Innocence Lost

Simulation Scenarios: Prospects and Consequences

Barry Dainton (2002, October), The University of Liverpool

Abstract
Those who believe suitably programmed computers could enjoy conscious experience of the sort we enjoy must accept the possibility that their *own* experience is being generated as part of a computerized simulation. It would be a mistake to dismiss this is just one more radical sceptical possibility: for as Bostrom has recently noted, if advances in computer technology were to continue at close to present rates, there would be a strong probability that we are each living in a computer simulation. The first part of this paper is devoted to broadening the scope of the argument: even if computers cannot sustain consciousness (as many dualists and materialists believe), there may still be a strong likelihood that we are living simulated lives. The implications of this result are the focus of the second part of the paper. The topics discussed include: the Doomsday argument, scepticism, the different modes of virtual life, transcendental idealism, the Problem of Evil, and simulation ethics.

1. The Simulation Menace

Imagine participating in a simple experiment. You are watching pre-recorded scenes from a televised soap opera unfold on a monitor in front of you; at the same time, in different rooms, nine other people are doing likewise. Or at least, they believe they are. In fact, only one of the screens is showing the original film featuring real actors; nine screens are showing a computer-generated film. The simulation is *very* good; so good, in fact, that the computer-generated images are visually indistinguishable from the originals. As is clear, if you had nothing but the onscreen images to go by, then (i) you would not be able to tell whether the people you are watching were real or computer-generated, and (ii) the odds that you are watching the film featuring real rather than virtual people are only one in ten.

Now consider an analogous case. As things stand, our abilities to create and control human streams of consciousness are severely limited. Let us suppose that in the future this changes, and it becomes possible to create human-type streams of consciousness, of any length, with any desired characteristics, very easily. Call the succession of streams which jointly compose the consciousness of a single person from birth until death, a *life-stream*. Despite their differences, your life-stream and mine, are of a certain general type: early 21st century human. Let us call these 'type-21 streams'. Now suppose that, for whatever reason, in the future *very* large numbers of type-21

streams will be created. To be more specific, suppose the total number of type-21 streams which exist after the year 2100 is ten times greater than the number which existed in the 21st century itself.

This scenario places you in a similar predicament as the first, but the consequences are rather more perturbing. Are you in a position to tell whether your experience is real or artificially generated? No. What are the odds that your experience is occurring when appears to be, in the early 21st century? Only one in ten. Although it seems to you that you are a normal human being living at the start of the 21st century, the subjects of all the many artificially produced type-21 streams have very similar impressions and beliefs. These subjects are all mistaken, and so might you be, for it is more likely than not that you *are* one of these subjects.

Following Bostrom, I will call this line of reasoning the *Simulation Argument*.[1] Although not everyone will find the possibility that their current lives are simulations something to be dreaded,[2] at the very least the argument threatens complacent assumptions about the status of our lives, and for this reason I shall sometimes refer to the *simulation menace* or *threat*. Many will no doubt be inclined to dismiss the argument as a mildly diverting but ultimately unthreatening curiosity, and for what might seem to be good reasons:

> *Could it be done?* The notion that future generations, or future civilizations, will be able to manipulate consciousness in the ways required is wildly implausible.

> *Would anyone bother?* Even if the required technology were to become available, our descendants would surely have better things to do than waste their time and energy producing realistic simulations of 21st century lives, at least in the vast numbers required for their existence to pose a significant threat to us.

In what follows I will argue that these objections carry less weight than might be supposed. Having established that the Simulation Argument should be treated with respect, I will move on to consider some of the implications of this.

2. Practicalities

I shall be using 'simulation' in a very broad way: any state or episode of consciousness is to be regarded as simulated if it is produced by non-standard methods in a controlled fashion (the degree of control may vary). Simulated experiences are of course real experiences in their own right, and while a simulated episode of consciousness may be a re-creation of an original non-simulated stretch of conscious life, it need not be. I shall say that a life (or part of a life) is *virtual* rather than *real* if it is entirely composed of simulated experiences.

Consciousness can be simulated in different ways, and to different degrees, and it will prove useful to have some of these differences in view before proceeding.

So far as degree or depth of simulation is concerned, we can contrast *complete* with *partial* simulations. The manufactured type-21 streams we encountered above are

---

[1] See Bostrom (2002b), and also the resources available at www.simulation-argument.com.
[2] E.g. S.R.L. Clark (1983).

examples of complete simulations: every part and aspect of experience is being generated by artificial means. In partial simulations, only *some* parts or aspects of experience are generated by artificial means. A simulation in which a subject is supplied with a wholly virtual environment (which here can be taken to include all forms of bodily experience) but retains their original psychology is one form of partial simulation. But we can also envisage cases in which the tampering is restricted to the domain of *inner* experience. Imagine having your psychology (e.g. memories, beliefs, desires, language skills, personality traits, and so on) replaced with a replica of Napoleon's, and then waking up in your own bed and perceiving your environment in the usual way. In what follows, unless otherwise stated, we will be concerned with complete rather than partial simulations.[3]

As for the *ways* in which consciousness can be simulated, it is important to distinguish what I will call *hard* (or *H-simulations*) from *soft* (or *S-simulations*). H-simulations result from directly tampering with the neural hardware ordinarily responsible for producing experience. S-simulations are streams of consciousness generated by running programs (software) on computers (other than the brain, if the brain is nothing but a computer).

The following scenarios outline some ways in which both H-simulations and S-simulations might exist in menacing numbers. The scenarios may seem far-fetched (to us, at any rate). Quite what we should make of them will be discussed in section 3.

Modal Realism and Other Worlds

According to the Modal Realist, all logically possible worlds are just as real as this one. Suppose David Lewis is right and Modal Realism is true. Even if creating human-type streams of consciousness in the actual world will always be a haphazard and time-consuming business, there are many logically possible worlds where this is not the case. There is, for example, a world which contains an infinite number of brains-in-vats, where each brain enjoys a recognizably human range of experiences. An infinite number of these vat-subjects, we can suppose, have menacing type-21 streams of consciousness, and your life-stream and mine are replicated many times over. Of course, this possible universe is not alone: there are infinitely many very much like it, differing only in the most trivial detail. There are also many other possible universes where other methods of simulating human-style consciousness are to be found. Suppose S-simulations are possible; if so, there are universes where every brain-generated consciousness is replicated by software running on computers. Suppose something like Cartesian dualism is true, and mentality is not purely material. It makes no difference: there are innumerable worlds, each containing innumerable disembodied immaterial souls, in which innumerable variants of type-21 streams of consciousness are instantiated. Needless to say, the subjects of these streams as completely unaware of their real predicament: they believe themselves to be embodied beings, living a normal life on 21st century Earth.

---

[3] More discriminating distinctions can be drawn. For example, it is possible for a subject to lead a virtual life – in the sense here defined – in the real world: think of the holo-doctor in Star Trek *Voyager*. This sort of case will not be relevant to what follows, where we shall be concentrating on simulations in which what is 'perceived' is *not* the real world.

If Modal Realism were true, the simulation menace would clearly be very real indeed. As for the odds of having a virtual rather than a real life, the calculation is not straightforward. Since there may well be an infinite number of worlds which are close replicas of (what we take to be) the actual world, there may well be an infinite number of worlds containing exact (or nearly exact) replicas of (what we take to be) our 21st century Earth. If so, then since every simulated type-21 stream could be paired off with a real type-21 stream, it could be argued that real and simulated streams are equal in number. In which case, the one's odds of one's life being real rather than simulated are at best only fifty-fifty.

Modal Realism is not the only potential source of simulations-in-other-worlds: some of the 'multiverse' theories encountered in the more speculative reaches of contemporary physics may have similarly menacing implications. Consider, for example, the kind of cosmos described by Smolin in which 'each black hole is a bud that leads to a new universe' (1997, 94). The variants of this model which have no beginning or end in time comprise an infinite number of sub-universes of varying character. A cosmos of this kind could easily contain an infinite number of Earth-like planets with Earth-like civilizations. Even if only some of the latter generate simulated type-21 streams, it could still be possible to pair off every real type-21 stream with a simulated type-21 stream.

S-Simulations

If all logically possible worlds are real, menacing simulations in truly vast numbers will inevitably exist. But even if only this world is real, menacing simulations may well exist in vast numbers – especially if S-simulations are possible.

Speculations as to what computers might one day be capable of are commonplace, but Frank Tipler takes things a good deal further than most. Tipler argues that if our descendants develop computers as far as they can be developed, given known physical constraints, we can all look forward to being resurrected in the far-future. Intriguingly, he suggests that our resurrection will not depend on our descendants having detailed knowledge of what our lives were actually like. The deduction runs thus:

(1) The computational conception of the mind is true. Any mental life, any stream of consciousness, can be replicated on a suitably programmed computer.
(2) There total number of possible human-like streams of consciousness (of finite duration) is finite.
(3) The processing power of the 'universal computer' that our descendants will develop will be effectively infinite.
(4) The universal computer will easily be able to simulate *every possible* human stream of consciousness (of finite length).[4]
(5) Hence our resurrection is all but inevitable: 'The dead will be resurrected when the computer capacity of the universe is so large that the amount of

---

[4] In fact, Tipler goes further. It is not just *minds* that will be simulated: 'an emulation of all possible variants of our world – the so-called visible universe – would require at most $10^{10^{123}}$ bits of computer capacity … this amount of computer capacity will be available in the far future.' (1994, 220)

capacity required to store all possible human simulations is an insignificant fraction of the entire capacity.' (1994, 225)

If the future turns out as Tipler predicts, the Simulation Argument has real bite: the probability that *your* life is a virtual life is extremely high. As Tipler himself notes: 'How do we know we ourselves are not merely a simulation inside a gigantic computer? Obviously, we can't know.' (1994, 207)

But Tipler's scenario rests on a good many colossal *ifs*. The future he describes may not be physically possible; even if it is, there is no reason to think it significantly likely that the future will turn out as he describes. However, there are far more modest forecasts which do not suffer from these flaws, but whose implications viz à vis the simulation menace are much the same.

If computer technology continues to advance at the rate it has for the past few decades, it will not belong before our most powerful computers equal, or exceed, the processing power and information storage capacity of a typical human brain. According to one of the more optimistic guesstimates, supercomputers should cross this threshold as early as 2010, with desktop machines of similar power arriving by 2030. A more conservative survey concludes that the breakthrough will certainly have been achieved by supercomputers around 2025, if present trends continue – and there is every reason to think they will.[5] Once such hardware becomes available it will be possible to simulate the computational activity of a human brain. Of course, this will require a fine-grained knowledge of our brains' structure and workings, but it may well be that the required knowledge will gradually be accumulated over the next few decades; significant strides in this direction have already been taken. On the assumption that mentality is a purely computational affair, computerized simulations of human brains could generate conscious mental lives that are subjectively indistinguishable from those generated by biological brains. S-simulations of this kind could be possible within the next half century or so.

A few such simulations pose no significant threat, but the situation becomes distinctly menacing if they start being produced by the billion or trillion. Such a situation could develop in at least two ways. The ability to produce menacing simulations could become very widespread, e.g., a hundred years from now everyone might own desktop (or handheld) computers easily capable of running them. If several billion computers were to possess this capacity, even if it were utilized only occasionally, menacing simulations would soon exist in disturbingly large numbers.[6] Alternatively, or in parallel, the capability of running large numbers of simulations might be found in the supercomputers of the not too far-distant future. Bostrom provides a Tipleresque illustration of the potential dangers:

---

[5] See Moravec (1999, 51-63). Moravec estimates the computer power of a single retina to be around 1000 MIPS (millions of instructions per second), which is about the power of a current PC. Since an entire brain is about a hundred thousand times bigger than a retina, he estimates the computing power of a brain to be around 100 million MIPS. Current supercomputers are capable of around 10 million MIPS. For the slightly more cautious appraisal see Bostrom (1997). Kurzweil is far more optimistic, arguing that a $1000 unit of computation in 2029 will have the power of approximately 1,000 human brains (1999, 220).
[6] Consider the massive current popularity of the computer 'God-game' *The Sims*, and suppose the simulated inhabitants are fully conscious – as they might be, in computer games of the future.

a rough approximation of the computational power of a single planetary-mass computer is 10^{42} operations per second, and this assumes only already known nanotechnological designs, which are far from optimal. Such a computer could simulate the entire mental history of humankind (call this an *ancestor-simulation*) in less than 10^{-7} seconds. (2002b, 3)

Less powerful devices, of the sort which might be available in the comparatively near future, would take somewhat longer, but the lesson remains much the same. If our descendants were able to run ancestor-simulations using only a small fraction of the computing resources available to them, they might very well do so, quite frequently. In such circumstances were to obtain, the risk that you and I are inhabiting a computer simulation would be high.[7]

H-Simulations

There is a further source of menacing simulations, one that has received less attention. Many of us have experienced fully realistic hallucinations, whether drug-induced or in ordinary dreams. Hallucinations produced by these means are typically *uncontrolled*: we cannot determine in advance the type of virtual world we will hallucinate, or the role we will play in the scenarios which unfold. This may very well change. Advances in brain science may make it possible to generate *controlled* hallucinations – or H-simulations – safely, easily and reliably. Almost inevitably, some of these controlled hallucinations will constitute menacing simulations.

One route to H-simulations requires the kind of neural implant and human-machine integration that is already familiar from science fiction. Interacting with computers mechanically – using screens, keyboards, etc. – is a cumbersome business, and a good deal of research is going into ways of facilitating the process. Among the methods already being considered, at least by the more adventurous researchers, are methods of connecting computers directly to brains. At present, such techniques are at a primitive stage of development, but this will no doubt change.[8] A hundred years from now, children could be growing up with implants buried deep in their heads, implants that both track and keep pace with their neural development, and allow their minds to interact directly with computers, on a number of levels, in a variety of ways.

It is not difficult to envisage some of the uses to which this sort of interface might be put. Your thoughts could be transmitted directly into someone else's mind – provided you were both hooked up to the same computer network. Forgetfulness would be largely

---

[7] There are other possibilities. Egan (1995) provides this intriguing illustration of how a significant simulation menace might arise through S-simulations: 'I was six years old when my parents told me that there was a small, dark, jewel inside my skull, learning to be me. Microscopic spiders had woven a fine golden web through my brain, so that the jewel's teacher could listen to the whisper of my thoughts. The jewel itself eavesdropped on my senses, and read the chemical messages carried in my bloodstream; it saw, heard, smelt, tasted and felt the world exactly as I did …. I thought: if hearing that makes *me* feel strange and giddy, how must it make *the jewel* feel? Exactly the same, I reasoned; it doesn't know it's the jewel, and it too wonders how the jewel must feel … - it too wonders whether it's the real me, or whether in fact it's only the jewel that's learning to be me.' The 'jewel' is a crystalline computer, programmed (in effect) by a living brain.

[8] See Nicolelis and Chapin (2002) for a useful survey of current work.

a thing of the past: your thoughts and experiences could easily be backed-up on a computer file, ready to be called on when required. More relevant to our purposes, fully immersive virtual reality would also be a possibility. There will be no need for you to wear a suit and visor to interact with machine-generated virtual worlds, your implants will perform the necessary tasks. Your sensory experience will be directly machine-controlled, via stimulation of the appropriate areas of the sensory cortex. The movements of your (simulated) body through virtual environments will be under your control, but there will be no need for you to actually move your physical body: the ways you *intend* to move your body will be detected by implants in your motor cortex and elsewhere, and this information will be used to generate corresponding movements of your virtual body. It will be possible to have a fully realistic experience of (say) flying a plane through narrow mountain passes while remaining motionless on a couch. You might even *believe* yourself to be an experienced pilot: your implants could ensure that a suitable set of false memories temporarily override your real memories. Alternatively, you might believe yourself to be an ordinary 21st century person, leading a typical life in a (virtual) 21st century environment.

If such technology were to be commonplace, it is by no means inconceivable that H-simulations would be generated in sufficient numbers so as to become menacing. People might take virtual reality 'trips' to the past quite frequently. They would certainly be used on an occasional basis during history lessons, and more intensively by historians, amateur and professional, with a particular interest in what it was like to live during certain periods of the past. But such trips might also be taken – far more frequently – for entertainment purposes. The soap operas of the future might well have an immersive/interactive character their present-day counterparts lack, computer games likewise. As is easily shown, the numbers soon add up.

Our descendants may 'visit' the past quite frequently, but since few are likely to want to spend significant portions of their lives in H-simulations, the concept of a *life-stream* introduced earlier is no longer appropriate as a basic unit of simulated consciousness. Something of briefer duration is required.[9] So, for present purposes, let us take *day-long* streams of uninterrupted consciousness – *D-streams* for short – as our working units. (An even shorter unit could be selected, but as will become evident, the upshot would not be greatly different.) We shall take as our class of *menacing* D-streams those simulated streams that resemble the sort of experiences enjoyed by actual inhabitants of the year 2002 – call these *MD-streams*.

Assuming the current population of the Earth to be six billion, there are just over 2×10^{12} D-streams for the year 2002. If a similar number of MD-streams of varying character exist in the future, the odds of the experiences you are currently having being simulated rather than original are around fifty per cent. Should the numbers of MD-streams created in the future be greater, your chances of living among the original inhabitants of the year 2002 will be correspondingly smaller.

In fact, the number of MD-streams created in the future could easily be far higher. Call the time at which H-simulations become commonplace occurrences the *C-threshold*. Let us suppose that from the C-threshold on, every future human being takes one virtual

---

[9] I assume here that our descendants have life spans not too different from our own – if their lives were far longer, the change from life-streams to day-streams would not be needed (my thanks to Stephen Clark for this point).

reality trip to the year 2002 during their lifetime, and that these trips are varied in character. If we now suppose that human civilization lasts for ten thousand generations after the C-threshold, and has an average population of ten billion, there will be 1.0×10^{14} MD-streams, compared with 2×10^{12} original D-streams. With fifty simulated streams for every real stream, you have a one in fifty chance of actually being alive in the year 2002. On more optimistic scenarios, your predicament is even more precarious. If humankind has a long history – one million generations exist after the C-threshold, say, with constant or improving technology – and a larger average population during this period – a hundred billion, say – then we can expect a total of around 1.0×10^{17} MD-streams to occur, which would reduce your chances of being alive in 2002 to around one in fifty thousand! In this case, even if only one in a thousand people ever take a virtual reality trip back to 2002, the chances that you are really living in 2002 are still only one in fifty.[10]

3 Assessments

Of the various ways in which a simulation menace might arise, Modal Realism is the most solid. It requires no second-guessing as to what is scientifically possible, or how the future might turn out, and the realm of *logical* possibility is so vast, so unconstrained, that menacing simulations are in inevitably in plentiful supply. But of course, Modal Realism is itself a highly controversial metaphysical doctrine, and few philosophers believe it to be true. That the doctrine entails a significant simulation menace will, for many, count as yet another reason for rejecting Modal Realism.

Somewhat similar considerations apply to scenarios involving multiverses of the menacing variety (i.e., those in which are such that it is reasonable to suppose that a significant proportion of all conscious lives are simulated).[11] Since the relevant physics is highly speculative and controversial, it is impossible at present to know how seriously theories of this type should be regarded. Nonetheless, the possibility that our universe has the character these theories predict cannot be ruled out, and the same can be said of the consequent simulation menace.

The issues raised by S-simulations are rather different. Those who have grown familiar with the claim that the human brain is the most complex object in the known universe may be surprised to discover that it will not be very long before we are able to construct machines of comparable complexity and computational power. But even if this is the case – and I suspect it is – the simulation menace posed by advances in computer technology is less severe than Bostrom and Tipler would have us believe. For S-simulations to constitute a threat they would have to be truly *conscious*. It is by no means certain that they would be.

---

[10] You might think: 'A few thousand years from now, would even one in a thousand people bother finding out what it was like to live in 2002? I think not.' This thought is understandable, but as we shall, there are also reasons thinking that the early 21st century might be 'visited' more frequently than most.

[11] If a single (non-branching) spacetime were infinite, menacingly large numbers of simulations could eventually be created by purely natural processes (e.g., vat-brains would spontaneously emerge from the quantum vacuum-field, or be emitted by black-holes – see Bostrom (2002a, 52-3)). But it is not yet clear that our spacetime is large enough for there to be a significant probability that extremely improbable occurrences such as these will actually ever occur.

Tipler takes a functionalist-cum-computationalist view of the mind as a given (1994, 124-7). If having a mind involves nothing more than possessing the right kind of causal organization, then in principle at least, minds can be implemented in physical (or non-physical) systems of radically different kinds, computers included. Given the way functionalists define mental states and properties – in entirely causal-functional terms – computers which replicate the causal-functional organization of a human mind simply cannot fail to be conscious. While this view certainly has its advocates, it also has many detractors. The reasons for this are well-known and familiar. Functionalism entirely overlooks, or ignores, the intrinsic qualitative features of experience – the very features which provide experiences with their *experiential* nature! Confronted with this objection, functionalists argue that the qualitative dimension of consciousness is irrelevant or illusory, but in the opinion of many, myself included, these heroic arguments are entirely unpersuasive. If functionalism is false, there is no guarantee whatsoever that computer-based simulations of human minds would be conscious.

This said, rejecting functionalism does not entirely eliminate the possibility of computer-based consciousness. Classical Cartesian dualism offers perhaps the most secure defence against this possibility, but another form of dualism offers no defence whatsoever. According to the doctrine of 'non-reductive' or 'dualist' functionalism, sympathetically explored by David Chalmers, experiences are non-physical, but they are nomologically correlated with functional organization; on this view, computer simulations of your brain would generate streams of consciousness indistinguishable from your own.[12] But as Chalmers himself would concede, non-reductive functionalism is at best a possible solution to the matter-consciousness problem. Materialism is another option, one that remains very much alive. Perhaps phenomenal properties are simply a certain kind of material property. If so, then it may very well be that human-type consciousness requires a human-type brain, or at least a biological system of a similar kind.[13] Of course, we cannot be certain of this. We do not know which parts or aspects of the physical processes in our brains are responsible for producing consciousness; consequently, we cannot rule out the possibility that the relevant physical processes could be replicated in very different physical systems – perhaps even silicon chips. But this is no more than a possibility, and in all likelihood, one that is quite remote if materialism is true.

So far as the computational challenge is concerned, the situation is clearly far from clear-cut. Whereas functionalists – of both classical and dualist persuasions – have good reasons for being very wary of future developments in computer technology, those who subscribe to different options in the philosophy of mind have far less reason to feel

---

[12] 'I will argue for a *principle of organizational invariance*, holding that given any system that has conscious experiences, then any system that has the same fine-grained functional organization will have qualitatively identical experiences … we might call [my doctrine] *nonreductive functionalism*. It might be seen as a way of combining functionalism with property dualism.' Chalmers (1996, 248-9) Bostrom notes that a weak variety of functionalism such as this is all that is required for S-simulations of the sort he introduces to pose a significant simulation menace: 'we need not assume that the thesis of substrate-independence is *necessarily* true (either analytically or metaphysically) – just that, in fact, a computer running a suitable program would be conscious' (2002b, 1).

[13] Lockwood (1989), McGinn (1991, 1999), Searle (1992), Strawson (1994) all defend version positions which combine materialism with a realist (or non-reductive) view of consciousness.

greatly concerned: brain-simulations that are not truly conscious pose no menace whatsoever.

By contrast, the threat posed by H-simulations looks to be very real indeed, irrespective of the position one adopts in the philosophy of mind. The simulations generated by the processes falling under this heading are indisputably streams of *consciousness*, rather than mere computation. Since the envisaged streams are generated by brains, rather than computers, even the most biologically oriented of materialists has reason to take their possibility seriously.

A materialist might argue that brain-computer connections, of the kind required for the scenario sketched above, would be so invasive and pervasive that their presence inside a brain would be incompatible with the production of conscious experience. Given our ignorance of the physical processes underlying consciousness, this possibility cannot be ruled out, but there is also little reason to suppose it very likely. After all, the envisaged interfaces would not replace neurons as experience-producers, they would merely provide ways of artificially controlling the triggering of neurons, or neural circuits – and we already know this to be possible on a small scale. Needless to say, the required nano-scale technology is far beyond anything we are capable of producing at present, and even if we could, our understanding of the brain's functioning is not sufficiently advanced for it to be deployed effectively. But anyone inclined to think this will continue to be the case should bear in mind Arthur C. Clarke's well-known dictum that any sufficiently advanced technology is indistinguishable from magic. A hundred years ago, now-routine procedures such as organ transplants and genetic engineering would have seemed miraculous. Is it not probable that a hundred years from now our descendants will be capable of similarly impressive feats?

It is not only materialists who should be open to the possibility of H-simulations, dualists should be too.

Even if our experiences unfold within immaterial substances, it is evident that our minds are profoundly dependent upon our brains. No contemporary dualist would be inclined to deny that the course of our sensory experience is dependent upon the neural activity within our brains, and this fact alone opens up the possibility of controlled hallucinations of a limited kind. But dualists should also recognize that appropriate neural manipulation could impact upon our conscious beliefs, intentions and desires. Intoxicants do not merely make it harder to control our bodily movements, they make it harder to *think* clearly, and there are numerous forms of brain damage that have more far-reaching (and often permanent) effects on our personalities and cognitive functioning, memory included. If brain damage can result in the permanent loss of certain memories, is it not likely that the memories to which we have conscious access depend on information stored in our brains? In which case, appropriate neural manipulation could lead a 23rd century person to have access to apparent-memories of the sort a 21st century person would have had.

But there is a further point to note, one that is relevant to materialists as well as dualists. Brain-computer interfaces of the kind I have been considering offer the possibility of very tightly controlled hallucinations, but there are undoubtedly other ways of inducing similarly life-like H-simulations, even if they offer rather less potential for fine-grained control. Ordinary, unaugmented, human minds are able to fashion richly-detailed and real-seeming virtual realities all on their own, almost effortlessly. Ordinary

dreams provide evidence both of this, and our ability to spin complex virtual worlds from limited and/or fragmentary evidence. I expect most of us have found ourselves having vivid dreams set in (say) the 17th century shortly after watching a film set in the same period. Although the dreamed-environment in such cases is inspired by what was seen onscreen, it often has a depth and complexity all of its own. Future methods of experience-induction could easily exploit these ordinary abilities. All that would be required is a safe and reliable drug which enabled people to enter a dream-like state at will, and also direct the general direction of the subsequent (fully life-like) hallucination – the framework for the latter could be supplied by a little prior reading, or the watching of video footage (e.g., of a 21st century televised soap opera). This method of controlling hallucinations could be put to the same uses – in, say, education and entertainment – as the computer-driven variant we considered earlier, and so is likely to be widely employed. So far as I can see, this method of inducing (partially) controlled hallucinations is not ruled out by any philosophical conception of the mind. It is also quite likely to prove attainable, perhaps quite soon.[14]

4. Consequences

Modal realism is metaphysically extravagant, developments in physics may lead to the Many-Worlds hypothesis' being discarded, and the computational conception of mind is highly suspect. But while those who reject these doctrines can take comfort in the idea that in so doing they are greatly diminishing the simulation menace, they would be wrong to conclude that the latter is non-existent. As we have seen, there are other ways of producing menacing simulations, ways that are harder to rule out. As one might expect, acknowledging the force of the Simulation Argument has a number of consequences; I will briefly outline a few of the more significant and intriguing.

WHERE DO WE STAND?

The simulation menace may be real, but it would be premature to suppose that we are all living virtual rather than real lives; after all, we do not know how the future will turn out. What the Simulation Argument does reveal, on the face of it at least, is a tension between the following propositions:

A. Humankind will have a long and successful future.

B. Technology will make it possible to manufacture realistic experiential simulations of any known type of human life, and these will be created frequently, in many and varied forms.

C. You and I exist in the early 21st century.

---

[14] Might *ordinary* dreams constitute menacing simulations? Perhaps, but I am inclined to think not, simply because I suspect my dream-experiences are somewhat less vivid and more course-grained than my ordinary waking experiences. I am not alone in this: see Flanagan (2000, 173-4). Of course, we cannot rule out the possibility that we are living in the dream-worlds of beings whose waking consciousness is far richer than our own.

Neither (A) nor (B) by themselves guarantee a serious simulation menace. Humankind could endure a very long time without ever developing (or deploying) simulation technology, hence the need for (B). But even if simulation technology is available and put to widespread and varied use, it may have to be in regular use for very long periods before menacing simulations are produced in sufficient numbers to constitute a significant threat to those of us living in the early 21st century– if so, (A) will be required in addition to (B). What seems clear is that the more confident you are that (A) and (B) are true, the less confident you should be that (C) is true. Proposition (C) is also in tension with:

(D) Modal Realism is true and/or a (menacing form of) multiverse theory is true.

Again, the more confident you are that (D) is true, the less confident you should be that (C) is true. [15]

DOOMSDAY POSTPONED

As far as (A) is concerned, the Simulation Argument undermines a line of reasoning which has led some to believe we humans *don't* have a long future ahead of us, and so is to this extent self-reinforcing.

The Doomsday Argument of Brandon, Carter, Gott and Leslie leads to the conclusion that the human race may well become extinct in the next hundred years or so. Why? Because you are an ordinary person, in no way untypical, and you find yourself alive at the start of the 21st century. Give your typicality, it is unlikely that you will find yourself alive at an exceptional period of human history, such as at a time when only a tiny percentage of all the humans who will ever exist are alive. It is much more likely that you will find yourself existing during a period when a when a sizeable percentage of all the humans who ever will exist are alive. Hence it is likely that *this* is such a time, and so unlikely that humankind will have a long and successful future. Our extinction is (probably) just around the corner. By analogy, suppose you wake up having forgotten your surname; it is much more probable that your surname will be found somewhere in the middle regions of a telephone directory, where most names are to be found, than on the first or last pages. In a similar fashion, other things being equal, it is more likely that an unexceptional person will find themselves being born in the period of history during which most other humans can be found.

The Doomsday reasoning has been criticized on technical grounds; there are those who believe the combination of anthropic and statistical reasoning it employs is

---

[15] Bostrom paints a slightly different picture (2002b, 5). He argues that at least one of the following propositions must be true: (1) The probability of humankind becoming extinct in the near future (or never developing powerful simulation technology) is close to one. (2) The probability of our descendants using their simulation technology to run ancestor-simulations is close to zero. (3) The probability that you are living in a simulation is close to one. But this only holds for those who share his theoretical commitments. Anyone who rejects the computational conception of consciousness could consistently believe that it is likely that our descendants will soon have computers capable of running huge numbers of ancestor simulations – and so reject both (1) and (2) – but also reject (3), on the grounds that the simulations in question will not be conscious, whereas we are.

illegitimate. This issue is complex, and still unresolved.[16] However, in the light of the preceding considerations, it is clear that even if the reasoning employed in the Doomsday Argument were valid, its conclusion would still be questionable, for it rests on a questionable premise, namely that *your* stream of consciousness *occurs when it seems*, in the early 21st century. Suppose you are optimistic about the prospects of humanity, you could reasonably argue thus:

> Since I am optimistic about the long-term prospects for humankind, the Simulation Argument leads me to believe that it is more likely than not that my stream of consciousness actually exists at a time *later* than the early 21st century, at some point in a (long) period during which the bulk of all the human-like streams of consciousness that ever occur do occur. This is precisely what the Doomsday reasoning predicts. Hence the Doomsday reasoning – in itself – does not provide me with any reason for reducing the probability I assign to the proposition 'Humanity will have a long and successful future'.

Propounders of the Doomsday Argument aim to persuade us that we (as a species) are less likely to have a long future than we (as individuals) would otherwise have thought. But if you find yourself believing it likely that humankind will have a long and successful future, the Simulation Argument should lead you to conclude that it is quite likely that you *actually exist* in the future (or at a time subsequent to the early 21st century), and since this is not at all a remarkable time to exist (assuming humankind endures for a long time), there is no reason for the Doomsday reasoning to affect your optimism about humankind's long-term prospects.

EXTINCTION MAY NOT SAVE US

Taking a broader, less parochial view, even if humankind were to become extinct in the comparatively near future – or survives but never develops (or deploys) simulation technology – it might not matter, for you and I could still be leading virtual rather than real lives. Modal Realism provides one route to this conclusion, as do certain multiverse scenarios, but there at least one further route: menacing simulations could be created by extraterrestrial civilizations.

There is no need for aliens ever to visit Earth. The information needed to generate simulations of human experience is even now starting to be broadcast to the stars, and this will no doubt continue for the next few years – or until our extinction invervenes.[17] Any alien listening post within a few dozen light-years of the Sun would be in a position to detect even fairly weak radio signals emanating from the Earth which, when decoded, would provide a detailed knowledge of our physical constitution (e.g., a

---

[16] Korb and Oliver (1998), Bostrom (1999), (2002a).

[17] The BBC's radio broadcasts from the 1936 Berlin Olympics were probably the first signals with the right wavelength to penetrate the ionosphere and sufficient strength to be detected in interstellar space; many more have been transmitted subsequently. In 1974 a simple message was directed at a globular cluster of some 300,000 stars with the *intention* of revealing our presence (but it will not be received for another 25,000 years or so). This idea is not new: as long ago as 1820 Gauss suggested carving mathematically meaningful patterns in the Siberian forests so as to alert extraterrestrials to our existence. (Drake and Sobel, 1991, 170)

typical human DNA sequence, which no doubt will soon be broadcast, if it hasn't been already) and our ways of life (e.g., radio talk shows, televised soap operas). This knowledge could, in principle at least, be transmitted to the alien home-worlds and used to generate simulations of typical human streams of consciousness. The fact that these worlds might be so distant that signals traveling at light-speed from our neighbourhood would take thousands (or millions) of years to arrive is of no consequence so far as the possible production of simulations is concerned – likewise the possibility that humankind might be extinct before the signals arrive.

This possibility, and others of the same ilk, should be of greatest concern to computationalists. Given the sort of information about ourselves we will shortly be providing, a technologically advanced alien species should have no difficulty in creating and running structurally accurate S-simulations of ordinary human beings, and ordinary human life. If simulations such as these would conscious in *our* computers, as computationalists believe, there is no reason to suppose they would *not* be conscious in *their* computers. It is not difficult to envisage scenarios in which human-type simulations are created in sufficient numbers so as to become menacing (e.g., alien historians might run a good many full-scale exploratory simulations of the period of human civilization in the period leading up to its demise, varying parameters to see how this unfortunate occurrence could have been averted).

Materialists and dualists have rather less to worry about, but neither camp should rule out this possibility: aliens could use the information about ourselves that we are transmitting to create *approximations* of human experience in controlled hallucinations.[18] It is not beyond the bounds of possibility that your current experience is a reflection of human life viewed through the (inevitably) distorting lens of a non-human consciousness.

These speculations may strike some readers as particularly extravagant.[19] But don't forget: there are thousands of billions of stars in our galaxy, and billions of other galaxies. Even if intelligent life is very uncommon, there may still be a lot of it out there, and since time and distance are not pressing factors so far as simulation (as opposed to communication) is concerned, a significant simulation menace from these quarters cannot be ruled out.

EVIL

The Simulation Argument goes some way towards providing a solution to the traditional Problem of Evil. The problem is not 'How could an omnipotent and benevolent God create a universe where bad things happen to some people' – given free will, some people will, on occasion, choose to act in ways which result in other people suffering. The problem is rather this: 'How could an omnipotent and benevolent God create a universe where there's so much *pointless* suffering, suffering that results from natural causes, rather than the actions of free agents.' But if your current life is a virtual life, then it

---

[18] These approximations need not be based on information gleaned from us. Aliens thoroughly acquainted with the relationship between material and phenomenal processes might simply simulate the sort of experience available to DNA based life-forms as part of a general investigation into the possible forms of experience. Thanks to Nick Bostrom for this suggestion.

[19] Of course, they merely involve the logical extension of a technique which features prominently in the *Jurassic Park* movies – see Crichton (1991).

could well be that the universe as a whole has (proportionally) a lot less pointless suffering than your virtual universe. Perhaps the unpleasantness to be found in the latter is due to the free choices made by some future (not very moral) human – the person who set up the parameters of the virtual world you inhabit – hence God is not *directly* responsible for the unpleasantness in question. The Simulation Argument thus provides an unexpected boost to the Free Will solution to the Problem of Evil.

A NEW SCEPTICISM

The Simulation Argument can be viewed as a posing a sceptical challenge, but the scepticism involved is of a novel kind. Ancient Greek sceptics argued that since our senses can deceive us we can never be justified in supposing that the world is how it seems, but the idea that there might not even *be* an external world never occurred to them.[20] For the latter hypothesis to be thinkable consciousness must be construed as a self-contained and potentially autonomous realm of existence in its own right. Descartes was the first to articulate this conception clearly, and drew the (now) obvious sceptical conclusion: our experience could be (subjectively) just as it is even if the reality external to our consciousness is very different from how we believe it to be on the basis of our experience, and consequently, we cannot be certain that the physical world exists.

Simulation scepticism (as we might as well call it) involves a rather different claim:

> *Even if* there is an external physical world, and this world has the character suggested by our experience, it is not only *possible*, but quite *likely* (given certain assumptions) that our current experience is hallucinatory rather than veridical.

So far as the existence of a mind-independent reality is concerned, Simulation sceptics are as one with their Greek predecessors: it exists. But Simulation sceptics are at one with Cartesian sceptics when it comes to the status of our current consciousness: both hold that it could be purely virtual, a detailed and convincing hallucination. However, for the Cartesian this conclusion relies on the reality external to our consciousness being very different from how it appears (e.g., a malicious Demon). Simulation sceptics, by contrast, derive this conclusion from the assumption that our experience is an accurate guide to the character of that portion of the external reality it seems to concern (in our case, the early 21st century).

Simulation scepticism is less radical than its Cartesian counterpart, but it is also less of a blind alley. Different hypotheses as to how the future may turn out, or what the universe may contain, render the hypothesis that we are leading virtual rather than real lives more or less likely, and these hypotheses can be refined, explored and evaluated. Unfortunately, this gain comes at a price. The threat posed by Simulation scepticism is far more *real* than that posed by its predecessors. Cartesian scepticism is hard to refute, but as Hume noted, it is also hard to take seriously: few of us spend much time worrying about the possibility that reality could be radically different from how it seems. Simulation scepticism reveals that even if reality *is* largely as we believe it to be, there could be a high probability that our actual condition is very different from our apparent

---

[20] Or so it has been argued, cf. Burnyeat (1982)

condition. As things stand, with simulation technology still at a primitive level, many will find Simulation scepticism as hard to take seriously as its Cartesian counterpart. This will no doubt change, as the technology advances.

Would taking Simulation scepticism seriously have any practical implications for how we should lead our lives? Like other modes of scepticism it has the potential to console. Anyone who has had cause to regret an action because of its consequences – and that includes most of us – will be cheered by the thought that these consequences might not in fact be real. The downside of this is that the same might apply to some of our most valued achievements and relationships. But since simulations come in different forms, and we cannot be certain that our lives *are* simulated, it would clearly be wrong to abandon all one's ambitions, or act without giving thought to the consequences.

Simulation scepticism differs from its kin in one further respect. Whereas everyone is equally vulnerable to Cartesian doubt, Simulation scepticism has the potential to be more discriminating. Anyone who thinks they are living an especially interesting life, and so having experiences which future simulators might be interested in sharing, will be led to the conclusion that their life has a greater than average chance of being virtual. It is hard to predict what effects this realization might have on (say) political leaders of the future, but they may not be wholly beneficial.

For those of us living only average lives, I suspect the practical consequences of Simulation scepticism will prove to be few. Even if we knew the probability of our lives being simulated, which we don't, this knowledge would be useless unless we also knew what sort of simulation was being run and the intentions and preferences of the simulators.[21] Until such knowledge is forthcoming, it seems best to continue much as we would otherwise do. Also, as Bostrom notes: 'Our best guide to how our … creators have chosen to set up our world is the standard empirical study of the universe we see' (2002b, 7). This seems right.

Not everyone will view simulation as a threat or menace. Despite the potential losses, there are people who would welcome the prospect of their current lives being virtual: there are many who would gladly exchange a life in the 21st century for a life in a later century, and all the wonders there to be found. This attitude is comprehensible, but a few cautionary words are in order. Only H-simulations offer this prospect, S-simulations do not. Also, H-simulations are themselves far from risk-free. If you were to emerge from your simulated 21st century life at a later time (or different world), you might well find that your 21st century personality was as much a simulation as your 21st century surroundings.[22] Moreover, you may find your new (non-virtual) circumstances

---

[21] The difficulty of attempting to second-guess the likely preferences of simulators is illustrated by Hanson's (2002) recommendations as to how one should act so as to reduce the risk of the curtains being brought down on one's virtual world. He concludes thus: 'If you might be living in a simulation then all else equal it seems that you should care less about others, live more for today, make your world look likely to become eventually rich, expect to and try to participate in pivotal events, be entertaining and praiseworthy, and keep the famous people around you happy and interested in you.' I am not sure that anyone could conform to *all* these injunctions simultaneously, but even attempting to do so might well make one so obnoxious as to hasten one's end.

[22] On some views of personal identity, radical psychological discontinuity (sudden memory loss or change, etc.) cannot be survived, so the person 'emerging' from you simulation would not be you. See Dainton (1995) for an entirely consciousness-based account of personal identity which does not have this consequence.

might leave much to be desired. After all, it is likely to remain the case that many of those inclined to indulge in simulatory practices will do so because they are less than fully satisfied with the circumstances in which they find themselves.

SOME VARIETIES OF VIRTUAL LIFE

The distinction between H-simulations and S-simulations reveals one way in which simulated lives can be subjectively indistinguishable but different in kind. Here are some others:

Active v. Passive (A-simulations v. P-simulations) The subjects of A-simulations are confined to virtual environments, but in all other respects they are free agents – or as free as any agent can be. Their actions are not dictated by the virtual-reality program, they flow from their own individual psychologies, even if these are machine-implemented. A P-simulation, by contrast, is a completely pre-programmed course of experiences. The subjects of P-simulations may have the impression that they are autonomous individuals making free choices, but unlike their A-simulation counterparts, they are deluded: all their conscious decisions are determined in advance by the virtual reality program. Such subjects have *apparent* psychologies – their consciousness is subjectively similar to that of someone with an active psychological system, so they have apparent memories, hopes, fears, etc. – but their real psychologies are entirely suppressed (or even non-existent).

Original Psychology v. Replacement Psychology (simulations$_{OP}$ v. simulations$_{RP}$) In A-simulations, a 'replacement psychology' is an artificially-generated system of beliefs, desires, memories, intentions, preferences, personality traits and so forth that supplants a subject's own ('original') psychology. The same applies in P-simulations, the difference being that the 'replacement' psychology is only *apparent*, in the sense just introduced. There is a sense in which the inhabitants of simulations$_{RP}$ are doubly deceived: not only is their environment not what it seems, neither are their minds. Obviously, the 'original' v. 'replacement' distinction only applies to subjects who have (or once had) a non-virtual life; the psychologies of subjects who only exist in virtual worlds are 'original' without exception.

Communal v. Individual (C-simulations v. I-simulations) A C-simulation is a virtual environment shared by a number of different subjects, each possessing their own distinctive individual psychology (even if these are machine-implemented). An I-simulation is restricted to a single subject. Of course, the subject of an I-simulation may meet what they take to be other people in their virtual worlds, but these 'others' do not possess their own individual autonomous psychological systems – they are not subjects in their own right, merely parts of a machine-generated virtual environment.

These options can be combined in various ways, e.g., a simulation of type AC_{RP} is active, communal with replacement psychology, whereas a simulation of type PI_{OP} is passive, individual with original psychology. There is a total of eight permutations:

AI_{OP}:   Active/Individual/Original Psychology
AI_{RP}:   Active/Individual/Replacement Psychology

AC_{OP}:   Active/Communal/Original Psychology
AC_{RP}:   Active/Communal/Replacement Psychology

PI_{OP}:   Passive/Individual/Original Psychology
PI_{RP}:   Passive/Individual/Replacement Psychology

PC_{OP}:   Passive/Communal/Original Psychology
PC_{RP}:   Passive/Communal/Replacement Psychology

Assuming that each of these modes could be generated by either H-methods or S-methods, we have a grand total of sixteen distinct kinds of (subjectively indistinguishable) virtual life. But the situation may not be quite so complex. A strong case can be made for thinking that a truly *communal* simulation of the passive variety is impossible. There is nothing impossible in the idea of a number of subjects simultaneously playing out roles in similar and coordinated hallucinations, but unless these subjects can causally interact with one another – something which cannot occur in P-simulations – they can scarcely be said to constitute a genuine community. For this reason it seems right to regard all P-simulations to be of the individual variety. This brings our grand total down to twelve.[23]

THE VIRTUAL AND THE REAL

Although I have been contrasting real and virtual modes of existence, some virtual realities are more real than others. Subjects inhabiting I-simulations – whether active or passive – are in a similar predicament to people suffering lasting and life-like delusions. Their environments may seem perfectly real, but they do not extend beyond the confines of their consciousness. The environments of the inhabitants of C-simulations are very different: they are shared, rather than private, and so well-founded distinctions can be drawn between 'appearance and reality' (and 'objective and subjective'). C-simulants can conduct empirical explorations of their virtual world, and agree and disagree on their findings, in all the ways available to the inhabitants of non-virtual worlds. Earlier, I characterized Simulation scepticism thus: '*even if* there is an external physical world, and this world has the character suggested by our experience, it may be quite likely that our

---

[23] To simplify, I overlook here the fact that the distinction between H-simulations and S-simulations may not be absolute: the consciousness of future humans (or post-humans) may be sustained by a combination of neural and artificial means, and neurons themselves may be genetically manipulated. I also ignore the fact that in some logically possible worlds, simulations are created by quite different means (e.g. magic). It should also be noted that simulants of different types can coexist, e.g., *The Matrix* features a combination of H-simulations (ordinary humans) and S-simulations (the 'agents'), both active, coexisting in a single C-type virtual environment.

current experience is hallucinatory'. We are now in a position to see that this requires qualification. The experiences of subjects inhabiting I-simulations *is* hallucinatory, in the usual sense of the term, but the predicament of those inhabiting C-simulations is less straightforward. Although their perceptual experience does not in fact reveal the real world (as they naively suppose), it does reveal a world which possesses at least some of the defining properties of 'reality'.

In Kantian terms, virtual worlds of the communal variety are *empirically real*, even if *transcendentally ideal*. The Simulation Argument thus injects new life into the Kantian claim that *our* world – the world we perceive and interact with – is no more than 'empirically' real. But bearing in mind the possibility that many simulations are I-type rather than C-type, our world(s) may also be less than empirically real, despite appearances to the contrary.

SIMULATION ETHICS

Since simulated lives are subjectively indistinguishable from the real thing, their creation is by no means a trifling matter, morally speaking. Even if our descendants develop the means of producing such simulations, might they choose not to do so? Might ethical scruples eliminate or at least diminish the threat posed by the Simulation Argument? It is hard to be sure – our descendants may be swayed by moral considerations we cannot now anticipate – nonetheless, there are reasons for thinking it unlikely.

It is easy to conceive of morally abhorrent simulations. An obvious example would be S-simulations of entire virtual worlds, of the sort computationalists believe possible, all of whose inhabitants suffer nothing but perpetual pointless torment. It is far from inconceivable that our descendants will forbear from creating such things; after all, ways might be found of controlling, or eliminating, the malicious tendencies which plague contemporary societies. But not all large-scale simulations are clearly morally abhorrent, far from it. Would creating an ancestor-simulation – a complete simulation of human-history up until the present time – be morally wrong? It is not at all obvious that it would. The sum total of human misery may be immense, but so too is the sum total of human happiness, and on balance, most people are glad to have had the opportunity of existing. Given this, what could be immoral about creating ancestor-simulations? Since the inhabitants of an ancestor-simulation would feel the same way about their lives as we do about ours, mightn't it be immoral *not* to create ancestor-simulations, if one had the means of so doing?

But the situation is by no means this straightforward. The fact that simulations need not be unpleasant does not mean their creation is morally unproblematic:

> **The Objection from Lesser Value** A real life has greater intrinsic value than a subjectively similar simulated life. Since it is wrong to impose on others a low-grade form of existence that one would prefer to avoid oneself, the creation of simulated lives is immoral.

This objection is weak. Even if virtual lives do possess less intrinsic value than their non-virtual counterparts, other things being equal, they can still be lives worth living, and hence lives that are worth creating. But there is a second point to note.

As Nozick's imaginary case of the experience machine reveals, the desirability of a life is not determined solely by the desirability of the experiences it contains. An experience machine will supply you with a (virtual) life of any kind you like, so by connecting yourself up to such a device you are guaranteed a very enjoyable (virtual) life, a life in which as many of your desires as you choose to come true will come true. But as Nozick notes, few of us would choose permanently to connect ourselves to an experience machine if we had the opportunity of so doing, and for good reason: 'What is most disturbing about them is their living of our lives for us.' (1980, 44) This lesson is important. But it is also of limited relevance.

The virtual lives sustained by experience machines are of the *passive* kind: they consist of solitary streams of consciousness that are completely controlled and pre-programmed. As we have seen, not all virtual lives need be like this. Of particular interest here are AC-simulations, i.e., virtual lives that are both *active* and *communal*, in the senses introduced above. Subjects in AC-simulations possess their own autonomous psychologies (whether original or replacement). They lead their own lives: their actions are not pre-programmed (they are as free as anyone can be). And they can causally interact with other subjects in their virtual environment (and these other subjects are autonomous individuals in their own right, rather than merely the appearances of such). Given all this, it is hard to see why life in an AC-simulation should be regarded as being inherently less valuable or worthwhile than a normal life. True, the inhabitants of AC-simulations are not physically embodied in the normal way, but they possess virtual bodies that are indistinguishable from the real thing. They are unable to manipulate physical objects, but they can manipulate virtual objects which *seem* physical. Why should the undetectable absence of a material environment significantly diminish the value of the lives of these subjects? I cannot see any reason why it should.[24]

These considerations further weaken the Objection from Lesser Value. Even those who find this objection persuasive would only have reason to avoid creating passive simulations; there is no reason why they should be reluctant to create AC-simulations.

However, there is a further, and potentially more serious objection to the fostering of virtual life:

> **The Deception Objection** The subjects of simulations are being deliberately deceived; their lives are virtual, but they believe them to be real. This deception is engineered and maintained by the relevant simulators. Such actions are clearly wrong.

Deception is not an inevitable consequence of simulation; there may well be simulants who are perfectly aware of their true condition. But since few contemporary humans believe themselves to be leading simulated lives, the Deception Objection does apply to

---

[24] Berkeley, who believed that we all have virtual lives (organized and coordinated by God) was perhaps the first to make this point, when he argued for the redundancy of mind-independent material reality. For an interesting fictional depiction of what life in an AC-simulation might be like see Egan (1998).

simulations of the menacing variety. This is not to say that it will have an impact on simulation policy. It is conceivable that future simulators will take the view that although deception is wrong, the kind of deception being perpetrated on simulants does not constitute a wrong that is sufficiently serious to outweigh the boon of existence. But equally, it is conceivable that future simulators *will* be swayed by the Deception Objection, and restrict their simulation activities accordingly. This may not seem likely, but since we can only guess at the ways ethical considerations will influence the simulation policies of our descendants, it cannot be ruled out.

However, it would be wrong to suppose Deception Objection has equal force against simulations of different types. It has considerable force in the context of long-term S-simulations of entire civilizations: anyone who creates an ancestor-simulation is responsible for the deceiving of billions of (virtual) people for thousands of (subjective) years. The situation is very different in the case of small-scale, short-term H-simulations. You are, let us suppose, feeling run-down by the demands of your 22nd century job, and decide to spend a couple of days in the (virtual) past to unwind; you employ the method of self-induced controlled hallucination, and 'wake up' in early 19th century England, in the midst of the Napoleonic wars. As you enjoy your adventure, are you the victim of a deliberate deception? In a sense, yes: you have opted for the fully-immersive trip, and so believe yourself to be a typical early 19th century person. But is the kind of deception involved in this case morally problematic? Surely not. Rather than one person imposing an uninvited delusion on another – as in the case of ancestor-simulations – we are dealing here with a person freely choosing to impose a short-term and harmless delusion *upon themselves*. Where is the wrong in that?

This implications of this point are by no means trivial, for as we saw earlier, given sufficient time, H-simulations might easily be created in sufficient numbers so as to be seriously menacing. But we are not yet done. There is at least one further reason why our descendants might avoid indulging in menacing simulatory practices, a reason that is pragmatic rather than ethical:

> **The Self-Interest Consideration** Future generations will be well acquainted with the Simulation Argument, and so will impose tight restrictions on simulation creation. They will realize that unless such restrictions are imposed, and enforced, no one – themselves included – will be in a position to rule out the likelihood that their lives are virtual rather than real.

I am not confident that such a policy will ever be adopted, for a number of reasons. (1) Simulation technology is certain to play an increasing role in recreational activities, and people will become accustomed to, and demand, ever more lifelike simulations – just as today there is a demand for every more life-like computer games. Since a ban on lifelike simulations would be unpopular with both the public and powerful commercial concerns, the prospects of one's being implemented are slim. (2) Most people will be unlikely to take the Simulation Argument seriously until they themselves have experienced what the technology can do, and taken a fully-immersive trip to the past or future. Should this point every be reached, billions of menacing simulations will have been created, and it will be obvious to everyone that it is already too late to consider a ban. (3) To have the desired effect, a ban on simulations would have to be continued into the indefinite future.

But even if an effective ban could be enforced in the present, we could never be confident that this policy would not be abandoned, or fail, at some future date. For this reason alone it is unlikely that our descendants would be willing to deprive themselves of all the benefits advanced simulation technology makes available.

There is a more general point. We are in the process of emerging from an age of innocence, an innocence that we are unlikely ever to recapture. Innocence was being able to believe that only sceptical possibilities of the most radical sort stood between ourselves and the world about us. This innocence evaporates on contact with the knowledge that even if reality is much as it seems, there is a significant likelihood that our current consciousness is simulated. Having to live with this knowledge may well be part of the normal lot of technologically advanced conscious beings the universe over. When this realization fully dawns on our descendants, attempting to recapture their lost innocence by imposing restrictions on simulatory practices will very likely strike them as futile. Since any restrictions on simulation creation can always be lifted subsequently, it will be obvious that their imposition would offer only meagre protection against the menace of simulation. But another factor will enter into the reckoning. Even if innocence once lost is impossible to regain, innocence can of course be *simulated*. If our descendants want to escape the shadow of simulation and experience for themselves what it was like to exist in more innocent times, they may have but one option: to embark on fully immersive virtual reality trips into the past. Not only does this further reduce the chances of restrictions on simulation creation being imposed, it is also bad news for our predecessors. It could easily be that the vast majority of people who find themselves living in more innocent times are simulants.

Our own predicament is only slightly better. Many of our descendants might be tempted by the prospect of finding out what it was like to *become* aware of the simulation menace; experiencing the first falling of the shadow might be irresistibly appealing prospect. If so, life in the early 21st century may be an even more fragile thing than it appears.[25]

References

Bostrom, N. (1997) 'How Long Before Superintelligence?'
http://www.nickbostrom.com/superintelligence.html

Bostrom, N. (1999) 'The Doomsday Argument is Alive and Kicking', *Mind*, Vol. 108, No.431, 539-50

Bostrom, N. (2002a) *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, New York: Routledge

Bostrom, N. (2002b) 'Are You Living in a Computer Simulation?' *Philosophical Quarterly* (forthcoming)

Burnyeat, M.F. (1982) 'Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed', *The Philosophical Review*, XCI, No.1, 3-40

Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press

Clark, S.R.L. (1983) 'Waking Up: a neglected model of the afterlife', *Inquiry*, 1983

Crichton, M. (1991) *Jurassic Park*, London: Arrow

Dainton, B. (1995) 'Survival and Experience', *Proceedings of the Aristotelian Society*, 17-36

Drake, F. and Sobel, D. (1991) *Is Anyone Out There?*, London: Simon and Schuster

Egan, G. (1995) 'Learning to be Me', *Axiomatic*, London: Orion

Egan, G. (1998) *Diaspora*, London: Gollanz

Flanagan, O. (2000) *Dreaming Souls: Sleep, Dreams, and the Evolution of Conscious Life*, Oxford: Oxford University Press

Hanson, R. (2001) 'How to Live in a Simulation', *Journal of Evolution and Technology 7* (http://www.transhumanist.com)

Korb, K. and Oliver, J. (1998) "A Refutation of the Doomsday Argument", *Mind*, Vol. 107, No. 426, pp. 403-410

Kurzweil, R. (1999) *The Age of Spiritual Machines*, London: Orion

Leslie, J. (1996) *The End of the World: The Science and Ethics of Human Extinction*, London: Routledge

Lockwood, M. (1989) *Mind, Brain and the Quantum*, Oxford: Blackwell

McGinn, C. (1991) *The Problem of Consciousness*, Oxford: Blackwell

McGinn, C. (1999) *The Mysterious Flame*, New York: Basic Books

Moravec, H. (1999) *Robot: Mere Machine to Transcendent Mind*, New York: Oxford University Press

Nicolelis, M.A.L. and Chapin, J.K. 'Controlling Robots with the Mind', *Scientific American*, October 2002, 24-31

Nozick, R. (1980) *Anarchy, State, and Utopia*, Oxford: Blackwell

Searle, J. (1992) *The Rediscovery of the Mind*, Cambridge: MIT

Smolin, L. (1997) *The Life of the Cosmos*, London, Weidenfeld & Nicholson

Strawson, G. (1994) *Mental Reality*, Cambridge: MIT

Tipler, F.J. (1994) *The Physics of Immortality*, London: Macmillan