This is the accepted manuscript version of: Proudfoot D. (2016) Heavenly computation: digital metaphysics and the new theology. *Metaphilosophy* 47(1): 147-153. http://dx.doi.org/10.1111/meta.12171. In citing, please refer to the published version.

CRITICAL NOTICE

HEAVENLY COMPUTATION: DIGITAL METAPHYSICS AND THE NEW THEOLOGY

DIANE PROUDFOOT

Eric Steinhart's *Digital Afterlives* is a thought-provoking contribution to a relatively new and rapidly-expanding genre: digital philosophy. Previous exemplars of the genre include Jim Moor, Hans Moravec, Nick Bostrom, David Chalmers, and Ray Kurzweil. The field marked out is an exciting one, representing a blend of writing about computer science, artificial intelligence, analytic philosophy, philosophy of religion, and sometimes more than a touch of science fiction. Steinhart's book is unique: it employs numerous and intricate arguments to provide a philosophical grounding for claims made by technological futurists (most of whom are computer scientists) and develops an elaborate 'digital theology'. The book unabashedly sets out the parallels between 'digital metaphysics' and traditional afterlife doctrines. In Steinhart's account, digitalism is to succeed where traditional theologies fail—in particular, by answering scientifically-oriented challenges from new atheists. If Steinhart is right, he has shown that core religious beliefs are consistent with a hard-headed materialism.

It must be admitted that many of the claims in this book are radical. Steinhart takes the simulation hypothesis (our universe is 'a software process') and the iterated simulation hypothesis ('we are living at the bottom of an infinitely high stack of computer-generated universes') very seriously (84, 98). The first hypothesis allows that 'uploaded people' may 'own earthly property' or 'be married to earthly people'; the second hypothesis allows that human beings themselves are just increasingly intelligent software and that every human life is the root of 'an endlessly ramified tree of better lives' including 'many infinitely long spans of lives, across many universes' (161). Gods, in Steinhart's account, are living computing machines. The most primitive digital god, Alpha, is 'the first cause of all things', 'the ultimate necessary being', and the 'root of the Great Tree' of successively smarter gods (114). The gods evolve according to the 'Divine Algorithm' and each successive god 'inherits' a 'cosmic script' which it uses to make a better virtual universe (143, 131).

For Steinhart, some of these claims are implied by what he calls the 'axioms' of digitalism, to which digitalists 'have every right' (no further justification of the axioms is offered) (47). Other claims are either elaborations of theories that digitalists 'certainly have the right to pick', or speculations that philosophy 'ought to take … more seriously', he says (21, 173). Steinhart also suggests that the opponents of digitalism are those dire characters, mysterians (8).

The fundamentals of Steinhart's metaphysics are as follows. What we usually think of as *the mind* is 'the sum of the minds of all the cells in the body'; each of these minds is 'a material connect-the-dots network in which the dots are molecular nano-machines and the connections are their functional interactions' (42). *Persons* are bodies, which can be replicated digitally. A biomolecular scanner can produce 'a perfect physiological ghost' of the body, which is also 'psychologically exact' (55, 56). Just as in many religions God's breath reanimates the dead, in Steinhart's account a computer 'reanimates' this ghost (10). This leads to the form of *afterlife* typically hypothesized by futurists (which I shall call the 'entry level' of digital afterlife): the digital ghost of your 'organic' or 'earthly' body is uploaded to a 'terrarium'—a computer system able to simulate a 'terrestrial biosphere' that is 'sufficiently earthlike for you to live a human life' (57, 61). The terrarium, Steinhart says, can 'house an enormous civilization—including every human who has ever lived' (58). This is the 'resurrection universe' (70). Life there is a great improvement on earthly life: the uploaded person 'will be healed in utopia' and 'will be rejuvenated to … a state of optimal youthful functionality' (73).

According to Steinhart, there is also a grander form of digital afterlife, which he calls 'promotion' (or 'ascent') to a 'higher' virtual universe (160). In the entry level of digital afterlife, animating my program produces a digital life which can be regarded as an continuation of my 'earthly' life. In 'ascent', however, my life begins over. My entire life is recorded 'in full biomolecular detail' to produce 'a 4D perfect physiological ghost', which is 'compresse[d]' into a '3D body'; this 'body' is the basis of my new life in another virtual universe (84, 85). As one digital god replaces another, each creating a better universe, this process is repeated, my life beginning anew in each successive universe. Steinhart says, 'You will be reborn, over and over again, in all possible ways, through all possible levels of biological excellence' (166). The gods will produce 'optimized' universes containing humans that are 'functionally superior to all earthly humans', 'idealized' universes containing humans who are 'as good at any task as any optimized carbon-based organism', and 'extended' universes containing humans who are 'saturated with

intelligence' (176, 183, 194). According to Steinhart, the Great Tree even contains gods that will create 'uncountably complex' humans possessing 'hyperperception, hypercognition, and hypervolition' (209). My future, then, is to be reborn as an 'uncountably infinitely powerful' being (210).

Alongside his origin story and soteriology, Steinhart provides digitalist versions of the traditional proofs of God's existence and also takes aim at many standard objections to monotheism (although the view he himself defends is polytheistic). His metaphysics is clothed in scriptural language. The purpose of digital resurrection is 'to glorify the flesh, to spiritualize the flesh, to intensify the visceral excellence of living computation'; and the Great Tree, the 'brilliance' of which is 'glory', grows up 'endlessly into ever brighter light' (69, 126-127). Digital afterlife allows human beings to move 'closer and closer to the Deity' (98). The Deity is an 'infinite self-programming computer' that is 'ultimately real', 'infinitely powerful', 'infinitely intelligent', 'everlasting both into the past and future', and 'generates all physical existence' (95-7). Steinhart's account even includes other heavenly beings: the terrarium—which is 'something like a divine computation in which the uploaded people live, move, and have their being'—is 'a utopia governed by angels', he says (75, 91). Audacious claims indeed, but surely no more so than many of those found in the major religions and debated by philosophers and theologians. Audacious, but consistent with materialism.

Technological futurists typically promise resurrection and even immortality—for Kurzweil, immortality will be available by 2045—but their standard account of persistence generates contradictions (see Proudfoot 2013). Kurzweil and others seem committed to 'patternism', typically the thesis that an organic human being *A* and a future uploaded 'mind file' *B* (to use Kurzweil's term) are the one person if and only if *B and A's brain instantiate the same 'pattern'* (see e.g. Kurzweil 2006). Assuming the transitivity of identity, this account faces the notorious 'duplication problem': if after *A*'s death we simultaneously upload *two* mind files, *B* and *C* (both duplicates of *A*), then *B* and *C* will be one person—even if in different places at the same time and having different experiences. And futurists typically do not forbid duplicates, even claiming that they are required in order to guard against hardware and software glitches. (I have argued that *fuzzy patternism*—the combination of patternism and fuzzy logic—may help futurists here, and also fits what they say about persistence. See Proudfoot 2013.)

In Steinhart's account each 'digital ghost' will be duplicated (and reanimated), although not to guard against glitches but because 'it is best of all to actualize every way to

improve your life' (70). He avoids the duplication problem and other puzzles by jettisoning the notion of identity over time, saying 'Any thing that persists through time changes into, turns into, or becomes some *other* thing at some later time. It *persists into* that other thing' (18-19). As Parfit and others have argued, persistence—in the sense of *survival as*—does not require identity. Steinhart opts for a temporal counterpart analysis of future or past properties: the statement 'After my organic body dies, I will live again' is true if and only if my 'digital ghost' is reanimated, this being my future counterpart. This analysis, however, only exchanges one notorious problem for another, originally posed about modal counterparts by Saul Kripke (1980, p. 45, n. 13): if I am not identical with my temporal counterparts, why should facts about these past or future entities be taken to verify past- or future-tense statements about *me*?

Steinhart's response to Kripke's objection to counterpart theory is that I should resist the desire to be identical to some person existing after my death, since this would condemn my post-mortem self to the miseries of my pre-mortem self. According to Steinhart, 'if you want to *overcome* your death, if you want to *prevail over* your death, if you want to *surpass* it' you must abandon 'your painful desire to stay the same, to remain self-identical' (28). We 'ought to want to' have better versions of our 'damaged earthly lives', he says; only if my life ends can my counterpart improve my life, by having properties that otherwise I would not have had (167). However, this response begs the very question at issue, by assuming that it is *my* life that is thereby improved.

In David Lewis's classic account of modal counterparts, *similarity* is the criterion of counterparthood. Your counterparts in other possible worlds, Lewis said, 'resemble you closely in content and context in important respects. They resemble you more closely than do the other things in their worlds' (1968, 114). If similarity is the criterion of temporal counterparthood, the results of applying it do not support Steinhart's predictions for my future: for example, the 'uncountably complex' person in a universe of 'hyperphysics' who 'can perform 'hypertasks' (compressing 'uncountably many operations into some finite volume of space-time') is hardly a ringer for *me* (209). Moreover, in Lewis's modal realism my counterparts and I are entities of the same kind, but my 'digital ghost' is an *avatar*—merely a representation or simulation of an entity of the same sort as me. Steinhart, though, describes avatars as if they were much more; for example, a digital ghost is 'physical', 'duplicates' the biological activity of the human body, has 'digital flesh', can suffer from bipolar disorder or cystic fibrosis, has 'fully realized senses of taste and smell', can 'enjoy eating' and 'appreciate fine wine', and is 'aroused by the smell of its lover' (56-60, 73). For

the digitalist the applicability of such predicates to computer simulations will likely be entailed by an 'axiom' to which digitalists 'have every right'; but to me these claims look like illicit anthropomorphism.

Moreover, Steinhart's criterion of temporal counterparthood involves *the soul*.[1] This fits with many traditional afterlife hypotheses, save that for Steinhart and other futurists the soul is the 'abstract pattern' of the body—the 'body-program' (50). My 'earthly soul', he says, is the program 'encoded in' my 'present earthly genetics', and the souls of my future counterparts are 'body-programs with better genetics' (232-3). Further, 'Since every body-program contains some mechanisms for self-correction, [my] body-program naturally entails the correction of every genetic defect' (51). On his view, the souls of my future counterparts are already implicit in *my* soul: these future souls are only 'a kind of debugging' of my own body-program (90). However, the fact that a program contains some self-correcting mechanisms hardly entails the correction of *every* defect, and intuitively my 'uncountably complex' counterpart in some future universe is no more a *debugging* of me than my iPhone is a debugging of a string telephone. The digitalist criterion of temporal counterparthood requires considerably more discussion than Steinhart offers in the book.

Steinhart claims that digitalism 'entails the salvation of every possible thing' (169). Futurists typically assume that superhuman-level artificial intelligences will choose universal salvation for human beings—because, for example, these AIs will be interested in acquiring knowledge about the past, or will feel obliged to the humans whose efforts led to the emergence of the AIs, or will simply respond benevolently to the fact that our lives are much poorer than their own. However, as in traditional theisms, this is to second-guess very different beings on the basis of our own, human attitudes—attitudes that may seem primitive and misguided to superhuman minds.

Steinhart has distinctive reasons for predicting universal salvation. These include the *Argument for Virtuous Engineers*, which aims to demonstrate that designers of universes possess 'superhuman benevolence' and will build terraria to be 'heavens rather than hells' (82, 71). According to Steinhart, building such sophisticated technologies requires 'long-term stability and rational purposiveness', and this in turn requires 'social harmony', which also requires 'virtuous individuals virtuously organized'. He concludes

[1] I take Steinhart to be saying that *similarity* of soul is the criterion of temporal counterparthood—just as, for other futurists, it is the criterion of identity of persons over time. Steinhart does, though, also suggest a *causal* relation between me and my counterparts: a causal 'pipeline' links me to future digital counterparts in different universes (13). A causal relation, however, seems inconsistent with his claim that universes are 'causally isolated' (138).

that such designers are 'virtuous persons in a virtuous society' and as such 'will design habitats for the flourishing of other persons' (71). However, this argument could not work in the case of the digital gods, since in Steinhart's account there is no *society* of gods: each god is alone and each universe is 'spatially, temporally, and causally isolated' from the others (138). This feature also undermines Steinhart's claim (if referring to the gods) that the designers of universes might save humans in order to avoid punishment from 'even higher-level' designers if they failed to do so (84).

Steinhart claims too that universal salvation is 'the consequence of the ultimate laws of nature' (169). The Divine Algorithm governs the evolution of the digital gods in such a way as to ensure better lives for human beings—'all the darkness of human life will eventually be turned into light', he says (172). But what reason is there for thinking that this is how the world is *in fact*? Steinhart's ultimate justification for his metaphysics is (what he calls) 'digital axiarchism'—including his 'Axiological Argument', which he says 'follows the pattern of Anselm's Ontological Argument' (214). This part of digital metaphysics receives relatively little examination, and Steinhart concedes that philosophers may find his 'brief sketch' of digital axiarchism 'far too fuzzy or metaphysical to take seriously' (213). His defence is that *every* philosopher must answer the 'terrifying Question' of why there is something rather than nothing—and digitalists, he says, are 'free to explore their own axiarchic answer' (212-3). However, this defence ignores the fact that many philosophers, following Hume and Russell, deny the very intelligibility of the 'terrifying Question'. Digital metaphysics needs more than this if it is to justify the forecast of universal salvation.

Futurists typically seem merely to tailor their afterlife hypotheses in order to fit traditional religious doctrines—or, like the proponents of those doctrines, in order to manage death apprehension. Not so Steinhart's version of the afterlife. In his view, 'intrinsic value is some type of density'; the gods 'accumulate perfection' just by becoming more complex, and *better* universes are just *richer* universes (120, 121, 139). For Steinhart this has the consequence that the 'best' universes contain 'the most suffering' and 'the most evil'; 'Digitalists are not hedonists', he says (140). So the 'utopia' that he promises will surely also contain suffering and evil. That's not my idea of heaven. If Steinhart's metaphysics is correct, perhaps one should hope for extinction—*strong* extinction, in which I have no future counterparts.

Digital theology is a thoroughly modern addition to philosophy of religion and to apologetics. Steinhart claims that his account has advantages for the theist: for example, the

Argument from Evil as traditionally construed fails to arise (since digital theology has no personal omni-God) and digital theism can agree with new atheists that God too must have evolved (since God is simply our 'local' designer) (126, 108). Steinhart can also be seen as offering the theist a novel response to the common view that current naturalistic explanations of religion successfully 'explain away' religion. Theists typically respond by claiming that experimental findings say nothing about the truth-value of supernaturalist beliefs, or that a supernatural deity is the likely source of any evolved disposition to religion. In contrast, Steinhart's theology rejects supernaturalism itself. This is natural theology for the computer age.

Yet is the digital theist not really an *atheist* in futuristic clothing? According to Steinhart, digitalists are 'hardly atheists', since they posit 'an infinite plurality' of gods (127). The traditional theist, however, is likely to reply that digital gods are hardly *gods*. For these theists gods are not 'hardware entities' (132). Steinhart's talk of 'resurrection', too, rewrites ordinary usage. He suggests that his conception of uploading 'very closely resembles' John Hick's replication theory of resurrection (67). While Hick does appeal to the cyberneticist Norbert Wiener's notion of the 'pattern' of an individual, this is only to argue for the possibility of God's creating 'an exact *psycho-physical* "replica" of the deceased' (Hick 1976, 282, 279, emphasis mine). In fact Hick's 'resurrection world' seems to be a *physical* world in the ordinary sense—one where my replica is, as he says, 'a psycho-physical being exactly like the being that I was before death' (ibid., 285). Digitalism is more radical than even the pluralist or non-realist metaphysics of some modern theologians.

Moreover, can the digitalist really approach the digital gods with 'veneration and reverence', as Steinhart claims (229)? Steinhart says, 'If algorithms can be *praised* or *worshipped*, then so can the digital gods' (229). That is a very big 'if'.

Hick, J. (1976) *Death and Eternal Life*. Collins: London.

Kripke, S. (1980) *Naming and Necessity*. Harvard University Press: Cambridge, MA.

Kurzweil, R. (2006) *The Singularity is Near: When Humans Transcend Biology*. New York: Penguin Books.

Lewis, D. K. (1968) Counterpart Theory and Quantified Modal Logic. *The Journal of*

Philosophy 65(5): 203-11.

Proudfoot, D. (2013) Software Immortals—Science or Faith? In A. Eden, J. Søraker, J. Moor, and E. Steinhart (eds) *The Singularity Hypothesis—A Scientific and Philosophical Assessment*, pp. 367-89. Berlin: Springer.

Diane Proudfoot
University of Canterbury
Private Bag 4800
Christchurch 8140
New Zealand
diane.proudfoot@canterbury.ac.nz