Flashback: Programming Free Will – creative robots

Björn Brembs

Blog post published August 29, 2013 in *bjoern.brembs.blog*
https://doi.org/10.59350/er44n-vab14

Abstract

*During my flyfishing vacation last year, pretty much nothing was happening on this blog. Now that I've migrated the blog to WordPress, I can actually schedule posts to appear when in fact I'm not even at the computer. I'm using this functionality to re-blog a few posts from the archives during the month of august while I'm away.

Keywords

Cognitive Neuroscience (Subfield), Blogarchives, Briegel, Free Will, Robots, Spontaneity

Recommended citation

Brembs, B. (2013, August 29). Flashback: Programming Free Will – creative robots. *Bjoern.brembs.blog*. https://doi.org/10.59350/er44n-vab14

Copyright

Copyright © Björn Brembs 2013. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*During my flyfishing vacation last year, pretty much nothing was happening on this blog. Now that I've migrated the blog to WordPress, I can actually schedule posts to appear when in fact I'm not even at the computer. I'm using this functionality to re-blog a few posts from the archives during the month of august while I'm away. This post is from September 13, 2011:
*

I wasn't planning to comment on Kerri Smith's piece on Free Will (probably paywalled) in the last issue of *Nature* magazine. However, this morning I read a paper on Free Will in robots (or rather 'agents'), which urged me to suggest some updates to the sadly (otherwise Ms. Smith is producing outstanding work, especially her podcasts!) outdated discussion in the *Nature* article.

Her article starts out with a modern variation of Libet's famous experiments. These experiments can be caricatured like this: "press a button whenever you feel like it and watch a clock while you're making the decision to tell us when you think you've made the decision". It is then little surprise that some form of brain activity (either electrical, in the case of Libet or blood flow, in the case of the modern fMRI studies) can be recorded before the time point when the study participants self-reportedly made the decision.

Detailed treatment of these experiments isn't really needed here, as any biologist realizes that all our thoughts are indeed based on brain activity and thus any conscious act or thought must be either simultaneous to or preceded by nervous activity. The amount of this time difference may vary with the task and the method of activity measurement. The fMRI brain scans allowed researchers to predict a dual choice to 60%, i.e., just above chance level. Clearly, even with modern brain scans a brain isn't even close to a system one might call 'deterministic' by any stretch of the word.

From the way I read the article, the most important point drawn by the researchers is that the thought process itself is based on brain activity. John Dylan-Haynes:

> *"I'll be very honest, I find it very difficult to deal with this," he says. "How can I call a will 'mine' if I don't even know when it occurred and what it has decided to do?"*

I'd counter that question with another question: what else then brain activity would you have expected when you peered into a brain? Dualism has been dead since Popper's and Eccles' "The Self and its Brain" in 1977. Why is this article still beating a dead horse?

About half way through the article, this exact issue is raised:

> *The trouble is, most current philosophers don't think about free will like that, says Mele. Many are materialists — believing that everything has a physical basis, and decisions and actions come from brain activity. So scientists are weighing in on a notion that philosophers consider irrelevant.*

Precisely! And yet, towards the end of the article, the dualism creeps back in, by the same philosopher who so rightly dismissed it:

> *Philosophers are willing to admit that neuroscience could one day trouble the concept of free will. Imagine a situation (philosophers like to do this) in which researchers could always predict what someone would decide from their brain activity, before the subject became aware of their decision. "If that turned out to be true, that would be a threat to free will," says Mele.*

Even if this prediction were possible, any decision would still be ours, as it would still not be possible to predict the decision from the time when the decision-task was initiated. In other words: one would need to observe the decision-making process for some time in order to eventually project where it is going to end up. I think it is very likely that we will be able to go rather far with this approach, but because our brain is still calling the shots, this has absolutely no relevance for the question on how free the decision was. We are not slaves of our brains, we *are* our brains. And this means an upgrade for our understanding of human nature, or you are vastly underestimating the abilities of brains.

But enough of the disappointing aspects of this article. I was reminded of it because of a very exciting article by a physicist in Austria, Hans J. Briegel: "On machine creativity and the notion of free will". It displayed a modern understanding of the scientific issues surrounding a materialistic (i.e., scientific) notion of free will and provided a proof of principle of how Free Will may be implemented in physical objects. And these objects don't even have to be biological in origin! As Briegel writes:

> *To put it provocatively, even if human freedom were to be an illusion, humans would still be able, in principle, to build free robots. Amusing.*

Amusing indeed! The paper by Briegel elaborates on a method to provide software agents with a degree of freedom without breaking any laws of nature, a method he calls 'projective simulation'.

Briegel claims that Free Will by projective simulation, could, "in principle, be realized with present-day technology in form of [...] robots." Projective simulation means that the robots have a flexible sort of memory that allows the agent to simulate situations that are similar, but not identical to, events that it has encountered before. There are rules according to which these 'projections' can be generated by the robot, so they're not arbitrary, but they contain a degree of randomness (or 'spontaneity') that allows them to "increasingly detach themselves from a strict causal embedding into the surrounding world". Briegel realizes that, in biological systems, much of the required random variability is readily available, but because we don't know how it is being used, we cannot say much about the relevance of it. In fact, with reference to Quantum Indeterminacy, he arrives at almost the same wording as I did in my Proc. Roy. Soc Article:

> *We may not need quantum mechanics to understand the principles of projective simulation, but we have it. And this is our safeguard that ensures true indeterminism on a molecular level, which is amplified to random noise on a higher level. Quantum randomness is truly irreducible and provides the seed for genuine spontaneity.*

It is gratifying to see how close we came of each other, without knowing of each other. Here is my way of putting it:

> *Because of this nonlinearity, it does not matter (and it is currently unknown) if the 'tiny disturbances' are objectively random as in quantum randomness or if they can be attributed to system, or thermal noise. What can be said is that principled, quantum randomness is always some part of the phenomenon, whether it is necessary or not, simply because quantum fluctuations do occur. Other than that it must be a non-zero contribution, there is currently insufficient data to quantify the contribution of such quantum randomness. In effect, such nonlinearity may be imagined as an amplification system in the brain that can either increase or decrease the variability in behaviour by exploiting small, random fluctuations as a source for generating large-scale variability.*

If this topic is of any interest to you, you really ought to read Briegel's paper!

Basically, the discussion about freedom today has progressed beyond the question of whether it exists (the dualistic notion, everyone agrees, does not), but how it has been implemented in a material world that is powerful and creative enough to not need any supernatural forces. It is sad that this was only briefly touched upon in the *Nature* piece, when it should have been the very core of the article.

Smith, K. (2011). Neuroscience vs philosophy: Taking aim at free will Nature, 477 (7362), 23-25 DOI: 10.1038/477023a