

Searle's wager

Neil Levy

Received: 20 November 2010 / Accepted: 9 February 2011 / Published online: 9 March 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Nicholas Agar has recently argued that it would be irrational for future human beings to choose to radically enhance themselves by uploading their minds onto computers. Utilizing Searle's argument that machines cannot think, he claims that uploading might entail death. He grants that Searle's argument is controversial, but he claims, so long as there is a non-zero probability that uploading entails death, uploading is irrational. I argue that Agar's argument, like Pascal's wager on which it is modelled, fails, because the principle that we (or future agents) ought to avoid actions that might entail death is not action guiding. Too many actions fall under its scope for the principle to be plausible. I also argue that the probability that uploading entails death is likely to be lower than Agar recognizes.

Keywords Agar · Searle · Consciousness · Radical enhancement · Uploading · Probability

1 Introduction

In his important new book *Humanity's End* (Agar 2010), Nicholas Agar presents a number of different arguments for the conclusion that we ought to reject radical enhancement

of ourselves. In this paper, I will focus on just one of these arguments. The argument, *Searle's wager*, is designed to show that it would be irrational to follow Ray Kurzweil's (2005) advice and attempt to upload ourselves onto computers; if it is successful, it generalizes to any attempt to replace the organic supervenience base of our consciousness with a non-biological substrate. I have no particular opinion one way or another concerning whether we should embrace radical enhancement, by these means or by any other. My purpose is to focus on this argument alone. I shall show that, like Pascal's wager, on which it is modelled, Searle's wager fails to be action-guiding. Agar's basic strategy is to argue that so long as there is a non-zero possibility that uploading might entail death, it would be irrational to choose uploading. But the doubts to which he appeals cannot be restricted to the choice of uploading or not. Instead, they extend to such a wide class of actions—and omissions—that the strategy cannot be used to argue for any particular action at all.

2 Searle's wager

Agar's aim in arguing against Kurzweil is to establish that it would be irrational 'to accept offers to replace the parts of our brains responsible for thought processes that we consider essential to our conscious experiences, even if the replacements manifestly outperform neurons' (Agar 2010: 65). His claim is that the risks of enhancement are exponentially greater than the benefits, no matter how great the benefits are. Agar's argument is explicitly modelled on Pascal's wager. Just as Pascal's wager—allegedly—establishes the prudential irrationality of atheism, so Searle's wager aims to establish the prudential irrationality of uploading your mind onto a computer.

N. Levy (✉)
Oxford Centre for Neuroethics, Suite 7, Littlegate House,
16/17 St Ebbes St, Oxford OX1 1PT, UK
e-mail: neil.levy@philosophy.ox.ac.uk

N. Levy
Florey Neuroscience Institutes, University of Melbourne,
Parkville, VIC 3010, Australia

It might be helpful to model both wagers in the form of pay-off matrices. Let us begin with Pascal's wager:

Pascal's Wager: pay-off matrix

	Believe	Disbelieve
God exists	Eternal bliss	Eternal torment
God does not exist	Small utility loss	Small utility gain

The important cell for Pascal is the top right-hand one: by failing to believe, we risk damnation. Since the worst we can do by believing is forgo a small gain of utility (by wasting our time on religious activities; bottom left cell), we ought to believe. The wager is powerful because it prescinds from the probabilities. We do not need to calculate the likelihood that God exists to calculate that we ought to believe because no matter what the probability of God's existing, *eternal torment* has a disutility that outweighs the utility gains from failing to believe in a non-existent God: any finite probability multiplied by infinity is infinite.

Searle's wager is so named because Agar builds it on Searle's Chinese room argument (Searle 1980, 1990). The Chinese room argument is supposed to establish that nothing could think unless symbols were intrinsically meaningful for it. Searle claims that because symbols can never be intrinsically meaningful for a merely syntactical device, computers can never think. That is, it is impossible to derive semantics from syntax.

The Chinese room argument is controversial. But Agar need not claim that the argument is certainly correct or even that it is very likely to be correct. Instead, just as Pascal's wager succeeds (if it succeeds) so long as there some non-zero probability that God exists, so (Agar claims) Searle's wager succeeds if there is some non-zero probability that Searle is correct. The Chinese room argument has come in for some heavy criticism, but there is surely *some* non-zero probability that it (or a suitably amended successor) is correct. Assuming this is the case, Searle's wager generates the following pay-off matrix for uploading ourselves into computers:

Searle's Wager: pay-off matrix

	Do not upload	Upload
Survival of consciousness	Forgo benefits of enhancement	Receive benefits of enhancement
Non-survival of consciousness	Status quo	Death

The pay-off matrix is generated by considering our options and their possible consequences. Either we upload

ourselves into computers or we do not. If we do not, then at worst we forgo the benefits of radical enhancement (top left cell): if we had uploaded ourselves and our consciousness survived the procedure, we would have benefited significantly (top right cell). If consciousness would not have survived, then we have had a lucky escape (bottom left cell): we continue in existence, avoiding the death that would have been our fate otherwise (bottom right cell). The important cell, for Agar, is the bottom right cell. By uploading ourselves, we risk death. But the very worst outcome of failing to upload is much less bad: it consists only in forgoing the benefits of radical enhancement.

There is one obvious disanalogy between Pascal's wager and Searle's wager. The worst outcome on Searle's wager is death. Now, while death is, I suppose, bad, it is not *that* bad. As a matter of fact, many people whom we regard as rational judge that there are goods in life worth risking death in order to attain. Some of these goods are extremely weighty ('give me liberty or give me death') but some are relatively trivial. Think of the goods pursued through risky activities—skiing, hiking, and so on. Many people think that it is worth the relatively small, but real, chance of dying in the pursuit of these goods. Even those who do not think these goods worth the risks involved, nevertheless routinely take some risks because doing so improves the quality of their lives. They cross streets, they go into crowded places despite the risks of infection, and so on. No one thinks that death is so bad that it is worth deferring at any cost. For Searle's wager to succeed, however, the disutility of death needs to be very bad. Recall, the argument is supposed to prescind from probabilities. If the worst outcome of uploading is not all that bad, then it does not prescind from probabilities because unlike eternal damnation, we cannot be assured that multiplying the badness of the outcome by its probability will always yield an enormous expected disutility. If death is not all that bad, the argument entails only that we ought to decide whether to upload by attempting to calculate the likelihood of the worst outcome, as well as the benefits of the best outcome.

Agar is well aware of this disanalogy between Searle's wager and Pascal's. Given how things stand *right now*, he concedes, it might be rational to choose uploading. The disutility of death will probably be reasonably high for most of us, but it might be more than offset by the potential gains of uploading. If you are convinced by one of the many different replies to the Chinese room argument that have been put forward over the years, you might calculate that this is a gamble worth taking. But as Agar points out, none of us face the choice of uploading right now. He argues that *for those people who might actually face the choice*, uploading will be highly irrational, because for them the expected disutility of death will be significantly higher than for us and the utility of uploading rather lower.

The disutility of death will be greater because by the time uploading is an option, human beings will be ‘experiencing many of the benefits of the genetic revolution’ (Agar 2010: 74). We might have already reached ‘longevity escape velocity’, which is to say that our expected lifespan will be increasing at a rate faster than one year per year. We might have beaten cancer, heart disease and infection. We will also already be greatly cognitively enhanced, by genetics and by neurophysiological prostheses. Of course, Agar concedes this is all rather speculative. It is certainly possible that we might develop the means of uploading without also developing the genetic, medical and technological enhancements that he envisages. But, he argues, we need not be convinced that these enhancements are imminent or even possible. His argument depends only on the ‘likelihood of their being achieved *relative* to the achievability of uploading’ (Agar 2010: 75). Uploading requires developments in computer science and in neuroscience that seem significantly harder to achieve than the medical and genetic breakthroughs needed for alternative enhancement mechanisms to be available.

If this is correct, then for those people facing the choice of uploading themselves, the disutility of death will likely be very much greater than for us. Given the radical extension of healthspan that enhancements will have brought and the high quality of life with enhanced intelligence, aesthetic sensibilities and what have you, the loss someone would risk by uploading will be much greater than the loss we would risk by making the same choice. Moreover, Agar argues that the benefits of uploading would also be very much smaller for these people than they would be for us. Agar concedes that the benefits might be large; ‘the enhancements compatible with the brain’s survival are likely to be significantly more modest than those enabled by uploading’ (Agar 2010: 76). Nevertheless, he argues that it would not be rational for the already greatly enhanced to trade in their currently enhanced state for a chance of this much more enhanced state. He points out that there is a diminishing return from some kinds of goods. Because of this diminishing return, it might make sense to refuse to trade a high probability of a smaller gain for a smaller probability of a large gain. Agar expresses this point in terms of refusing to trade a smaller gain for a larger with expected utility held constant, but the point can be strengthened: even if the expected utility of the larger reward is higher than the expected utility of the smaller, it might be rational to refuse to trade. I would prefer a 100% chance of \$1,000,000 to a 1% chance at \$150,000,000, and I do not think my preference is irrational. Agar argues that the subjective value, for the person facing the choice of uploading, is likely to be significantly smaller than its objective magnitude.

Why should this be true? After all, cognitive enhancement does not involve a good like money, which might

diminish in value because it is a positional good. Enhancements might be intrinsically valuable and therefore not subject to diminishing marginal returns or subject to diminishment at a much slower rate. Agar claims that, nevertheless, it is likely that people will not desire the more radical enhancements all that strongly. This is because ‘we have comparatively few desires that correspond specifically’ with the goods such enhancements might make available (Agar 2010: 77). In our unenhanced state, we cannot even fully understand the desires we would have were we radically enhanced. We have to make our choice from here, where there is an enormous gap between what we can really grasp and the goods that radical enhancement might make available to us. Agar notes that for a billionaire, the preference for a 100% chance of a further \$1,000,000 rather than a 1% chance at \$100,000,000 might be irrational, because the billionaire has desires that can be satisfied only with the larger sum. But the fact that we know that there is a standpoint from which our preferences would reverse is irrelevant to what it is rational to prefer from our current perspective. Our current desires and goals are ‘contingent on our current levels of cognitive powers’ (Agar 2010: 70). We want to protect and promote our *current* relationships and our *current* moral and political ideals. Were we radically enhanced, we might no longer have these desires, so choosing radical enhancement would be making a choice which would not promote what is most important to us right now.

3 Assessing Searle’s wager

I turn now to assessing Agar’s argument against radical enhancement. I shall argue that there are a number of considerations that suggest that Agar’s confidence in the strength of the central contention—that uploading entails death—is misplaced. The likelihood that uploading entails death is more remote than he thinks, because it is not just one controversial philosophical issue—the success of the Chinese room argument—that has to be resolved in his favour for uploading to entail death. Instead, a whole host of philosophical issues has to come down in his favour: the right account of personal identity and its cessation must be true, and the right account of the badness of death must be true, and so on. I concede, however, that there is a non-zero chance that all these issues might come down in the way Agar requires. Nevertheless, I will argue that Agar is wrong in thinking that any positive probability of uploading entailing death is sufficient to show that uploading is irrational and that therefore the fact that the probability is lower than he recognizes is directly relevant to the assessment of the rationality of uploading. My strategy for showing that a non-zero probability of entailing death

cannot suffice to show that uploading is irrational will consist in showing that far too many actions, and for that matter, omissions have a non-zero probability of entailing death and that as a consequence the principle to which Agar appeals cannot be action guiding.

In making his case for the irrationality of choosing uploading, Agar oscillates between two different perspectives: the perspective of those who face the choice of uploading sometime in the future (at $t1$), and our current perspective (at t). In arguing that the disutility of death, for the purposes of Searle's wager, is very high, Agar argues that we ought to adopt the perspective of the highly enhanced. For them (though not for us), it would be irrational to run the risks of uploading, because for them (though not for us) death has been defeated or at any rate very greatly deferred (as have the infirmities of the flesh). But in arguing that the benefits of uploading would be small, he adopts our current perspective. It would be irrational for us (though not for them) to choose to upload ourselves because we (though not them) have few desires that uploading would satisfy. Now, this is not a straight out inconsistency, since the perspectives affect different cells of Searle's wager (top right and bottom right, respectively). However, it does entail that the task of showing that uploading is rational is easier than Agar seems to think. When we consider the problem holding our temporal perspective fixed, either the disutility of death is great or the utility of uploading is relatively small, but never both at once. Hence closing the gap—demonstrating that the expected benefits of uploading outweigh the risks—will be easier than Agar thinks. Those who have the most to lose also have the most to gain, and those with the least to gain also have the least to lose.

It seems to follow at once that Searle's wager is not analogous to Pascal's. There is no temporal perspective from which the gap between expected benefit and expected cost is so great that the probabilities cannot matter. In other words, in deciding how to act, we have two different payoff matrixes to consider. Let us, therefore, address the two temporal perspectives independently.

4 Our current choice

As Agar notes, right now, we do not face the choice of uploading or not. But we do face a choice: how are we to invest our scarce resources? Right now, the costs of death are not so great that uploading is obviously irrational for us, but Agar argues, since the benefits of uploading are relatively small right now, it is rational for us to invest our resources in other ways. Because we have few desires now that would be satisfied by radical enhancement, we do better to invest in less radical means of enhancing

ourselves. As a way of avoiding assessing the probabilities, I think this argument fails. Notice that we commonly believe that it can be rational to invest in future goods, even when we fail to desire those future goods at the time of the investment. It is sufficient that we believe that there is a high (enough) probability that we will come to desire those goods. Thus, for instance, we believe that it is rational to spend time and effort to acquire skills in order to compete in the job market, even though we might lack the desire to compete in the job market at the time of the investment. Children are often forced by their parents to spend time doing things they have no desire to do and which will enable them to do other things which, again, they have no current desire to do, and most of us think that these actions are permissible and quite likely even obligatory. The fact that the children can be expected to become adults who would *then* regret their wasted opportunities suffices to rationalize the choices made on their behalf.

If, therefore, there is a reasonable probability that we will become people with desires which can be satisfied only by uploading, then it is an open question whether we ought to direct a proportion of our research dollars into uploading research. *Pace* Agar, whether we should or should not do so ought to be assessed on the balance of probabilities. We need to know (a) how likely it is that we will become agents with desires that can only be satisfied by uploading; (b) how likely it is we will survive uploading; (c) what we forgo by investing in uploading rather than alternatives; and (d) is it really the case that we now have few or no desires that would be satisfied by radical uploading?

Because settling the question whether we ought now to invest in research that might lead to uploading in the future requires assigning a probability to (b) the questions we ask ourselves *now* are not independent of the questions that future agents will confront (though not vice versa, if future agents confront the question whether to upload or not, the question how we ought to act now will be irrelevant to their considerations). In claiming that there are two questions we need to ask, I am not claiming that the questions are not related to one another; my claim, rather, is that there are some considerations that are relevant to only one of the options and not the other and that Agar illegitimately mixes together these considerations. The fact, *if* it is a fact, that we now have few desires that can be satisfied by uploading does not entail that it is highly improbable that we ought to invest in uploading. Depending on the answers to questions (a) to (d), it might be rational to invest in uploading research.

5 The choice facing the already highly enhanced

Agar believes that *independently* of the question of what we ought to do now, the options facing the already

enhanced are sufficiently similar to those figuring in Pascal's wager to rule out the possibility that it might be rational to choose uploading. Though he concedes that it is an open question whether the Chinese room argument succeeds, a vanishingly small probability of Searle being right is sufficient to establish the irrationality of uploading. That is, the costs of uploading if Searle is right are so high that even a tiny probability of his being right is sufficient to establish the irrationality of uploading: 'if there is room for rational disagreement you should not treat the probability of Searle being correct as zero...This is all that the Wager requires' (Agar 2010: 71). There are many replies to the Chinese room argument extant (the virtual minds reply, the systems reply, the robot reply, and more; see Cole 2009 for review); many people, including me, find one or more of these replies persuasive. But given that there is a non-zero chance that Searle might be right (or that some other argument for the same conclusion might be right), it would be irrational to upload because the costs would be so high.

Before assessing the choice facing the already highly enhanced, I will set out some complications that Agar does not recognize. These issues complicate matters, but do not fundamentally alter the choice facing the enhanced. However, as we will see, these complications make our task much easier *if* (as I shall go on to argue) Agar's claim that we can prescind from the probabilities turns out to be false.

First, we need to recognize that even if it is successful, the Chinese room argument does not obviously entail that uploading equals death. Agar interprets the Chinese room argument as an argument about consciousness. In fact, Searle seems to be arguing for two different, though related, claims: the first of which was more prominent in earlier formulations of the argument and the second of which subsequently came to the fore. The first is about *understanding*: Searle's claim is that you cannot get semantics from syntax (Cole 2009). The second is about *consciousness*. It is this second issue which is Agar's focus, but both must be considered.

On one view, then, there is a non-zero probability of an upload failing to be conscious; on another, there is a non-zero probability of an upload failing to understand its own informational states. On neither of these views, however, is there precisely the same probability that uploading = death. That depends upon what account of personal identity and of its cessation is correct. On many views—a suitably modified organism view, the closest continuer view, and others—either scenario is compatible with the continuation of the uploaded person's existence. For uploading to entail death, a psychological account of personal identity must be correct; moreover, it will need to be a psychological account of just the right kind. My zombie continuer would be me, at least on some psychological accounts of personal identity. Still, I suppose that there is a

non-zero probability that Searle's claim is correct *and* that the right account of personal identity is correct. Further, there seems to be a non-zero probability that even if the loss of consciousness is consistent with survival, its loss would sufficiently impoverish life as to make it as bad as death (see Siewert (1998) for relevant discussion). Though the issues are trickier than Agar recognizes, I think we ought to concede that these complications do not by themselves render the analogy between Searle's Wager and Pascal's invalid.

Other complications can be dealt with in similar ways. Agar is apparently committed to a particular account of the badness of death. Some philosophers have rejected the claim that death is a harm at all. For Agar's claim that uploading risks a terrible fate to be correct, it is not sufficient that they be wrong. In addition, the right account of the badness of death must be a forgone good account, such as that defended by Kamm (1998). Only if the right account of death is a foregone good account, can Agar's claim that death would be terribly bad for those facing the choice of uploading be true. Since he holds that death is so bad for the highly enhanced because of the quality of the life they forego, he commits himself to this claim. Again, though, while this greatly complicates the issues, it does not entail that the analogy between Searle's wager and Pascal's wager is invalid. Though many things must come down Agar's way for uploading to entail death—it must cause the cessation of consciousness or of understanding; this cessation must entail the death of the person, and the right account of death must be a foregone goods account—it is plausible that there is a non-zero probability that all of this can go his way. If, however, Agar turns out to be wrong in claiming that we can avoid the assessment of probabilities, the fact that so many controversial issues need to be resolved in just the right way will be bad news for him, raising the probability that uploading is rational.

Agar's argument that uploading is irrational depends on the claim that we can never be sure that uploading \neq death. He notes that given that uploads will be very good information processors, they will pass any Turning-type test, but argues that this will not be sufficient to reduce the probability that uploading = death to zero. No matter how plausible the claims of an uploaded person that it is conscious, no matter how much it insists on being subject to qualia, there is some probability that it is an entirely unconscious information processor. The problem for Agar here is that sceptical doubts like this one come cheap. They are easy to generate and hard to dispel—so hard that it is plausible to think that there is a non-zero (though perhaps small) probability of many of these doubts being true. Consider Agar's own preferred alternative to uploading. Because the mind is modular, he claims, we should enhance ourselves using (inter alia) neuroprostheses that

replace modular functions while leaving central processes unaffected. Since modules operate unconsciously, utilizing these prostheses will not threaten us with death, he suggests. But there is a non-zero probability that Agar is wrong. After all, the modularity thesis is controversial; moreover, even if it is true, it might be that Fodorian modules (somehow) support consciousness. Indeed, given that (a) it is controversial whether modules are neurologically localized—they might be spread out across the brain—and given also that (b) brain regions play multiple roles, it seems reasonably likely that in replacing or supplementing brain-based modules with neuroprostheses, we risk losing the basis of consciousness. *Mutatis mutandis*, the same seems to be true of understanding. Hence, it seems that it is not just uploading that falls within the scope of Searle's wager: we should not be enhancing ourselves by this method either.

It follows immediately from these considerations that Agar will have a harder time than he realizes in establishing that risking death through radical enhancement is so terrible. Its badness, recall, rests on the quality of the life foregone by the person, but if we dare not enhance ourselves with neuroprostheses, then the quality of life that is risked is somewhat lower. If neuroprostheses of the kind Agar regards as acceptable are too risky, then the choice facing future agents will more closely resemble the choice we would face, were uploading available now, and as Agar concedes, choosing uploading in those circumstances may not be irrational. But worse is to come: it is possible to generate sceptical doubts for *all* the enhancements on Agar's list. Take genetic enhancements. Given the uncertainties which characterize the current state of our knowledge concerning consciousness, we cannot assign a probability of zero to the proposition that genetic enhancement will cause the cessation of consciousness (or of some other property that is necessary for personal identity or that makes life worth living). Indeed, Agar's own speculation regarding how uploading ourselves might cause the cessation of consciousness seems to apply straightforwardly to genetic enhancements. Agar argues, as we have seen, that non-modular central processes are the seat of consciousness. He points out that these central processes are commonly taken to be much slower than modular processes. Might not it be, he asks, that consciousness actually depends on this slowness of processing? In that case, a gain in speed might imperil consciousness. Perhaps, who really knows beyond all doubt? In that case, however, genetic enhancements imperil consciousness. Who knows how close we are to the speed limit right now: perhaps, a small increment in speed will spell perpetual darkness.

So, no genetic enhancements for us. Searle's wager applies more widely than Agar wants it to. Agar is no bioconservative, yet the conclusion we seem to be heading for

is a bioconservative one: no enhancements, lest we lose everything. Actually, though, the scope of Searle's wager is broader than even the bioconservative could want: it rules out non-technological enhancements as much as technological. Perhaps increases in IQ are incompatible with the continuation of consciousness. Given the continuation of the Flynn effect—the gradual rise in IQs across time—perhaps we are in danger right now. The explanation for the Flynn effect is still controversial. One theory is that the rise is due to better nutrition (Mingroni 2004). We had better cut back on our fruit and vegetables, just to be on the safe side. Another explanation of the Flynn effect is that it is due to better and longer education (Neisser 1997). Flynn himself thinks the cause is modernization, with the increasing complexity of daily life it brings (Dickens and Flynn 2002). Again, we had better take steps to simplify and stupefy life.

It might be objected that we need not take these dramatic steps. The Flynn effect is about rising average IQs, and *you*, you might think, have had a good education, good nutrition and a complex environment, thereby benefiting from the cognitive advantages these things bring, and have nevertheless preserved your consciousness. We can therefore extend these benefits more widely without fear. All we need do is to ensure that we do not raise anyone above the level that you enjoy. There are several problems with this suggestion. First, you cannot be sure—you cannot assign a probability of zero to the hypothesis—that idiosyncratic factors in your environment preserved your consciousness or that bringing everyone up a little further will not (say) make the environment too complex for the preservation of consciousness. For all you know, widespread zombification is already occurring. Perhaps, you are the only conscious person for miles around. Second, and more radically, are you *sure* you are conscious?

Consider the following argument:

1. It is possible, in some sense of 'possible', that we might each have a zombie twin: a functional duplicate of ourselves (which might or might not also be a physical duplicate) that lacks phenomenal consciousness.
2. Your zombie twin would believe (falsely) that it is conscious.
3. Since, by hypothesis, your zombie twin has the same beliefs as you, your belief that you are conscious cannot be caused by your being conscious (for if it were, your zombie twin would not be a functional duplicate of you).
4. It follows that you would believe you were conscious whether you were conscious or not.

Conclusion (1) your belief that you are conscious is not justified. Hence;

Conclusion (2) you do not know that you are conscious.

Now, this argument may be wrong. There are numerous ways of replying to it. You might deny that zombies are possible. Or, like Chalmers (2003), you might deny that our zombie twins have the same phenomenal beliefs as ourselves. Be that as it may, surely there is some non-zero chance that the argument (or some suitably modified version of it) is correct or perhaps that though our zombie twins would lack our phenomenal beliefs, this difference—since it cannot be introspected by zombies—cannot be appealed to in order to establish even for each of us that we are conscious (for were we zombies, it would seem to us, in some sense of ‘seem’, that we were appealing to it to rule out our being zombies). Perhaps it is already too late, or perhaps it is not too late yet, but stimulate your brain any further—say, by reading just one more word ... Oops.

Of course, it ought to be pointed out that Agar thinks we are not now in the position of the person in Searle’s wager, because by becoming zombies we would lose much less than would the radically enhanced. But he might be wrong! Perhaps becoming a zombie is losing *everything* (though the value of consciousness is hard to pin down, there is some intuitive appeal to the idea that it is of supreme value or that it enables everything that is of value, such that in losing it we lose everything).

6 Conclusions

I conclude that Searle’s wager fails. We cannot argue from the fact that there is a non-zero chance that radically enhancing ourselves would entail death to the conclusion that we ought not to enhance ourselves. That argument extends too broadly. Like the precautionary principle, whose problems it mirrors, it entails that we ought never to do anything to ourselves. But like the precautionary principle, it also seems to entail that we risk too much in not taking action either. It therefore fails to guide action at all.

There is no argument against radical enhancement that prescind from the probabilities. Deciding whether we ought to upload ourselves therefore requires that we assess the risks and benefits of doing so. I have suggested that the

likelihood of uploading entailing death—or some other equally bad fate—is likely to be relatively low, because so many controversial issues have to be settled in the right way for the conclusion to follow. It may nevertheless be the case that we ought neither to upload ourselves nor to invest in technologies which might make uploading a future possibility. I remain agnostic on that question here. I want only to insist that this is a question that requires ordinary cost-benefit analysis to answer.

Acknowledgments I would like to thank Nick Agar for his helpful comments on this paper. I am also grateful for the support of the Wellcome Trust.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Agar N (2010) *Humanity’s end: why we should reject radical enhancement*. The MIT Press, Cambridge
- Chalmers D (2003) The content and epistemology of phenomenal belief. In: Smith Q, Jovic A (eds) *Consciousness: new philosophical perspectives*. Oxford University Press, Oxford, pp 220–272
- Cole D (2009) The Chinese room argument. Stanford encyclopedia of philosophy. <http://plato.stanford.edu/entries/chinese-room/>
- Dickens WT, Flynn JR (2002) The IQ paradox is still resolved. *Psychol Rev* 109(4):764–771
- Kamm FM (1998) *Mortality mortality*, vol 1. Oxford University Press, Oxford
- Kurzweil R (2005) *The singularity is near: when humans transcend biology*. Penguin, London
- Mingroni MA (2004) The secular rise in IQ: giving heterosis a closer look. *Intelligence* 32:65–83
- Neisser U (1997) Rising scores on intelligence tests. *Am Sci* 85(5):440–447
- Searle J (1980) Minds, brains and programs. *Behav Brain Sci* 3(3):417–457
- Searle J (1990) Is the brain’s mind a computer program? *Sci Am* 262:26–31
- Siewert C (1998) *The significance of consciousness*. Princeton University Press, Princeton