

From: Ari Gesher <[REDACTED]>
To: Joscha Bach <[REDACTED]>
Cc: Greg Borenstein <[REDACTED]>, Sebastian Seung <[REDACTED]>, Joi Ito <[REDACTED]>, takashi ikegami <ikeg@saclan.e.u-tokyo.ac.jp>, Kevin Slavin <[REDACTED]>, Martin Nowak <[REDACTED]>, Jeffrey Epstein <[REDACTED]>
Subject: Re: MDF
Date: Mon, 28 Oct 2013 02:55:31 +0000

> On Oct 26, 2013, at 18:47, Joscha Bach <[REDACTED]> wrote:

>

>> Am 24.10.2013 um 02:56 schrieb Greg Borenstein <[REDACTED]>:

>>

>> I think this issue of the changing definition of intelligence being a moving goal post is absolutely critical, Joscha. And it's one that long-predates 20th century digital computation-based AI efforts.

>>

>> Recently, I've been reading the work of Jessica Riskin, a Stanford historian who studies the long history of AI and Artificial Life. Specifically, Riskin's been writing about a strange phase in the history of mechanical automatons that happened in the second half of the 18th century. Previously, automatons had always been built with their mechanism in one place (i.e. in a hidden box or platform) that then drove their figures via a series of rods or connectors. The figures, the representative part of the automaton were like the birds in a cuckoo clock with no relation to the mechanism that made them move.

>>

>> Then, suddenly, in the second half of the 18th century, a series of automaton makers started to produce automatons that were built in a way that was analogous to the thing they represented. (...)

>

> Yes! This represents a shift from a depiction to some kind of "essentialist" mechanist world view. This essentialist thinking is still alive today, even if the notion of the essence has changed. Searle believes that thinking is impossible without the "intrinsic properties of biological neurons", Penrose wants the intrinsic properties of quantum systems, and a lot of people seem to believe that minds are chemical or biological or environment-interaction phenomena. The common denominator of these positions is that we need to recreate the actual makeup to produce the property (intelligence, mind, cognition).

> At the moment, the most fruitful position is probably functionalism, i.e. the idea that the property is the result of the underlying functionality (with respect to that property), and that there is no essence. If we would replace all biological neurons with little machines that perform exactly the same functions and interact with the rest of the body in the same way, the resulting mind would be unchanged.

> Contemporary functionalism, however, is mostly still built on a the foundation of mechanism, in the form of materialist physicalism. We are asking ourselves: what kind of mechanism is the mind? What are the mechanics of neurons?

While I think the notion of functionalism stands as a thought experiment, building equivalent systems that "perform exactly the same function" as the original is pretty elusive. I remember hearing from a researcher at UW trying to build a mechanical finger to study human movement for cybernetic purposes. The first prototype they had incorporated four degrees of freedom; that didn't work so they took a closer look at the finger. They settled on 39 degrees of freedom as the correct model. (sorry if [REDACTED] sounding petulant here, [REDACTED] just always suspect of this sort of reduction - while not really offering an alternative or staking a flag for essentialism, per se; usually, [REDACTED] good at making muddy waters even more muddy).

It feels like the scruffy vs. neat tension again (out of curiosity, where you do place yourself on that spectrum, Joscha?). The value I see in the essentialist approach is that natural, evolved systems use all the subtlety, the complexity of their medium. The neat approach keeps trying to add complexity until epsilon hits zero on its

integral. But isn't a zero epsilon actually asymptotical outside of the clean confines of math? Or at least elusive (in this arena) until we actually understand what level of physical reality that neurons are essentially operating in?

> Epistemologically, I think that matter and mechanics cannot be primary. All we have is information (discernible differences at our systemic boundaries), and our ideas of matter, or causal mechanics, are encodings over these patterns of information. Information is primary, the universe is a pattern generator, minds are a class of systems that (in very particular and distinct ways) identifies and manipulates structure in that information. "Computation" means, simply put: meddling with information. Minds, in this computationalist perspective, are certain types of information processing systems. (Ones that implement functions for integrating, representing, interpreting, combining, anticipating... information, and that have a motivational system to set goals that direct these functions.)

Strong agreement with all of this. All of science and human understanding is a best-fit fiction that matches the narrative of the sense stream that is our only proxy for that thing we assume exists: some sort of shared, objective reality. Or as I like to put it: there is no such thing as an electron but it sure makes a good story.

>
> Most computer scientists are computationalists by instinct: to us, everything is a computer program in some sense. (Physics, for instance, is the endeavor to find a possible implementation that could produce all known observable phenomena.) Most other people on the planet, including quite a few philosophers, are not. To them, the idea of "reducing" mind and universe to regular and stochastic changes in patterns of information (aka computation) might even sound offensive.

Hah. I guess I never thought of just how weird that makes us to the rest of the world, but yes. Without the supernatural (which seems to lacking in any sort of proof), any other conclusion is absurd.

So let's take it way out there: the bedeviling factor here might be how much of the dynamics that make up mind reside outside of the brain or even outside of the body. The favorite fictional device in the make-believe that is scientific understanding is that of the system boundary. A very useful approximation, to be sure, but we've already seen the idea of rigorous differentiation and sub-system boundaries in the brain evaporate as we learn more about how it works. And at the physical particle level of approximation there is no real difference between inside our skin and outside; the coherency of our bodies as systems is true but that's really just a line of credit against the eventual entropy that eats us all - we expend huge amounts of energy to maintain the pattern that is our corporeal self (thus hastening heat death).

Which, if I may be permitted an aside, seems to really be the definition of life: systems trading temporal and spatial local stability and order at the expense of increased overall entropy. I guess some patterns are more equal than others.

So while I believe that a functionalist rebuilding is possible, I think we underestimate just how entwined we are in our environment. The logical extreme is that you couldn't perfectly simulate a human mind without including the rest of the universe. The open question, I guess, is just where on the spectrum between large single-all-encompassing system and small, closed, minimal complexity does mind lie.

■ love to hear your thoughts on that.

>
> By the way, the first AI optimist was probably LaMettrie. His small, witty and much maligned book "L'homme machine" (1747) is full of modern insights, such as the continuum between humans and great apes, the futility of the scholastic method (building on authority instead of experiment), the nonsense of dualism, the idea that machines need not be physical but can be mathematical, and so on. When he predicts that Vaucanson's automatons herald the imminent arrival of machines that will actually speak and understand, he sounds almost like Ray Kurzweil ;-)

Has anyone voiced the worry that building AGI might make us aware of larger structures in the universe that have the right level of connectionism, dynamism, and complexity to also support emergent minds? That it might lead us to to god (in an areligious sense)? We already have conjecture around the internet itself, the Gaia hypothesis before that.

Another question: where does the AGI crowd sit on the question of animal cognition? What is the lowest high creature?

>> This strikes me as very much like the process we go through with the defining AI tasks like chess or Jeopardy or car driving. We start off believing that these are tasks that only human intelligence can achieve. Then we build computational systems that can do them. Those systems are often inspired by the way humans achieve the tasks, but in the end work in extremely non-human ways. Google's self-driving car uses massive satellite data and laser scanning to drive, Deep Blue doesn't play chess like a human does.

>

> Applied AI has the big benefit that it does not offend anyone. The applications you mentioned are usually straightforward engineering, i.e. they do not even attempt to mimic human intelligence, but only look for ways to solve the task at hand in the best possible way.

> If we want these applications to teach us something about the mind, we need to impose additional constraints. For instance, in robotic soccer, we tended to brute-force problems with sensing the robot's environment with hardware (laser scanners, sonar, additional cameras, better CPUs). The exception was the Aibo league, which had to make do with Sony's dog robots. With only a terrible, shaky camera, little memory and a slow CPU, the programmers had to come up with attentional processing, anytime algorithms that deliver better results over time, resource allocation etc. I found that the most constrained robots enforced the most interesting solutions (of course, without getting much credit for it).

That's certainly my experience in surveying industry practices. The fitness functions on these things tend to very practical and it's more about problem-solving as fast, cheaply, and safely (from a completion standpoint) as possible. New ways of doing things are invented but only by necessity.

> Another problem with the typical applications is that they usually replace or reproduce human performance, i.e. they need to start out with the abilities of a trained adult. Instead, we might want to look at reproducing the path that human children take towards intelligence, the autonomous process by which children learn to make sense of the world, acquire a language, visual and conceptual grammars and so on. After about 3.5 years of swimming in massive flows of data, a child won't do well at Jeopardy, traffic navigation or chess yet, but can already watch the first Star Wars movie and afterwards explain that Darth Vader destroyed princess Leia's planet, and that he needs a laser sword. All cognitive development afterwards is probably trivial ;-) I think this is the kind of performance we should be looking for when we try to build AI.

Yeah, I heard this point echoed by the cybernetics researchers I mentioned above and I think it's an important one. Learning to tie your shoes takes something like 250,000 hours of training (four years) for the brain to learn. That was something I noticed in the unsupervised learning paper. I was sad that they used so little data and didn't let it run longer. With the results they got, I would think that a much larger scale test could yield even better results.

Cheers,

Ari