

---

**From:** Joscha Bach <[REDACTED]>  
**Sent:** Sunday, October 27, 2013 1:48 AM  
**To:** Greg Borenstein  
**Cc:** Sebastian Seung; Joi Ito; takashi ikegami; Ari Gesher; Kevin Slavin; Martin Nowak; Jeffrey Epstein  
**Subject:** Re: MDF

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 could 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?

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.)

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" =ind and universe to regular and stochastic changes in patterns of =nformation (aka computation) might even sound offensive.

By the way, the first AI optimist was probably LaMettrie. His small, =itty and much maligned book "L'homme machine" (1747) is full of =odern insights, such as the continuum between humans and great apes, =he futility of the scholastic method (building on authority instead of =xperiment), the nonsense of dualism, the idea that machines need not be =hysical but can be mathematical, and so on. When he predicts that =aucanson's automatons herald the imminent arrival of machines that =ill actually speak and understand, he sounds almost like Ray Kurzweil =-) LaMettrie had relatively little impact in his time; his open atheism and =ejection of souls, vitalism, etc. made him a persona non grata. Perhaps not entirely unlike Minsky among continental philosophers of =ind... And the impact of the Lighthill report has probably a lot to do =ith the offense that people take at the notion of mind as machine.

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

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

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

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

Cheers,

Joscha

```
<?xml version="0" encoding="TF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN" "http://www.apple.com/DTDs/PropertyList-1.0.dtd">
<plist version="0">
<dict>
  <key>date-last-viewed</key>
  <integer>0</integer>
  <key>date-received</key>
  <integer>1382838458</integer>
  <key>flags</key>
  <integer>8623750161</integer>
</dict>
```

```
<key>gmail-label-ids</key>
<array>
  <integer>22</integer>
  <integer>2</integer>
</array>
<key>remote-id</key>
<string>355708</string>
</dict>
</plist>
```