From: Joscha Bach [redacted]
Sent: Sunday, July 10, 2016 4:43 AM
To: Jeffrey Epstein
Cc: Martin; moshe hoffman
Subject: Mechanisms for learning

Dear Jeffrey,

thank you for your support and encouragement, even where I fail.
Sorry for being such an embarrassment today. I will spell out today's =rgument a bit better and cohesive when I get to it. Also, I should have =ecognized that the main point I tried to make would trigger Noam (who =as as always very generous, patient, kind and humble on the personal =evel, even though he did not feel like conceding anything on the =onceptual one). Almost all of Noam's work focused on the idea that =umans have very specific circuits or modules (even when most people in =is field began to have other ideas), and his frustration is that it is =o hard to find or explain them.

I found Noam's hypothesis very compelling in the past. I still think =hat the idea that language is somehow a cultural or social invention of =ur species is wrong. But I think that there is a chance (we don't know =hat, but it seems to most promising hypothesis IMHO) that the =ifference between humans and apes is not a very intricate special =ircuit, but genetically simple developmental switches. The =ootstrapping of cognition works layer by layer during the first 20 =ears of our life. Each layer takes between a few months and a few years =o train in humans. While a layer is learned, there is not much going on =n the higher layers yet, and after the low level learning is finished, =t does not change very much. This leads to the characteristic bursts in =hild development, that have famously been described by Piaget.

The first few layers are simple perceptual stuff, the last ones learn =ocial structure and self-in-society. The switching works with something =ike a genetic clock, very slowly in humans, but much more quickly in =ther apes, and very fast in small mammals. As a result, human children =ake nine months before their brains are mature enough to crawl, and =ore than a year before they can walk. Many African populations are =uite a bit faster. In the US, black children outperform white children =n motor development, even in very poor and socially disadvantaged =ouseholds, but they lag behind (and never catch up) in cognitive =evelopment even after controlling for family income.

Gorillas can crawl after 2 months, and build their own nests after 2.5 =ears. They leave their mothers at 3-4 years. Human children are pretty =uch useless during the first 10-12 years, but during each phase, their =rains have the opportunity to encounter many times as much training =ata as a gorilla brain. Humans are literally smarter on every level, =nd because the abilities of the higher levels depend on those of the =ower levels, they can perform abstractions that mature gorillas will =ever learn, no matter how much we try to train them.

The second set of mechanisms is in the motivational system. Motivation =ells the brain what to pay attention to, by giving reward and =unishment. If a brain does not get much reward for solving puzzles, the =ndividual will find mathematics very boring and won't learn much of it. =f a brain gets lots of rewards for discovering other people's =ntentions, it will learn a lot of social cognition.

Language might be the result of three things that are different in =umans:
- extended training periods per layer (after the respective layer is =one, it is difficult to learn a new set of phonemes or the first =anguage)
- more layers

- different internal rewards. Perhaps the reward for learning =rammatical structure is the same that makes us like music. Our brains =ay enjoy learning compositional regular structure, and they enjoy =aking themselves understood, and everything else is something the =niversal cortical learning figures out on its own.

This is a hypothesis that is shared by a growing number of people these =ays. In humans, it is reflected for instance by the fact that races =ith faster motor development have lower IQ. (In individuals of the same =roup, slower development often indicates defects, of course.)

Another support comes from machine learning: we find that the same =earning functions can learn visual and auditory pattern recognition, =nd even end-to-end-learning. Google has built automatic image =ecognition into their current photo app:
=ttp://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people=as-gorillas-showing-limits-of-algorithms/

The state of the art in research can do better than that: it can begin =o "imagine" things. I.e. when the experimenter asks the system to =dream" what a certain object looks like, the system can produce a =omewhat compelling image, which indicates that it is indeed learning =isual structure. This stuff is something nobody could do a few months =go:
=ttp://www.creativeai.net/posts/Mv4WG6rdzAerZF7ch/synthesizing-preferred-i=puts-via-deep-generator-networks

A machine learning program that can learn how to play an Atari game =ithout any human supervision or hand-crafted engineering (the feat that =ave DeepMind 500M from Google) now just takes about 130 lines of Python =ode.

These models do not have interesting motivational systems, and a =elatively simple architecture. They currently seem to mimic some of the =tuff that goes on in the first few layers of the cortex. They learn =bject features, visual styles, lighting and rotation in 3d, and simple =ction policies. Almost everything else is missing. But there is a lot =f enthusiasm that the field might be on the right track, and that we =an learn motor simulations and intuitive physics soon. (The majority of =he people in AI do not work on this, however. They try to improve the =erformance for the current benchmarks.)

Noam's criticism of machine translation mostly applies to the Latent =emantic Analysis models that Google and others have been using for many =ears. These models map linguistic symbols to concepts, and relate =oncepts to each other, but they do not relate the concepts to "proper" =ental representations of what objects and processes look like and how =hey interact. Concepts are probably one of the top layers of the =earning hierarchy, i.e. they are acquired *after* we learn to simulate = mental world, not before. Classical linguists ignored the simulation =f a mental world entirely.
It seems miraculous that purely conceptual machine translation works at =ll, but that is because concepts are shared between speakers, so the =tructure of the conceptual space can be inferred from the statistics of =anguage use. But the statistics of language use have too little =nformation to infer what objects look like and how they interact.

My own original ideas concern a few parts of the emerging understanding =f what the brain does. The "request-confirmation networks" that I have =ntroduced at a NIPS workshop in last the December are an attempt at =odeling how the higher layers might self-organize into cognitive =rograms.

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>conversation-id</key>
	<integer>75449</integer>

```
        <key>date-last-viewed</key>
        <integer>0</integer>
        <key>date-received</key>
        <integer>1468125782</integer>
        <key>flags</key>
        <integer>8590195717</integer>
        <key>gmail-label-ids</key>
        <array>
                <integer>6</integer>
                <integer>2</integer>
        </array>
        <key>remote-id</key>
        <string>626407</string>
</dict>
</plist>
```