps things not a bit when the new guy struck it rich by inventing a cool tech gizmo and now wants to do "real science." Jeff Hawkins is such a man, asserting that he has studied the esoteric findings of brain researchers, and has new insight about how it all fits together.

Hawkins' credentials as a creative person with plenty of his own natural intelligence are stunning. Every review of Jeff Hawkins' new book begins with a note that he is the brilliant entrepreneur and computer expert who founded Palm Computing and Handspring. He invented the PalmPilot, the Treo smart phone and other gadgets. One of his gifts is being able to make technology simple and accessible, and that goes for his writing as well.

In "On Intelligence," Jeff Hawkins presents a new theory about how the brain works and how we can finally build "intelligent" machines. Of course, machine intelligence has been a goal of computer science for decades. He discusses the success of Artificial Intelligence in the 60's and 70's that led to an irrational exuberance when the popular press started hyping the possibility of creating artificial brains. In the 90's the bubble burst when the fickle press discovered those wild expectations were unrealistic, and researchers nearly had to stop using the term "AI".



Progress continued to be made in the foundations of AI of course, and it was good for the field when the hype died down. Fuzzy logic chips appeared in washing machines, automated planning systems controlled NASA space probes including the Mars Rover, and expert systems are used by banks to decide who to loan money to. Neural networks have had success in pattern recognition. But although great strides have been made in machine learning, and Moore's Law has made many AI algorithms practical, it is clear that AI alone will not result in a robot that can do your errands for you or babysit your children.

Hawkins points out that the Turing Test is not, after all, a good indicator of intelligence. John Searle, a philosopher and cognitive scientist who created a thought experiment called the "Chinese Room," indicated that while a computing device could indeed reply to questions in such a way that made it indistinguishable from a human being, it did not understand the conversation and there was no meaning attached to its replies. Although Searle's experiment is controversial, it suggests that Turing's test is faulty and misleading.

Hawkins believes that our best bet for learning to build truly intelligent machines is to learn how the brain actually operates. In particular, it is the neocortex, the center of higher thought, which is the focus of his attention. But he also implies that neuroscientists are lost in the complexity of mapping out neural pathways, and are not coming up with compelling overarching theories that begin to explain how we think and learn.

Hawkins believes there is enough evidence now to posit a common cortical algorithm, as first proposed by Vernon Mountcastle, a neuroscientist at Johns Hopkins, in 1978. The algorithm is hierarchical, with lower layers encoding data from a sensory organ, but higher layers dealing with abstract signals that bear little resemblance to the sensory signals. Hawkins asserts that brain researchers got sidetracked partly due to the experimental difficulty of taking measurements. The standard approach is to present a static sensory stimulus and take readings of resulting cortex activity. It is too difficult to work with dynamically changing stimuli, so researchers have missed a point that Hawkins believes is crucial: the brain can only perceive dynamic stimuli.

Since the author believes that brain research is mired in complexity, a higher level theory is needed to provide a top-down pressure to guide the field. By the way, he makes no mention of Minsky's "Society of Mind," a metaphorical / philosophical thought experiment less constrained by brain research than what Hawkins had in mind.

Hawkins' theory, called "Memory Prediction Framework," defines *Intelligence* as "the capacity of the brain

[1] mhoward@hrl.com, HRL Laboratories, LLC

to predict the future by analogy to the past." According to him, there are four key attributes of neocortical memory that differ from computer memory:

- All memories are inherently sequential.
- Memory is auto-associative; a partial memory can be used to retrieve the full memory.
- Memories are stored in invariant representations.
- Patterns are stored in a hierarchy.

Support for the theory is most concretely expressed in chapter six, the meatiest part of the book. This is where the author describes in some detail his vision of how the neural circuitry in the layers of cortex works. The description is compelling, but takes more work to follow than the other chapters.

Chapter six ends with several fascinating observations that are built on top of the neural circuitry described earlier. It emphasizes that perception and behaviour are highly interdependent because they both originate in a detail-invariant representation that is then transmitted through both motor and sensory cortex. Also, although many researchers have discounted it, Hawkins argues that feedback and the importance of distant synapses in cortex is essential to explain the Memory Prediction Framework theory, and should be reconsidered. The theory includes the broad principles of how hierarchical learning of sequences explains how children first learn letters, then words, phrases and finally sentences, and as adults we can speed-read without needing to study every letter. The author believes that the memory of sequences re-forms lower and lower in cortex, allowing higher layers to learn more complex patterns. Finally, the hippocampus is briefly described as logically residing at the top of the cortical hierarchy: the short-term repository of new memories.

An impressive result of the speculations in chapter six is a list in the appendix of 11 specific, testable predictions made by the theory, which is an invitation to brain researchers. And Hawkins founded a company, Numenta, to develop the Hierarchical Temporal Memory concept based on the theory.

Chapter six also hints at how daydreaming or imagining occurs, when predictions from layer 6 of a cortical column are fed back to layer 4 of the same column. Cortical modeller Stephan Grossberg calls this "folded feedback". In chapter seven the book expands on philosophical speculation about the origin of consciousness and creativity that arise from the Memory Prediction Framework theory. Creativity is defined here as "making predictions by analogy". As the author says, there is a continuum of creativity, from mundane extrapolations from learned sequences in sensory cortex to rare acts of genius. But they have a common origin. This is how a piano player can quickly figure out how to play simple melodies on a vibraphone, or a customer in a strange restaurant can figure out that there is probably a restroom in the back. Creativity is so pervasive that we hardly label it as such, unless it violates our predictions like an unusual work of art. There are practical suggestions in this section for how to train oneself to be more creative, and an interesting story of how Hawkins conceived the handwriting recognition system, Graffiti.

Chapter seven ends in speculation about the nature of consciousness, imagination and reality in response to the inevitable questions to which this type of work gives rise. A review on the Amazon website by Dr. Jonathan Dolhenty takes issue with what he describes as "plain old-fashioned metaphysical materialism and, probably, old-school psychological behaviourism," which are largely discounted theories today. Dolhenty is a philosopher who thinks human intellect at the higher abstract and conceptual levels cannot be described by such a simple extrapolation of the Memory Prediction Framework. But this reviewer found the connections made between brain theory and "mind" reassuring. Leave it to others to build on this foundation. In fact, Hawkins does hint at a broader source of the mind in chapter seven, where he says that it is influenced by the emotional systems of the old brain and by the complexity of the human body.

The last chapter in the book contains another vision, of how intelligent machines might be built in the future. This is back into the Popular Science mode. Unlike many current roboticists who believe humanoid robots will be needed to interact with humans, Hawkins believes humanoid form is pointless and impractical. He advocates working from inside out, by building sensing mechanisms and attaching them to a hierarchical memory system that works on cortex principles. Then by training the system he believes it will develop its own representations of the world. This system can be built into any sort of machine, and the sensors can be distributed if desirable.

The technical challenges of building an intelligent machine include *capacity*, which by analogy to the brain, at 2 bits per synapse, would require 8 trillion bytes of memory or about 80 hard drives. *Connectivity* is a larger problem, since it would be impossible to provide dedicated connections. Hawkins believes the answer would be some sort of shared connections, like in today's phone network, but this is still a challenge.

As an aside, there is no mention of the Cyc project, which has been working since 1984 to build a mammoth semantic knowledge base. But unlike the automatically learned representations in Hawkins' proposed artificial brain, the ones in Cyc are hand-input in a preconceived structure as a vast quantity of terms related by assertions.

The last chapter ends with a very positive view of the potential of intelligent machines to solve problems humans cannot, because they can be equipped with custom senses, immense memory, and even be networked to form hierarchies of intelligent machines. Hawkins believes that intelligent machines will be a hot topic in the next ten years. It is easy to get caught up in his excitement.

<u>Related Websites</u>:
http://www.onintelligence.org/
http://redwood.berkeley.edu/
http://www.numenta.com/