Ratio (new series) XXVI 3 September 2013 0034-0006

FREAK OBSERVERS AND THE SIMULATION ARGUMENT

Lyle Crawford

Abstract
The simulation hypothesis claims that the whole observable universe, including us, is a computer simulation implemented by technologically advanced beings for an unknown purpose. The simulation argument (as I reconstruct it) is an argument for this hypothesis with moderately plausible premises. I develop two lines of objection to the simulation argument. The first takes the form of a structurally similar argument for a conflicting conclusion, the claim that I am a so-called freak observer, formed spontaneously in a quantum or thermodynamic fluctuation rather than through ordinary processes of evolution and growth. The second rejects the basic line of reasoning of both arguments: the sort of evidence they cite is not capable of supporting either the claim that I am a simulant or the claim that I am a freak observer. The evidence that simulants or freak observers exist is not a reason to think that I am one of them.

The simulation hypothesis claims that the whole observable universe, including us, is a computer simulation implemented by technologically advanced beings for an unknown purpose. The simulation argument concludes that the simulation hypothesis is almost certainly true. A proponent of the argument suggests that we should believe with only modest confidence that all of its premises are true, and so should believe the simulation hypothesis with only modest confidence, e.g. 20% (Bostrom 2005).[1] The simulation argument is important not only because of its stunning, quasi-theological conclusion, but also because it is an ingenious attempt to support a conclusion about the fundamental nature of the world that is neither a traditional armchair metaphysical exercise (it draws on empirical evidence, albeit quite speculatively) nor an interpretation of fundamental physical theory. Here I press two objections to the argument; one plays by

[1] The standard version of the simulation argument is Bostrom's (2003, 2009). Its conclusion is a three-way disjunction of which the simulation hypothesis is one disjunct. What I call the simulation argument in effect includes a denial of Bostrom's two other disjuncts, a denial which he endorses with low confidence externally to his official argument. The plausibility of this denial is what warrants calling this the *simulation* argument.

the same rules, and so constitutes another example of an argument in this exciting metaphysical niche, and the other is a skeptical objection that challenges its basic reasoning.

The first objection takes the form of an argument, structurally similar to the simulation argument, in support of a conflicting conclusion, which I call the 'fluctuation hypothesis'. This hypothesis claims that I am a so-called freak observer. Freak observers are conscious systems formed spontaneously by quantum or thermodynamic processes rather than ordinary processes of evolution and growth (or, for that matter, intelligent design); most are solitary, transient, minimally ordered, and massively deluded. The 'fluctuation argument' concludes that the fluctuation hypothesis is almost certainly true. The fluctuation hypothesis is outrageous even by the standards of the simulation argument, but the fluctuation argument has defensible premises. Moreover, the details suggest that, depressingly, it pre-empts the simulation argument. The latter's line of reasoning should, if anything, lead me to suspect not that I am a simulant, but rather that I am a freak observer.

The second objection, which I develop more briefly, contends that I have no (or almost no) reason to believe either hypothesis, even with low confidence. This objection is based on, but distinct from, the complaint that the simulation argument is 'epistemically unstable' (Dainton 2003), an objection also raised against the fluctuation argument (Carroll 2010). I think that epistemic instability itself is not a decisive objection, but that it does point to a fundamental problem with these arguments, a sleight-of-hand they perform with their empirical evidence. The key evidence that they enlist is evidence that simulants or freak observers are plentiful in the universe, but 'the universe' here means *my* universe, the universe within which I am a regular human being, whether that universe is simulated or dreamed or real, so it is nonsensical to imagine that I might be a simulant or a freak observer in that universe. But that is the only universe to which this evidence pertains.

The Simulation Argument

I understand the simulation argument to contain four premises:

> P1: Consciousness is multiply realizable, and the execution of a suitable program by a suitably powerful computer would instantiate a conscious mind or minds.

P2: Out of all the occurrences of complex life throughout space and time, at least some tiny fraction become advanced civilizations that a) have the technological capacity to simulate worlds containing zillions of conscious minds, and b) exercise that capacity for scientific, entertainment, or other purposes.

P3: We lack any evidence (other than the facts supporting P1 and P2) that could help us determine whether we are simulated or 'real' (physically ordinary) observers.

P4: The indifference principle.

P1 implies that simulants are possible in principle; it is supposed to be fairly uncontroversial, and in any case is simply assumed. P2 implies (with 'zillions') that even if simulators are a minuscule fraction of all complex life, their simulants vastly outnumber physically ordinary minds. There is currently no way to reliably judge the probability of P2. Bostrom offers some back-of-the-envelope calculations supporting the computational viability of the simulation hypothesis, and Prince (2005) develops some novel technical arguments in support of what I am calling P2. We might also note humanity's own increasing enthusiasm for, and success with, large-scale neural simulation (Markham *et al.* 2011; Eliasmith *et al.* 2012). In general, given the vastness of the universe, P2 seems not unreasonable. P3 is entailed by an appropriate description of the simulation scenario. P4 says that if there are n observers in the world, and I have no basis for thinking that I am one kind of observer rather than another, then I assign a probability of $1/n$ to the possibility of being any particular observer.

If the universe has r real and s simulated observers ($n = r + s$), and P3 is true, then the probability that I am a simulated observer is $s/(r + s)$. If P2 is true ($s >> r$), then that probability approaches 1. So if P2 is the only contentious premise (although P1 will be rejected by some), then I should believe the simulation hypothesis approximately to whatever degree I believe P2.

The Fluctuation Argument

The simulation hypothesis is not the only highly surprising claim that may be supported with the pattern of reasoning just outlined. The fluctuation hypothesis holds that I am a freak observer, and the fluctuation argument that I will develop in parallel to the simulation argument concludes that the fluctuation hypothesis is

almost certainly true. Freak observers are often called 'Boltzmann brains' (e.g. Albrecht and Sorbo 2004) after Ludwig Boltzmann, who saw that the newly developed statistical mechanics predicted not only the familiar evolution of any closed system toward equilibrium, but also the occasional reversal of this entropic trend. Arbitrarily large departures from equilibrium can and, given enough time, will occur. The probability of any ordered macroscopic system randomly fluctuating into existence is utterly minuscule but still non-zero. In an infinite, eternal cosmos, even objects as complex as brains in states instantiating conscious experiences ('observations') appear. Of course, these freak observers are not necessarily organic human brains, if other sorts of systems, in particular less ordered (i.e. less improbable) ones, can also be conscious observers.

Freak observers are predicted by more recent developments in physics, as well. Quantum mechanics assigns, in various circumstances, a non-zero probability to the spontaneous formation of any macroscopic object. Bostrom (2002) considers the case of Hawking radiation, emitted by 'evaporating' black holes. Quantum freak observers also tend to appear in cosmology's reigning inflationary model of the universe (Gott 2008; Davenport and Olum 2010). A standard version of inflation predicts that dark energy increasingly dominates future phases of the universe, eventually flattening it out into an empty, endless de Sitter space (a vacuum solution to the field equations of General Relativity). In a de Sitter space, quantum fluctuations produce particles and occasionally structured macroscopic systems. Freak observers in every possible conscious state randomly fluctuate into existence somewhere, sometime.

Disturbingly, both Boltzmannian and inflationary cosmology seem to imply not only that freak observers exist, but that I should expect to be one. There are a couple of ways to think about this. Statistical mechanics alone, without the added assumption that low entropy at the Big Bang supplies an overall entropy gradient to the universe, seems to imply that all our apparent records of the past, including our 'memories', are almost certainly spurious. No matter how improbable it would be for all these records to randomly fluctuate into existence, this would be vastly less improbable than a 'normal' evolution from a state of even lower entropy. This ruthless logic promptly collapses my whole world. What statistical mechanics really says is that I should expect to find no more order than is entailed by my observations. Thus, not only the

history I think I remember, but everything I think I presently observe is almost certainly an illusion. I should expect to be the smallest (least improbable) departure from equilibrium that is sufficient for my present observations: a bare brain floating in the void, or perhaps something even less ordered. The freak observer problem is particularly acute for inflationary cosmological models that give every present region of space an eternal de Sitter future with infinite freak observers. Even if ordinary observers appear with some frequency in the era of stars and galaxies, there are infinitely more freaks.

The simulation and fluctuation arguments are structurally analogous except for the former's P1, the assumption of multiple realizability. Corresponding to P2 of the simulation argument is the claim that some mechanism produces freak observers that hugely or infinitely outnumber ordinary observers. Corresponding to P3 is the claim that we have no direct evidence for or against the fluctuation hypothesis. With these premises, the indifference principle dictates near certainty (or certainty, with an infinite preponderance of freaks) that I am a freak observer.

Andrei Linde, one of the architects of the inflationary model, notes that 'none of us wants to believe that he or she is a [freak observer]' (2007, p. 3). Indeed, no one wants to believe this, and probably no one of sound mind could believe it. Existential preferences and psychological limitations aside, however, what exactly is wrong with the argument? It might be claimed that our best physical theories of space and time do not, in fact, predict infinite freak observers. Indeed, not all cosmological models do; however, this is at least partly because cosmologists have developed or modified theory specifically in order to be rid of freak observers. Linde's (2007) model of eternal inflation, in which they exist but are forever outnumbered by ordinary observers, is conservative compared to Carlip's (2007) suggestion that the fundamental constants of nature (the fine structure constant, the electron-proton mass ratio, etc.) are slowly changing to values that suppress freak observer production. More startlingly, Page (2008) argues that the whole universe will decay (a quantum process resulting in the absolute annihilation of the 'bubble' of spacetime we inhabit) within 20 billion years, since this is projected to be a kind of point of no return, after which an eternal future full of freak observers is inevitable.

Theoretical innovations such as these are motivated by a simple bit of reasoning. Physicists generally adhere to the Copernican, or

mediocrity, principle, which states that I should expect to be typical among observers capable of the same general sort of cognition as I am (Gott 1993). But many assume that the premise of the fluctuation argument corresponding to the simulation argument's P3 is false, that I do have strong direct evidence against the fluctuation hypothesis. Hence, theory had better not predict that the typical observer is a freak (Davenport and Olum 2010, note 2). If the simulation argument could be turned on its head in this way, one could claim to have found evidence for the alternatives Bostrom considers (near universal extinction of civilizations before they become technologically advanced, or near universal loss of interest in world simulation) or for something more exotic, such as a heretofore unsuspected physical principle limiting computation in ways that prevent world simulation. But P3 of the simulation argument seems secure. Is the fluctuation argument worse off?

Gott (2008) suggests that I have evidence against the fluctuation hypothesis in the very fact of being conscious. This is supposed to be because, while the physical structure of a momentarily healthy and functional brain could form, it could not pass the Turing test. The reasoning here is murky. Davenport and Olum, discussing Gott's claim, suggest that a freak observer would fail the Turing test because it would be physically impossible 'for a brain arising as a vacuum fluctuation to send out any durable signal of its existence' (2010, p. 5), and hence impossible for anyone to administer the test. Gott himself seems to think that the problem is the duration of a freak observer – the vast majority are extremely brief. Even if one could administer the Turing test, 'no matter how many questions are answered successfully, the [freak observer] you see is likely to fail to answer the next one successfully (either by vanishing or by answering in nonsense)' (2008, p. 12).

It is not clear why the mere expectation that some entity will fail the Turing test should count against its being consciousness. I may communicate with another human whom, for whatever reason, I expect to die in the next moment, and I don't decide they aren't conscious now. Moreover, both Gott and Davenport and Olum appear to regard possible, and perhaps actual, success in the Turing test as a necessary condition for consciousness. This is at best a minority view (Oppy and Dowe 2011, sec.4.1); supporters of the Turing test regard passing it as a sufficient, not a necessary, condition for consciousness. Gott's suggestion also seems overly

fixated on a specific physical description of the freak observer as a brief brain. It is whatever minimal physical system is sufficient to be a conscious observer. If, for example, a 50ms bare brain isn't conscious, then that isn't an observer at all. But then entities lasting some longer duration will be.[2]

The very brief lifespan of almost all freak observers may be relevant for more straightforward reasons. Gott proposes the following test to confirm his ordinary observer status: 'I will wait 10 seconds and see if I am still here. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 . . . Yes, I am still here' (2008, pp. 12–13). Davenport and Olum endorse the same 'simple test' (2010, p. 2), and John Leslie (personal communication) suggests something similar in proposing that the consequent near certainty of instant demise 'smashes' any attempt to take the fluctuation hypothesis seriously. This objection has some appeal, but the simple test appears to be simple in the way that Samuel Johnson's 'refutation' of Berkleyan idealism was – it does not provide any real evidence. The whole point of the freak observer is that its memories, even (usually) its very recent ones, are illusory. Impatience or exasperation in the face of a ludicrous proposal is understandable, but the simple test begs the question.

Carlip and Page both claim that surplus order in my present experience is nearly decisive evidence that I am not a freak observer: 'it seems obvious that our observations and thoughts would be very unlikely to have the order we experience if we were vacuum fluctuations, since presumably there are far more quantum states of disordered observations than ordered ones' (Page 2008, p. 1). If there really is substantially more order in my present experience than there might have been, and if this use of the Copernican principle is legitimate, then this is a strong objection. But is there, and is it?

Consider first the claim of surplus order. If its truth seems obvious, I think this may be at least partly because the idea implicitly invokes a supposed recent *history* of ordered observation. An

[2] There may be more promising ways to run this objection. A freak observer would be a sort of quantum Swampman (Spaceman?) (Davidson 1987), and a representationalist such as Dretske (1995) would claim that, lacking a history, it lacks functions, and so lacks representations, and so lacks phenomenal consciousness. Assuming that phenomenal consciousness is something that I can know that I have, I can therefore know that I'm not a freak observer. Of course, many will reject the claim that type identical systems could differ phenomenally. Furthermore, other versions of representationalism might be flexible enough to allow a freak observer to acquire functions relatively easily (Adams and Dietrich 2004), in which case the minimal complexity, duration, etc. that is sufficient for a system to be a freak observer may simply be greater than is often supposed.

aspect of my present experience is my sense of having just experienced certain ordered observations, and of my present experience emerging or flowing out of those experiences. I feel confident that I have (or am) an extended stream of consciousness characterized by a high degree of continuity and stability. This supposed order, projected to be 'out there' beyond, say, a 500ms window of the psychological present, must not be surreptitiously included in the tally of the order in my present experience. Perhaps my confidence in immediately past order itself contributes to the order of my genuinely present experience, but I see no particular reason to think that it does, and have no clear sense of how the question could be decided.

Intuitively, the surplus order objection points to an abundance of conscious detail. And undeniably, at this moment, I do seem to introspect more detail than it seems that I might have. Again, though, there are obstacles to deploying this intuition as evidence. Introspection may not be a good way to discover the nature and extent of the detail represented in consciousness (Noë 2004; Schwitzgebel 2008), for one thing. But that aside, such detail as there is must not only be quantified somehow, but must be compared with a *lower limit* for inclusion of any mind in the reference class. Recall that Gott holds that the members of the reference class are those beings who are cognitively comparable to me. This is a difficult standard to use if I'm attempting to consider myself as only this momentary experience – I don't have many cognitive capacities to speak of. If I try to imagine a reference class of something vaguer, such as 'minds like mine', I'm not sure that I can make sense of the idea that it includes, for example, a mind consisting of a single observation of a monochrome ganzfeld, or a more complex but chaotic percept. The appeal to obviousness here, both about how much detail I experience and about how little I could have experienced, is less than we might hope for in a decisive objection to the fluctuation argument.

Even if my present experience does contain a genuine surplus of order, however, the objection's use of the Copernican principle may not be legitimate.[3] Hartle and Srednick (2007) argue that it

[3] A possible objection that I will not press is that the scenario being considered is one in which there are infinite freak observers of every level or degree of order, and it is senseless to speak of bigger and smaller infinities. Leslie (2001, p.28) addresses this claim in another context and gives the example of a dart board with infinitely many points in each section. Plainly, a dart has a lower probability of hitting the bull's eye than of landing elsewhere.

amounts to what they call the 'selection fallacy', the crypto-dualist mistake of supposing that I have been randomly associated with (as though 'implanted into') the perspective I occupy by some kind of universal lottery mechanism. But 'I' just am a physical system described within a certain set of observational data, and any two theories that predict that data with equal probability are indistinguishable on the basis of that data, no matter how atypical either theory might imply that I am. On this view, even considerable surplus order in my present experience can do nothing to reassure me that I am not a freak observer. In other words, the Copernican principle cannot be used to test theories. In particular, it does not tell me to reject a theory that predicts both that I am a freak observer and that there are myriad highly disordered observers for every one with an experience as ordered as mine (however ordered that really is).

Cosmologists with Copernican scruples reject Hartle and Srednick's heresy. Two of them, Garriga and Vilenkin (2008), clarify the 'ideal' reference class: those observers whose total observational data is identical to mine. By definition, I can have no evidence bearing on the question of my typicality within this class. Adhering to the Copernican principle here is a very minimal commitment. The principle is not being used to reject theories that imply the existence of a reference class within which I would be atypical. This is in keeping with the way practicing scientists regard the principle: as an empirical hypothesis that could itself be tested with observations. For example, astronomers have tested the classic version of the principle, that the Earth is not the centre of the observable universe, using measurements of the cosmic microwave background (Caldwell and Stebbins 2008; Zibin *et al.* 2008). As it happens, the data suggest that the Earth is not located at a privileged position in the cosmos (nor, from what we can tell, is there any such position), but the experiments presupposed that a genuinely anti-Copernican finding was possible.

So the Copernican principle is not in general regarded as too basic to test. A finding contrary to mediocrity would not have compelled astronomers to revise cosmological theories in order not to predict a reference class containing so many other, non-special positions at which we might have been located. A more banal case: I don't expect to win a raffle, and if someone tells me I've won, I will be surprised and seek confirmation. But if I have won, I simply discover that I am atypical, and I don't revise my worldview to restore that typicality. Someone had to win the raffle,

and I am that someone. Allowing that my present experience has surplus order while at the same time thinking that probabilistic considerations favour the fluctuation hypothesis is similar to this, except that there is nothing equivalent to buying a ticket, and no opportunity to form an expectation about the outcome of the draw. I have not, as Hartle and Srednick remind us, been entered into a universal lottery and won it (at least, relative to many other freak observers!) by getting to exist as an unusually ordered freak observer – 'I' just *am* this ordered freak observer.

I sum up this long objection with four quick points. First, the structures of the simulation and fluctuation arguments correspond closely, so if the former is acceptable as a general pattern of reasoning, then so is the latter.

Second, the simulation hypothesis and the fluctuation hypothesis conflict. While it is conceivable that I am a simulated freak observer, this could be the case only if there is a universe within which there is a simulation of another universe within which I am a freak observer. I cannot be a simulant and a freak observer in the same 'level' of reality.

Third, the fluctuation argument enjoys a comparable level of support for its premises. It has no premise corresponding to P1 of the simulation argument, so insofar as there is any doubt about multiple realizability of consciousness, the fluctuation hypothesis is better supported than the simulation hypothesis. As in the simulation argument, one premise predicts that non-ordinary observers exist and vastly outnumber ordinary ones. Certainly this is not the settled view of theoretical physics. However, freak observers do appear to be a stubbornly recurrent feature of mainstream cosmological theory, and the current cosmological consensus is the result of decades of rigorous and competitive experimental and theoretical work. Moreover, we've seen that some models in which freak observers either do not exist or do not outnumber ordinary observers have been developed at least in part in order to be rid of them, rather than on independent grounds. Credence in P2 of the simulation argument, in contrast, depends on considerable speculation, extrapolation, and imagination both about the technological feasibility of world and person simulation and about the biological, psychological, and social development of life in the universe. The best reason for rejecting the fluctuation but not the simulation argument is that the claim of no direct evidence bearing on the fluctuation hypothesis is false. I have argued that this objection can be resisted; it is

not supported by a careful and serious consideration of the available evidence, and it misuses the Copernican principle.

Finally, the fluctuation argument pre-empts the simulation argument. If I believed all the premises of both arguments, then I should believe the fluctuation hypothesis instead of the simulation hypothesis. Whereas the simulation argument supposes only that simulants would massively outnumber ordinary observers, the vacuum fluctuated freak observers of the eternal de Sitter future of any region of the universe infinitely outnumber any ordinary or simulated observers who exist in the early phases of that region. So, depressingly, the indifference principle dictates belief in the fluctuation hypothesis over the simulation hypothesis.

Epistemic Instability and Evidential Realms

Notwithstanding all of the above, I do not believe the fluctuation hypothesis, and not only because it is psychologically impossible to do so. I think that probably I have no (or almost no) reason to believe the simulation or the fluctuation hypothesis. This brief objection takes its lead from the troubling feature of the simulation argument that Dainton (2003) calls 'epistemic instability'.

Bostrom remarks that the argument 'adopts as its starting point that things are the way they seem to be and that science gives us reliable information about the world' (2005, p. 87). It is on the basis of this information (cosmological, biological, psychological, social, technological, etc.) that we would accept P1 and deem P2 somewhat probable. The problem is that, to whatever extent all this evidence gives us reason to believe the simulation hypothesis, it thereby gives us reason to believe that this evidence pertains only to a simulated universe rather than to the real world, the world that would have to actually produce the simulation. But then if there's no longer a reason to accept the conclusion, the premises are secure (as secure as they ever were), after all. And so on. Hence the instability. The fluctuation argument invites basically the same objection, although some who mention it do not press the point (Davenport and Olum 2010, p. 5).[4]

Dainton tentatively defends the simulation argument by suggesting that our scientific evidence, even within the context of the

[4] Carroll is one physicist who seems to think that instability may be the last best hope for defusing the argument, aside from his protest that 'there is no sensible way to live and think and behave' if one allows oneself to be persuaded (2010, p. 233).

argument, could still be judged a somewhat reliable guide to the real world because most human-crafted fiction, films, video games, etc. resemble our world (at least implicitly) in ways relevant to the argument. Bostrom may have something similar in mind in describing the simulations as 'ancestor simulations' – presumably ancestors and their descendents would inhabit worlds with substantially the same properties.

No such defence against epistemic instability is applicable with the fluctuation argument, but I think that doesn't matter, since this is not a compelling solution to epistemic instability in the first place. Is the linchpin of the simulation argument really to be the conjecture that the scientific banality of the bestseller rack and the multiplex is evidence for the limited imaginations of beings capable of simulating an entire universe? Whatever the plausibility of that claim, the proposal faces a problem akin to the problem of induction: the tendency of human creative productions to resemble our world is good evidence about the nature of the real world only if our world resembles that of our simulators, which is just what that evidence is needed to establish. Of course, the whole matter of how confident we can be that our world resembles that of our simulators would seem to be irrelevant if we are already supposing that there *is* a world of our simulators.

I think that epistemic instability is probably a fatal problem for the simulation argument in that it neutralizes the evidential support for the simulation hypothesis. However, I'll conclude by suggesting that the problem of evidence in the simulation argument can be posed as an objection in which instability per se plays no role.

The paradox of epistemic instability arises because of the hierarchical structure of universes implied by the simulation hypothesis, and the problem of relying on evidence collected from within (what is hypothesized to be) one 'level' of that structure. Consider how an analogy lacking such a hierarchical structure clearly misleads. Bostrom (2003, p. 8) imagines a scenario in which you discover that a certain bit of 'junk DNA', whose possession neither causes nor is correlated with observable characteristics, is prevalent in a population of which you are a member. The evidence you have for the general prevalence of the DNA sequence is a reason for you to think that you possess it (unless a personal genetic analysis reveals otherwise).

To see the problem with this analogy, imagine that you and another person are in a room, and you learn that only one of you

has the DNA sequence. You have no more reason to think that it is you than you have to think that it is the other person. But if you and another person are in a room, and you learn that one of you is a simulant and the other is a real observer, you have no trouble telling which is which. The simulant is the one living in the flickering CPU registers of the computer under the desk, and the real observer is the one sitting in the chair beside that computer. If you are in the chair, then you know that you are the real observer.

The simulation argument cites evidence, collected from within a certain realm called 'the observable universe', that this realm contains multitudes of simulants. But this evidence gives me no reason to think that I am one of them. Relative to this evidential realm, I just *am* a human being; 'the observable universe' refers to a realm containing the human being that is me. I am a real observer and the simulants, if they exist, live in an alien's computer somewhere. This is not dogmatism. The claim is not that I can 'just tell' that I'm not a simulant. I cannot tell that. In this sense, I am not denying P3. It is perfectly possible that I am a simulant – if this whole realm, including any alien computers sustaining simulants, is a simulation. But it makes no sense to suppose that I might be one of the simulants *of which I have evidence*. If I am a simulant, then they are 'meta-simulants', embedded in a simulation executed somewhere within this simulation. As before, I have no trouble telling which is which. And the numbers make no difference. Suppose I learn of the existence of one single simulant living out its deluded life in an alien computer somewhere in a faraway galaxy. I do not say to myself, 'Well, there's just the one, so it's very unlikely that I am it.' This response would be absurd. I am *not* it – I am me and it is it. This should be my response regardless of whether my evidence suggests that one or zillions or infinitely many such beings populate this realm, either presently or throughout spacetime.

None of this is to deny that I could acquire evidence that I am a simulant. But such evidence would have a very different character from the evidence cited by the simulation argument. For instance, Barrow (2007) suggests that error correction algorithms and coding shortcuts might have detectable effects on fundamental physical laws within a simulated universe, and Beane *et al.* (2012) describe how other presumed computational underpinnings of a simulation might have measurable effects on processes in high energy physics. Such evidence would be a glimpse of a

realm larger than this one. The evidence cited by the simulation argument, in contrast, is evidence that this is a larger realm than someone else's realm. The simulants in that realm, if it exists, may be in the same position as I am. Depending on the interests or whims of their simulators, they might experience a world much like this one, and they might hit upon the simulation argument and be tempted to suspect that they are simulants. But they would be wrong to be moved by the argument even though, ironically, its conclusion would be true of them.[5]

Simon Fraser University
Burnaby, BC, Canada
LyleCrawford@gmail.com

References

Adams, F. and Dietrich, L. (2004). 'Swampman's revenge: squabbles among the representationalists'. *Philosophical Psychology* 17 (3), 323–40.
Albrecht, A. and Sorbo, L. (2004). 'Can the universe afford inflation?' *Physical Review D* 70: 63528.
Barrow, J. (2007). 'Living in a simulated universe'. In B. Carr (Ed.) *Universe or Multiverse?* (Cambridge: Cambridge University Press), 481–86.
Beane, S., Davoudi, Z. and Savage, M. (2012). 'Constraints on the universe as a numerical Simulation'. arXiv:1210.1847v2 [hep-th].
Bostrom, N. (2002). 'Self-locating belief in big worlds: Cosmology's missing link to Observation'. *Journal of Philosophy* 99 (12): 607–23.
—— (2003). 'Are you living in a computer simulation?' *Philosophical Quarterly* 53 (211): 243–55.
—— (2005). 'Why make a Matrix?' In W. Irwin (Ed.) *More Matrix and Philosophy* (New York: Open Court), 81–92.
—— (2009). 'The simulation argument: some explanations'. *Analysis* 69 (3): 458–61.
Caldwell, R. and Stebbins, A. (2008). 'A test of the Copernican principle'. *Physical Review Letters* 100: 191302.
Carlip, S. (2007). 'Transient observers and variable constants, or repelling the invasion of the Boltzmann brains'. *Journal of Cosmology and Astroparticle Physics*. arXiv:hep-th/0703115v5.
Carroll, S. (2010). *From Eternity to Here* (New York: Dutton).
Dainton, B. (2003). 'Simulation scenarios'. Unpublished presentation at http://www.simulation-argument.com/daintonpower.ppt.
Davenport, M. and Olum, K. (2010). 'Are there Boltzmann brains in the vacuum?'. arXiv:1008.0808 [hep-th].
Davidson, D. (1987). 'Knowing one's own mind'. *Proceedings and Addresses of the American Philosophical Association* 60, 441–58.
Dretske, F. (1995). *Naturalizing the Mind* (Cambridge, MA: MIT Press).
Eliasmith, C. *et al.* (2012). 'A large-scale model of the functioning brain'. *Science* 338(6111), 1202–05.

[5] For their helpful and encouraging comments on versions and portions of this paper, I thank John Leslie, Jeff Foss, Owen Mackwood, Jillian McIntosh, Michael Tooley, and Sean Carroll.

Garriga, J. and Vilenkin, A. (2008). 'Prediction and explanation in the multiverse'. arXiv:0711.2559v3 [hep-th].
Gott, J. R. (1993). 'Implications of the Copernican principle for our future prospects'. *Nature* 363 (6427): 315–19.
—— (2008). 'Boltzmann brains: I'd rather see than be one'. arXiv:0802.0233 [gr-qc].
Hartle, J. and Srednick, M. (2007). 'Are we typical?' arXiv:0704.2630 [hep-th].
Leslie, J. (2001). *Infinite Minds*. (Oxford: Oxford University Press).
Linde, A. (2007). 'Sinks in the landscape, Boltzmann brains, and the cosmological constant problem'. *Journal of Cosmology and Astroparticle Physics*. arXiv:hep-th/0611043v3.
Markham, H. *et al.* (2011). 'Introducing the Human Brain Project'. *Proceedings of the 2nd European Future Technologies Conference and Exhibition* 7, 39–42.
Noë, A. (2004). 'Is the visual world a grand illusion?' *Journal of Consciousness Studies* 9 (5–6), 1–12.
Oppy, G. and Dowe, D. (2011). 'The Turing test'. *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition). E. Zalta (Ed.) URL =http://plato.stanford.edu/archives/spr2011/entries/turing-test/.
Page, D. (2008). 'Is our universe likely to decay with 20 billion years?' *Physical Review D* 78, 063535.
Schwitzgebel, E. (2008). 'The unreliabity of naive introspection'. *Philosophical Review* 117(2), 245–73.
Zibin, J., Moss, A. and Scott, D. (2008). 'Can we avoid dark energy?' *Physical Review Letters* 101: 251313.