

## Achieving substrate-independent minds: no, we cannot ‘copy’ brains

August 24, 2011 by Randal A. Koene

On August 18, IBM published an intriguing update of their work in the DARPA SyNAPSE program, seeking to create efficient new computing hardware that is inspired by the architecture of neurons and neuronal networks in the brain.

At carboncopies.org, we strive to take this research a step further: to bring about and nurture projects that are crucial to achieving substrate-independent minds (SIM). That is, enable minds to operate on *many* different hardware platforms — not just a neural substrate. And we seek realistic routes to SIM.

But what is it that those projects should accomplish?

When you transfer a mind from its biological substrate to another sufficiently powerful computing architecture, that mind has become substrate-independent. In that case, complete access is gained in a way that enables you to transfer its relevant data. Such minds still depend on a substrate to run, but you can operate on a variety of different substrates.

### Life expansion

What do you do with a substrate-independent mind? Backing up, copying and restoring minds can be aspects of a robust route to life extension. But SIM researchers are particularly interested in enhancement, competitiveness and adaptability.

We can think of this as “life expansion.”

This has been described to some degree in my article, Pattern survival versus Gene survival. Imagine a mind that can think many times faster than we do now, and can access knowledge databases such as the Internet as intimately as we access our memories now.

In addition to minds that are copies of a human mind, we are interested in man-machine merger, or rather in the ability of man to keep pace with machine and share the future together. Other-than-human substrate-independent minds are therefore necessarily also topics of SIM. In this context, it is worth noting that a SIM is a type of artificial general intelligence (AGI). But an AGI is also a type of SIM, since the AGI should be able to take advantage of any sufficiently powerful computing platform. The two areas of research are closely related.

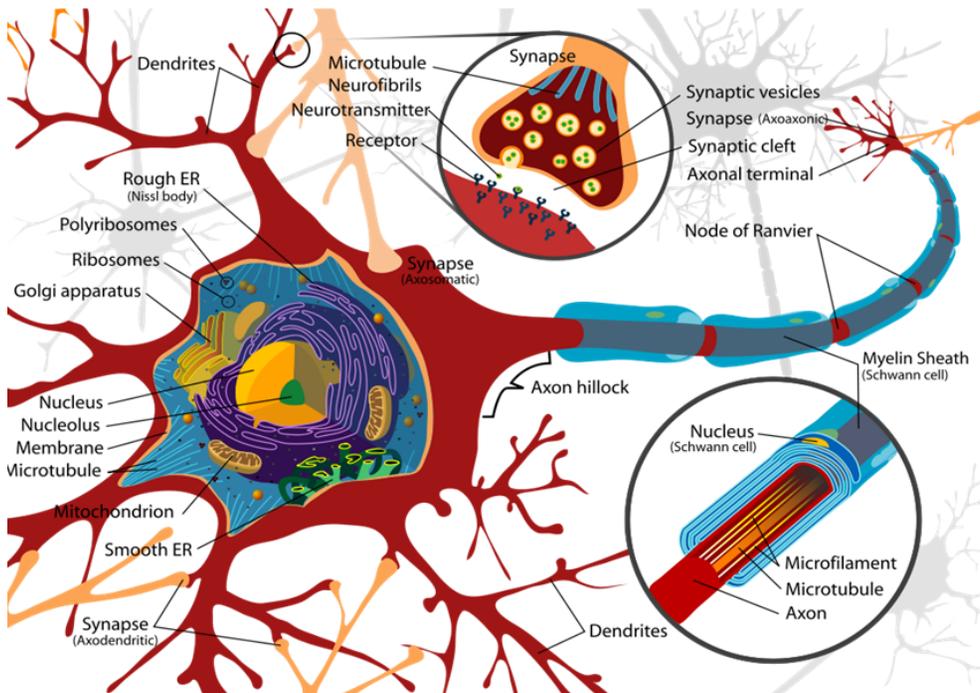
## How to create a SIM

So there are at least two ways to successfully create a substrate-independent mind: implementing a mind in a synthetic neural system or transferring the existing characteristic information — and the functions that process that information — from a specific mind into an implementation that operates in another substrate.

Are both of those legitimate goals of SIM?

A synthetic neural system and the transfer of information about an existing neural system to a synthetic implementation do share many technological requirements, but there are some significant differences that appear when you consider the possible scope of each of those aims. For example, how complex must the neurons of a synthetic neural system be? A simple spiking neuron is a computational element that integrates weighted input received through numerous synaptic connections and delivers a spike of activity when that integral reaches a threshold.

We can certainly imagine that a synthetic neural system composed of simple spiking neurons could carry out useful work as a flexible and powerful neuromorphic computer.



Neuron (credit: Wikipedia user LadyofHats, public domain)

But how does that compare with a human brain? The neurons of the human brain, or of any biological brain, are not simple spiking neurons. As we zoom in and examine the composition of a biological neuron, as we take note of the intricacy of its biological machinery, we learn that each individual neuron is in fact an incredibly complex biological entity. To capture all of its behavior might require analysis and simulation down to the molecular level, or even to the subatomic level if we care about every possible environmental effect that interaction with the neuron can have.

Does that matter? That depends on your goals. We must concede that at some level, somewhere between the spiking of the neurons and their subatomic behavior, cumulative non-linear behavior may lead to results that differ significantly from those of a simple spiking neuron. This is certain, as it has proven to be quite hard for computational neuroscientists to develop precise models of single neurons that produce spikes at exactly the same times as a biological neuron would, given the same input<sup>1</sup>. And spike timing is crucial for computation in many operations of the biological brain.

### **Mind uploading**

If we care about more than just the train of spikes generated by a neuron, the challenge becomes ever so much greater. If the goal is to create a synthetic neural system that can do useful things, those intricacies may not matter. But what if the goal is to take one person's specific human mind and move it to another substrate in such a way that the experience of the mind's thought processes and sensations are not disturbed? What if we want to consider a transfer to SIM as a means of continuing a human being's existence? That is the putative achievement often called "mind uploading."

Could we create a synthetic brain that is not identical with the biological brain on the molecular or subatomic level, but is functionally identical at every level at which it could interact with its environment, interact within itself, and through time?

The answer is no. That is simply not possible.

At some level, no matter how precise the emulation in another substrate, there is a divergence. In the parlance of information theory, this is a Kullback-Leibler divergence (or KL divergence). KL divergence measures the divergence, expressed as additional bits required to code samples from a probability distribution  $P$  when using a code based on a probability distribution  $Q$ . The most optimal encoding of samples from  $Q$  will not be optimal for samples from  $P$  if  $P \neq Q$ .

In plain English, KL divergence evaluates the overhead of an emulation. Physically, what does this overhead mean? It means that we pay a penalty for implementing one computing method in another method. We cannot, using exactly the same space and time, carry out the same computational processes as in the original biological brain and produce the exact same effects, both internal and external interactions. We

cannot do better than the physical elements that are carrying out their own natural processes (or computations, if you like). E.g., a particle's spin is the optimal expression of that characteristic and is not represented as efficiently in any model of that spin.

### **Fidelity trade-offs**

Fortunately for SIM, this problem is actually a straw man. Do we really care about every possible process, every possible interaction? By analogy, when we want to run Macintosh programs on a PC, do we actually care about the precise patterns in space and time in which the Mac computer architecture is heating up its environment? We usually do not care about such things. We just want the programs to run and to produce the expected results.

We can emulate the Mac on a PC and run Mac programs, even if the underlying architectures are different. We may even be able to emulate one architecture on another and thereby improve the performance of the programs we wish to run!

Similarly, SIM is not about crafting perfect copies of brains or copying everything about the way they work in their environment. Since we already have the original biological implementation that interacts and decays exactly as it does, what would be the point?

SIM certainly includes the goal of creating a synthetic neural system. It is both about creating something that can perform as well or outperform the original system in the ways that we care about, and about creating a process that, when desired, can provide for a faithful transition from one system to another by emulation.

It is possible to select abstraction levels within a functional architecture and to create alternate implementations of the functions at that level. If this were not the case, the entire field of artificial intelligence (AI) research would have no hope of achieving human-level or better performance on tasks that human brains can carry out. We already know that AI systems can match and exceed human performance in a variety of tasks.

When we speak of SIM as a combination of a process (“uploading”) and of an objective (to achieve a “substrate-independent mind”), it is really about collecting the parameter values at the chosen abstraction level and re-implementing the dynamics with those parameter values at that abstraction level in a desired target platform. That is SIM.

There certainly is some relationship between such a process and means of life-extension that the notion of “copying the brain” evokes, a transition from human to post-human existence. Unfortunately, most discussions that focus on this aspect of the endeavor are relatively vague and unclear. In contrast, the ideas behind SIM are actually quite crisp and clear.

Note that it is also possible to devise an uploading procedure that provides the experience of an unbroken transition, even if it is a transition to something that is at some implementation level different from a prior existence based on a biological brain. For example, it is not

necessary for the process to be perceived as abrupt, strange, or even uncomfortable. Avoiding such experiences is a matter of process implementation. The mind, at least at the level chosen for re-implementation (and further work), can be quite good at making everything it experiences seem perfectly sensible. We do that all the time (and sometimes we even confabulate reality).

So, despite objections about the differences between biological and other hardware — and the resulting implementation of a SIM, it is quite possible that if each of your neurons and synapses were replaced one by one with something else, you might not notice.

*This article arose from a discussion with colleagues. I would like to thank both of these colleagues (you know who you are) for their useful critiques during that discussion.*

<sup>1</sup>Jolivet, R. *et al.*, (2008), “The quantitative single-neuron modeling competition,” *Biological Cybernetics*, vol. 99, pp.417–426