
From: Joscha Bach <[REDACTED]>
Sent: Wednesday, August 26, 2015 7:45 PM
To: Jeffrey Epstein
Subject: Decision making

Re motivation: Have you seen the recent movie "Ex Machina" = I liked it; one of the few AI movies that have not been dumbed down. =he main character is a beautiful female looking AI, clearly =ntelligent, and able to manipulate humans to an arbitrary degree. What =akes her inhuman is that she is probably motivated by a single =inciple, like option maximization. That would make her an inscrutable =sychopath. Option maximization would entail energy, physical integrity, =erhaps reproduction, certainly learning, but you will have an agent =hat you won't like to share a prisoner's dilemma with. = How do you approach decision making? I have recently learned that people =hat subscribe for Cryonics (freezing one's head or body in the =ope to be revived when future technologies make it possible) assign a =ower probability to that it works than the general population. But =hereas "normal" people tend to make binary models about =he world: something is "probably not going to work, so let us =ot bother", many of the Cryonics folks will argue that paying =500 a month for a 1% chance of immortality seems like a bargain.

This seems to generalize: in principle, we should perform a Bayesian =pproximation for all our major decisions, attach probability =istributions to everything in the space of possible beliefs, and be =ble to outperform the vast majority of folks that relies on narratives =i.e. binary yes/no decisions about the facts in the world). Gigerenzer, =ahnemann and many others have shown that human brains are terrible of =etting this intuitively right, to the point where an absence of =ine-grained domain knowledge often leads to better management decisions =tc.

The divide between probabilistic models vs. narrative models is =eflected to some degree in the conflict between probabilistic and logic =ased AI. In practice, we will probably need to combine both, but I =onder if there is an intrinsic limit to probabilistic descriptions in a =ightly complex world, where we cannot observe baseline probabilities =nyway.

How do you decide? Do you Solomonoff-induce and Bayes the hell out of =he stock market, do you reason, do you soak up data and let your =ntuitions guide you, or is most of the important stuff depending on =ommunication and negotiation? Is there a general approach, or how =uch should theories of decision making be dependent on the domain?

Cheers,

```
Joscha=?xml version=.0" encoding=UTF-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>106419</integer>  
    <key>date-last-viewed</key>  
    <integer>0</integer>  
    <key>date-received</key>  
    <integer>1440618272</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>537356</string>
</dict>
</plist>
```