From: Joscha Bach <[redacted]
Sent: Friday, November 1, 2013 6:31 AM
To: Ari Gesher
Cc: Greg Borenstein; Sebastian Seung; Joi Ito; takashi ikegami; Kevin Slavin; Martin Nowak; Jeffrey Epstein
Subject: Re: MDF

Ari,

sorry for the slight delay until I got around to answering your latest =ail. The discussion is very interesting and inspiring to me! With =espect to our original discussion of intelligence with respect to =ooperation, competition and deception, it is mostly tangential (all =ecipients, please be warned, we spun off towards the metaphysics of =I).

> While I think the notion of functionalism stands as a thought
> =xperiment, building equivalent systems that "perform exactly the same =unction" as the original is pretty elusive. I remember hearing from a =esearcher at UW trying to build a mechanical finger to study human =ovement for cybernetic purposes. (...) The value I see in the essentialist approach is that natural, evolved =ystems use all the subtlety, the complexity of their medium. The neat =pproach keeps trying to add complexity until epsilon hits zero on its =ntegral. But isn't a zero epsilon actually asymptotical outside of the =lean confines of math? Or at least elusive (in this arena) until we =ctually understand what level of physical reality that neurons are =ssentially operating in?

I do not think of functionalism as a recipe to build something, but as =n epistemological position. Functionalism recognizes that we construct =ur concepts (including the concept of mind and intelligence) based on =hat things do, not on what they 'essentially' are. A mind is not an =ntrinsic power animated by a soul with no empirical properties, but a =ausal arrangement that processes information in such a way that it is =ble to participate in discourse, control a body, reflect upon itself, =ind creative solutions, imagine and dream, and so on. I do not suggest =o reduce any of these properties away, but to focus on the right level.
This level is not the neural level, for instance. I suspect that look =olely at neurons would be akin to explain flight by looking at =eathers, instead of aerodynamics. Chances are that we learn things from =eathers, but we will also be intimidated by the trillions of tiny stems =hat interconnect them in intricate ways, etc., and if we replicate =hem, chances are that we end up with a penguin. In other words, I =uggest looking at what functions neurons compute, and how that =ontributes to the set of abilities that we want to replicate and/or =nderstand.

> It feels like the scruffy vs. neat tension again (out of curiosity, =here you do place yourself on that spectrum, Joscha?).

I am pragmatic. I suspect that nature is mostly scruffy (but not =ntirely so, a lot of our physiology is very clear-cut), and that there =re limits to both a completely scruffy and especially an entirely neat =pproach.

>> Most computer scientists are computationalists by instinct: to us, =verything is a computer program in some sense. (Physics, for instance, =s the endeavor to find a possible implementation that could produce all =nown observable phenomena.) Most other people on the planet, including =uite 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.
>
> Hah. I guess I never thought of just how weird that makes us to the =est of the world, but yes. Without the supernatural (which seems to =acking in any sort of proof), any other conclusion is absurd.

That is even true if we would include the supernatural. Imagine that we =ere living in a dream (i.e. that our experience of matter does not =eflect anything outside our minds, which is in some sense what the =agic or esoteric world views eventually come down to), we would still =e information processing systems that perform computations, i.e. =anipulate information, and could be described and modeled as cognitive =rchitectures.

> So let's take it way out there: the bedeviling factor here might be =ow much of the dynamics that make up mind reside outside of the brain =r even outside of the body. The favorite fictional device in the =ake-believe that is scientific understanding is that of the system =oundary. A very useful approximation, to be sure, but we've already =een the idea of rigorous differentiation and sub-system boundaries in =he brain evaporate as we learn more about how it works.

I would like to invite you to entertain for a moment the idea of giving =p on the essentialism here. The system boundary is merely a conceptual =hoice, which then determines the functional properties of the resulting =ystem. System boundaries are part of the map, not the territory. For =nstance, Andy Clark suggests that we should add tools (such as cars and =otebooks) and even a slice of the environment to our concept of mind. I =hink that this idea of the 'extended mind', as he calls it, makes a =ot of sense, but it won't change a bit of what we do as AI =esearchers: yes, we want to build our system to be able to use tools, =nd to integrate their mental model of these tools into its =roprioception and self-model. The only difference is in what part of =he resulting functionality we call 'mind', it is mostly =erminological.

The attempt to strictly align the conceptual system boundary with the =unctionality has lead Maturana and Varela to their idea of =autopoiesis', a brilliant, intriguing and utterly poisonous notion that =as killed both cybernetics (—> second order cybernetics) and systemic =ociology (—> Luhmann) by turning them from proper sciences into =umanities.

> So while I believe that a functionalist rebuilding is possible, I =hink we underestimate just how entwined we are in our environment. The =ogical extreme is that you couldn't perfectly simulate a human mind =ithout including the rest of the universe.

Why? For instance, imagine a brain that is connected a birth to a =omplex game in with a physics simulation, and learns to interact with =hat environment. Why would the resulting system not qualify as a mind? =ecause it fails to simulate a particular human mind? (The latter does =ot strike me as the interesting task here, just as understanding flight =oes not amount to the exact simulation of a hummingbird.)

> The open question, I guess, is just where on the spectrum between =arge single-all-encompassing system and small, closed, minimal =omplexity does mind lie.
>
> I'd love to hear your thoughts on that.

I am sympathetic to Turing's original idea, to conceptualize minds as =ystems capable (at least) of intelligent, meaningful discourse. In my =iew, this necessitates a certain kind of generality of concept =cquisition and control that includes the equivalent of autonomous =erception, motivated action, associative and syllogistic reasoning and =o on. While I am not convinced that embodiment is a necessary condition =or having a mind, I think that a mind must be able to make use of a =ody when given one (i.e. a general AI architecture must be able to =ddress embodiment).

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

That idea is obvious and inevitable. I like the idea of framing the =eligious perspective as the assumption that the universe, or a =eta-structure beyond human organizations is intentional, self-aware and =artial towards our personal existence and toil. (But why should it.) On =he other hand, we can conceive of micro-minds, sub-structures of human =r animal minds that themselves present the functional properties that =efine mind-ness. In each case, however, we will have to make these =roperties explicit when we ask the question: mind as an atomic, =ssential concept is pretty useless here, and we will have to ask =urselves, whether a large or small structure in the universe is capable =f intentional action, concept formation, creative problem solving, =eflection, self-awareness and so on. That case is probably extremely =ard to make for any empirically given arrangement of stuff that is not = large brain, a complex (social or economic) organization or a suitably =esigned computer, even if it might suit our spiritual needs or =an-psychic intuitions.

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

I doubt that there is a well-defined consensus, but most people I know =n the field would probably agree that all animals with sufficiently =omplex brains (including mammals, birds, cephalopods) are cognitive =gents. For instance, no-one publicly objects when Aaron Sloman rejoices =bout the smarts of Betty the crow.
On the other hand, very few species are capable of mastering linguistic =nd visual grammars to some interesting degree. Even among humans, there =re classes of problems that cannot be solved by all people, e.g., it =eems that not all people (with normal intelligence) can be taught how =o program. Even people are not generally intelligent, in the strict =ense.

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

To me, it seemed that the Ng/Google experiment got to the level of =bject discrimination of perhaps a six month child, with the equivalent =f about ten years of visual input. That is quite good, considered that = toddler can dramatically improve its classification by actively =esting hypotheses (which the Youtube frame processor was not allowed). = somehow doubt that the system would get dramatically better than =emonstrated while staying at a crude and passive model of the visual =ortex. Beyond that, you'd want goal directed and social concept =ormation, some basic reasoning capability and so on.

Bests,

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>conversation-id</key>
	<integer>270440</integer>
	<key>date-last-viewed</key>
	<integer>0</integer>
	<key>date-received</key>
	<integer>1383287474</integer>
	<key>flags</key>
	<integer>8590195713</integer>
	<key>gmail-label-ids</key>
	<array>
		<integer>6</integer>
		<integer>2</integer>
	</array>
	<key>remote-id</key>
	<string>357709</string>
</dict>
```

</plist>