

THE SIMULATION ARGUMENT: SOME EXPLANATIONS

Prof Nick Bostrom
Future of Humanity Institute
Faculty of Philosophy & James Martin 21st Century School
Oxford University
(2008)

www.nickbostrom.com

[*Analysis*, in press (expected publication date July 2009)]

Anthony Brueckner, in a recent article, proffers “a new way of thinking about Bostrom’s *Simulation Argument*”.¹ His comments, however, misconstrue the argument; and some words of explanation are in order.

The simulation argument purports to show, given some plausible assumptions, that at least one of three propositions is true.² Roughly stated, these propositions are: (1) almost all civilizations at our current level of development go extinct before reaching technological maturity; (2) there is a strong convergence among technologically mature civilizations such that almost all of them lose interest in creating ancestor-simulations; (3) almost all people with our sorts of experiences live in computer simulations. I also argue (#) that *conditional on* (3) you should assign a very high credence to the proposition that you live in a computer simulation. However, *pace* Brueckner, I do not argue that we should believe that we are in simulation.³ In fact, I believe that we are probably not simulated. The simulation argument purports to show only that, as well as (#), at least one of (1) - (3) is true; but it does not tell us which one.

Brueckner also writes:

“It is worth noting that one reason why Bostrom thinks that the number of Sims [computer-generated minds with experiences similar to those typical of normal, embodied humans living in a Sim-free early 21st century world] will vastly outstrip the number of humans is that Sims ‘will run their own ancestor-simulations on powerful computers they build in their simulated universe’ (Bostrom 2003: 253). ... However, since a Sim cannot *really* build a computer—he

¹ Brueckner (2008)

² Bostrom (2003); see also Bostrom (2005).

³ Brueckner writes (p. 224): “Nick Bostrom has argued that given some plausible assumptions, we should believe that we are not humans but rather conscious computer simulations of humans (Bostrom 2003).”

merely *seems* to build one within his simulation—it follows that he cannot *really* create another human-like Sim mind that is instantiated in the programming of a *real* computer that he *really* builds.”⁴

There are several problems with this passage.

First, as already noted, the simulation argument does not purport to show (and I do not believe) that the Sims outnumber the [non-Sim] humans.

Second, Brueckner misquotes me. The phrase that he attributes to me does not occur in the paper he cites. Instead, what I wrote is that “[i]t may be possible for simulated civilizations to become posthuman. They may then run their own ancestor-simulations on powerful computers they build in their simulated universe. Such computers would be ‘virtual machines’, a concept familiar from computer science.”⁵ Notice that “*may* run” is quite different from “*will* run”! (And in fact, in the paragraph following the quoted passage, I went on to adduce one consideration against this possibility.)

Third, I doubt—even if I had believed (which I do not) that the Sims vastly outnumber the humans, *and* even if I had maintained (which I do not) that the Sims would run their own ancestor-simulations—that one of my reasons for the former would have been the latter. The computational resources that the Sims devote to creating their own ancestor-simulations ultimately must be provided by the non-Sims that are not themselves simulated. If each first-level ancestor-simulation run by the non-Sims requires more resources (because they contain within themselves additional second-level ancestor-simulations run by the Sims), the non-Sims might well respond by producing fewer first-level ancestor-simulations. Conversely, the cheaper it is for the non-Sims to run a simulation, the more simulations they may run. It is therefore unclear whether the total number of ancestor-simulations would be greater if Sims run ancestor-simulations than if they do not.

The fourth problem with Brueckner’s passage concerns a more substantial issue. He claims that:

“The idea that ‘stacked virtual machines’ can give rise to more and more conscious Sims seems to be just a confusion. ... Just as a brain in a vat is incapable of *really* building another brain in a vat, a Sim is incapable of *really* building another computer which instantiates another human-like conscious Sim mind.”⁶

Italicized “*really*”s notwithstanding, the idea of stacked virtual machines is actually quite unproblematic. And given the substrate-independence assumption, which underpins the

⁴ P. 224, italics in the original.

⁵ Bostrom (2003), p. 253.

⁶ P. 224, and fn. 1.

entire simulation argument and which Brueckner does not challenge, the idea that such stacked virtual machines, implementing suitable programs, could produce more conscious Sims does not seem to present any special difficulty.

The analogy with a hallucinating brain in a vat is misleading. When you dream or hallucinate that you are programming a computer in such a way as to create conscious Sims, of course no Sims are actually created. But if a Sim (virtually) runs a suitable program using the resources of her (computer-simulated) environment, then she might well succeed in creating new Sims. The difference is that while *your dreaming* that you are programming (or building brains-in-vats, or going to the bank) does not produce any programming (or envatted brains, or bank visits), the execution of a computer program can, and often does, emulate the execution of another computer program. In such cases the computation specified by the emulated program is (*really*) performed. Thus, when I run a Java applet on my computer, the computer runs a program that implements a virtual machine which in turn runs the applet. It is perfectly possible (albeit somewhat pointless) to create a Java applet that implements a virtual machine that runs another Java applet. This can be iterated: we can create virtual machines that implement virtual machines that implement virtual machines. Now, these virtual machines are not like dreams or hallucinations that create the false appearance of something that is not there. Instead, each computation that any of these virtual machines implements is really being implemented; and if we were to look closely at what is going on inside my computer, we would see all of these computations being instantiated by the electrons shuffling through its circuitry.

The actions of Sims have real physical consequences. If a Sim lifts her left arm, this corresponds to a different computational pattern being implemented than if she lifts her right arm. Similarly, if a Sim sits down at a (simulated) computer terminal and writes some program and implements it on her (virtual) computer, this corresponds to a different computational pattern being implemented (in the computer at the more basic level of reality that is running the simulation in which the Sim resides) than would otherwise have been the case. If the program that the Sim writes and runs is of a suitable type, it will instantiate an ancestor-simulation of its own, complete with its own Sim inhabitants. All these Sims have a real physical instantiation in the computer located at the “basement-level” of reality. The fact that some of these Sims are running on virtual machines, or on virtual machines emulating virtual machines (etc.), in no way undermines their claim to consciousness. It would be an odd form of computationalism indeed that claimed that brains can be conscious, and that computer emulations of brains can be conscious, but that computer emulations of computer emulations of brains cannot be conscious!⁷

⁷ I am grateful to Anders Sandberg, Toby Ord, and Rebecca Roache for comments.

References

Bostrom, N. 2003. Are We Living in a Computer Simulation? *Philosophical Quarterly*, Vol. 53: 243-255.

Bostrom, N. 2005. The Simulation Argument: Reply to Brian Weatherson. *Philosophical Quarterly*, 55: 90-97

Brueckner, A. 2008. The Simulation Argument again. *Analysis* 68.3: 224-226.