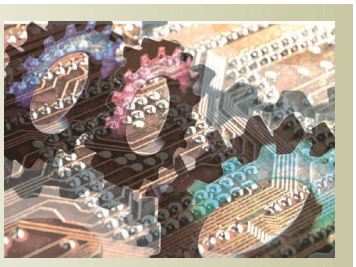
Machine Intelligence Meets Neuroscience

Bob Colwell

When I was finishing high school in the early 1970s, the closest normal people got to computers was using a teletype. These electromechanical monsters could convert bits received at glacially slow (110 baud) telephone rates into a veritable cacophany of sound. This acoustic din was produced by lots of metal parts slamming into a paper roll somewhere inside the machine. The paper was the same consistency and smoothness as the paper towels you find in restrooms, and the bit rate was caused partly by the constraints of the acoustic couplers used in those days. I probably still have the pieces of the 300-baud modem I built out of two op-amps and two tennis shoes.



With each passing year, the goal of achieving "artificial intelligence" seems to recede further into the distance.

THE EARLY DAYS

My personal introduction to computing occurred when a friend suggested that we sneak into the school's "computer room" to see what a computer looked like. After suitable furtiveness, we used the key I had that opened almost everything at the school to enter the room. We sidled up to the weird contraption and peered inside.

The keyboard looked more or less familiar, but there was a lot more stuff in there than on any typewriter I'd ever seen. I also noticed that the "computer" was making a humming sound, almost like electronic breathing. Just as I moved closer for a better look, the entire lashup started clattering, bouncing, gyrating, and emitting crashing waves of sound. I didn't know if that was normal, but it sure didn't sound like anything I'd ever heard before.

I knew that I hadn't touched anything—although I probably did have some vague plans in that direction—and I hadn't seen my friend touch anything either. So now I was confronted with an even bigger puzzle: What caused the beast to leap into action? And what was it doing, anyway?

With hands over our ears, we looked closer and saw that words were being printed on the endless paper towel rolling out of the top. At the time, we thought the machine was printing those words at a prodigious rate, but I'm sure whatever speed it was would be taken as a joke today.

I left in wonderment. What was making the machine print out recognizable English words, even if I couldn't then understand the importance of the patterns or why the word "Basic" appeared on every page? The way the machine printed for a while, then seemed to think for a while, then printed some more almost seemed intelligent.

After thinking about it for a while, I guessed that an agent outside the room was somehow controlling this machine, but how that was happening wasn't at all clear. That hypothesis explained the observable phenomena, but I still wondered how smart the thing I couldn't see was.

WHAT IS INTELLIGENCE?

Then I wondered what "smart" was in the first place. I knew that I sometimes had come up with some good ideas, but I also knew that I was perfectly capable of doing things universally regarded as Truly Stupid Maneuvers. So is "smart" something we *are* or a something we *do*?

More broadly, what is this notion of intelligence, and where did it come from? To the extent that we are made of physical materials that work in predetermined ways—a useful if minimalist definition of a machine—it's already clear that machines can be intelligent.

I'm purposely skating quickly past topics such as solipsism (you're all a figment of my imagination), mysticism (which might be right but doesn't lead to any fundamental analysis), and infinite digressions such as quantifiability, creativity, and dolphins. And if we do constitute an existence proof of an intelligent machine, why aren't we intelligent enough to make machines that aren't carbon-based wetware?

This all sounded very logical, almost inevitable, and I was entering college right at the beginning of it. Perhaps I could help make this new intelligence happen. It was all very exciting.

Published by the IEEE Computer Society

A few years later, I heard about the efforts at many universities where they were trying to achieve something called "artificial intelligence." That sounded like fun to me: We could learn a lot about ourselves, maybe gain useful insights into various mental problems, and make more capable machines in the bargain. It seemed reasonable that whatever kind of intelligence these machines exhibited, it would probably be different than human intelligence, in the same way that aircraft don't fly like birds, and boat propulsion doesn't mimic a fish.

What artificial intelligence might achieve for us seemed boundless. When humans write computer programs, for example, they make a lot of errors. Why not let intelligent machines write the programs perfectly every time? They could also write poems, stories, books, music. Give them the rules of chemistry, and they would spit out new drug formulations. They would solve Fermat's Last Theorem, ponder physics' Grand Unification Theories, and tell us who was hiding on the grassy knoll. What an amplification of the human mind.

But that was more than 25 years ago, and it's not the way things actually played out. AI researchers learned many useful things about what intelligence is not, harnessing computers to perform tasks in ways that are useful and illuminating but ever-more-divorced from the original goal of creating new intelligence. I never completely lost faith that maybe we'd eventually get there somehow, but with each passing year, the goal seemed to recede further into the distance.

A NEW UNDERSTANDING

In October 2004, Jeff Hawkins (of Palm Pilot fame) rode into AI town and shot up the place. The front cover of his new book titled *On Intelligence* (Times Books, 2004) proposes that "a new understanding of the brain will lead to the creation of truly intelligent machines."

Just glancing at the cover instantly revived neurons that hadn't had much stimulation since 1977, and it was great fun having Hawkins kick-start them again. It turns out that Hawkins has been a neuroscience aficionado for decades, but only recently got to devote enough time to the topic to really think some things through and put his arguments in writing.

In addition to providing a retrospective of the AI and neuroscience fields, Hawkins outlines his own theories of how the neocortex produces intelligence and the nature of intelligent machines that will be viable if we can develop a model of neural connectivity and operation (perhaps his own) and find ways to map that model onto hardware.

AI researchers learned many useful things about what intelligence is not.

Intelligence and intelligent behavior

Hawkins begins by carefully dissecting the connection between intelligence and intelligent behavior. This point resonated with me twice, once in having seen through the illusion of my high school's teletype (which seemed to print a little, think a little, print a little more), and the other in Hawkins's deconstruction of the famous Turing test.

Like Hawkins, I have never been satisfied with Turing's formulation that intelligence is whatever fools an intelligent agent into thinking that it, too, is intelligent. At first, Turing's proposition struck me as deeply insightful, but as the years rolled by, I gradually came to regard it as banal and useless, more of a copout than a useful definition.

Hawkins recounts John Searle's "Chinese Room" thought experiment, which places an intelligent agent in a room, where he mindlessly follows rote instructions that result in the correct answers to a set of questions (written in Chinese, a language the agent doesn't understand) being issued outside the room.

Searle pointed out that the Chinese recipient of the output would conclude that the agent in the room must have understood Chinese, possibly even exhibiting considerable insight. But the agent himself was doing nothing that a machine couldn't do—he was simply following instructions written by someone else.

Searle said he didn't know what intelligence was, but this experiment showed that computers didn't have it. Searle's thought experiment clearly showed—to me, anyway—that Turing's test is easily fooled. There must be more to intelligence than some subjective judgment.

Others have argued that somehow the combination of agent, instructions, and room taken together exhibited intelligence—anything will do to avoid having to face the fact that we really don't know much about intelligence, despite having worked on it in several branches of science for a very long time.

I don't want to reawaken this controversy because I don't think it's very instructive in light of where Hawkins wants to take us. I mention it here mostly because it's a useful starting point in considering what intelligence is versus what its physical manifestations or outward appearances usually are.

The brain's behavior

Hawkins believes that intelligence is an emergent behavior of a large group of specialized neurons, which use a memory-based world model to make a continuous series of predictions of future events. He argues that time itself is a crucial component of what the brain does and how it does it. He believes there are three crucial aspects of a brain's behavior: The brain works on time-sequenced streams of inputs, there is a lot of feedback involved (as evidenced by the way neural nets are organized in the brain), and there is a pattern to the hierarchy of real networks that seem to be important to their function.

I particularly like Hawkins's argument that the brain is an intelligent device that is essentially independent of our limbs and our senses. He reminds us that the only way the brain can perceive or consider the outside world is through our senses, each of which sends patterns corresponding in some way to real-time measurements of the world. We think of sight, sound, and touch as being very different, but they all represent sequences of patterns to our brains. We are products of millions of years of evolution, which has so tightly integrated our brains with our senses and our motor control that we easily fall into the Turing trap of thinking that "if it looks intelligent, it is intelligent."

Hawkins points out the example of Helen Keller, who learned language and became an excellent writer despite being both blind and deaf. Yes, our brains are wired to help us make sense of the high-bandwidth information arriving via our optical and audible transducers, but intelligence clearly can still be much in evidence even when thoses senses aren't present.

Hawkins suggests viewing the neural nets of the neocortex as a distributed memory of pattern sequences, accessed associatively, stored in an invariant form, and arranged as a hierarchy.

That model makes great sense to me. I think it helps explain why many bright people think in terms of analogies: They see deeper patterns in seemingly disparate things, and when they unconsciously do the associative search, many surprising correlations surface. As Hawkins predicts, these people may be storing a single invariant representation of a concept and then adapting it as necessary to cover multiple ideas that might strike anyone else as a brilliant leap of intuition. I've often felt that there are only a handful of ideas that are truly unique to computer science, for instance; most of what we learn is merely new names for concepts we already have stored in our heads, whether we make the connection or not.

NEURAL NET STORAGE

As I read this book, I found myself nodding in agreement with many observations that Hawkins makes. For example, to help argue his case that neural nets store sequential patterns, he suggests that readers think about something that happened to them and imagine telling the story to a friend.

This exercise reveals that the story must be told in sequence, beginning to end. You can't start in the middle, and you can't tell the story backwards. Each phase "wakes up" the next phase. This is also true for musicians, who can play a long complex piece from memory, but if you ask them to start in the middle they struggle.

Most of what we learn is merely new names for concepts we already have stored in our heads.

I suspect this same phenomenon is responsible for the way most people remember the titles to songs: They mentally sing the song until the line with the title in it scrolls by their mental simulation. Again, there's that sequential aspect to storage that Hawkins is highlighting.

Hawkins also points out that whether we recite a memorized story, or type it, or write it, we're pulling it from an invariant form of storage. That information is then coordinated with other brain functions to make our hands or mouth move appropriately, speaking for recitations, making hand and arm motions for writing, and so on.

According to Hawkins, intentionality—one of the hallmarks of intelligence—seems to fit well with a model that proposes that the basic function of all neural nets is to make continuous predictions and then correct the stored world view coming back from the senses.

WRITING A WINNING STRATEGY

I once took a course on list processing in which the course project was to write a program that played tic-tac-toe. I spent an hour thinking about how to identify and codify the best strategies and tactics for that game, but it felt like I was taking the hard way. Then I realized that the game is so simple, I could have it "learn" as it went.

Essentially, I wrote the program in such a way that it would always take the "most adjacent" space in response to any move. That may seem like a poor strategy if winning is your aim, and if that is all the program did, it would have lost every game. But this was a course in list processing, so I had the program keep lists of all moves either side made for all games that it lost. After identifying the simplest adjacent move, the program would check all lists of games it had previously lost and refuse to make the same move again.

In effect, you could beat this program at tic-tac-toe, but never in the exact same way twice. Because tic-tac-toe is such a simple game, it only took 10 or 20 games for the program to become unbeatable.

Halfway through that project, it became clear to me that not only was this program not intelligent, no program even remotely like it would ever *be* intelligent. The computer was not really playing tic-tac-toe—I was. It was simply following its programming, which it had to do because it was not broken.

INTENTIONALITY

What was lacking, I felt, was any *intention* on the machine's part to accomplish something. So I was particularly enthused by Hawkins' emphasis on intention as the hallmark of neural activity.

Think of this another way. Imagine a human who was hired to guard the door of a gymnasium during a school event. The guard's job is to keep outsiders out and insiders in. But there is a context here that a human will understand without being told: The

guard's real job is to protect the people in the gym, and the building itself, while not interfering unnecessarily with what the people inside are doing.

If something unexpected happens, such as the fire department showing up, a human guard will defer to that authority immediately, no deep cognition required. Now, imagine a computer guarding that same door—it will reliably do its duty during nominal conditions, well past the point where a bored human falls asleep. But change the conditions, such as a real fire breaking out, or a police officer requesting entry unexpectedly, and the computer is likely to do the wrong thing. It has no real understanding of the context and no intention of fulfilling the unstated mission of the job. Intention matters a great deal.

SOMETHING'S MISSING

After carefully building his case, anticipating objections, and offering quick thought models to reel in the reader, about 80 percent of the way through this book, Hawkins changes directions. He takes up the topic of what an intelligent machine would look like, the obstacles to building these machines, and what they would be good for.

A famous cartoon by Sydney Harris pictures two scientists standing at a chalkboard with two distinct groups of equations, separated by the words "and then a miracle occurs." The caption reads "I think you have to be more explicit in step 2." I feel like I quantum-tunneled through this part of the book, and a chapter is missing. In contrast to his measured, steady, relentless marshalling of facts and careful explication of his theories, the last part of the book feel like a wild ride through someone's fantasies.

To be fair, I don't think Hawkins should have to solve every last problem before publishing a book, and it could well be that the challenges he has taken on dwarf the problems that may be lying in wait in "step 2"—reducing ideas on neural net neocortex organization to silicon-based machinery.

Maybe I'm just responding to my machine intelligence neurons that Hawkins has reawakened. I *want* Hawkins to be right. I think a future in which machinery stops being so obdurately stupid and starts working with us would be enthralling. Would personalities emerge? Would continuous evolution be possible or even avoidable?

Even asking these questions seems like science fiction, but reading Hawkins's book confers the right to dream these dreams again. Go feed your neurons. ■

Bob Colwell *was Intel's chief IA32 architect through the Pentium II, III, and 4 microprocessors. He is now an independent consultant. Contact him at bob.colwell@comcast.net.*