

# Evaluating the Posthuman Future – Some Philosophical Problems

---

SØREN HOLM

Centre for Social Ethics and Policy, School of Law, University of Manchester, Manchester M13 9PL, UK; Center for Medical Ethics, HELSAM, University of Oslo, Oslo, Norway; Institute for Health Science and Technologies, University of Aalborg, Aalborg, Denmark.  
Email: [soren.holm@manchester.ac.uk](mailto:soren.holm@manchester.ac.uk)

Imagining a future scenario where human beings have evolved in ways so that they are no longer human but post- or transhuman has been a recurrent trope in science fiction literature since the very inception of the genre. More recently, the possibility of a future including posthumans has received significant philosophical attention due to the emergence of activist ‘transhumanism’. This paper will analyse some of the philosophical problems in evaluating whether a posthuman future is a good future that we ought to pursue. It will first briefly describe the transhumanist conception of the posthuman, and the different routes envisaged from the current human condition to the future posthuman condition. The second part will then present and analyse some fundamental philosophical problems we encounter when we try to assess whether and to what extent the posthuman future is good and/or desirable; and it will be concluded that assessing the ethical desirability of the posthuman future is close to impossible.

## Introduction

Imagining a future scenario where human beings have evolved in ways so that they are no longer human but post- or transhuman<sup>1</sup> has been a recurrent trope in science fiction literature since the very inception of the genre. In the early SciFi novel *The Time Machine* by H.G. Wells, the protagonist travelling from our time to the far future finds himself on an earth where humanity has split into two distinct species: the peaceful, but weak Eloi, and the strong, but brutal underground-dwelling Morlocks. Both are descendants of humans, but have followed different evolutionary paths.<sup>2</sup> In other SciFi novels the transition from human to something else is not driven by evolution, but caused by alien influences or deliberate technological intervention.<sup>3,4</sup> In many of the SciFi novels envisioning a posthuman future this future is dystopic and it is used as a foil for reflection on current social conditions, but there are also

some novels that reflect a more utopic view of the posthuman future, for instance by seeing it as a necessary step in mankind's evolution towards even better futures.<sup>4</sup>

More recently, the possibility of a future including posthumans has received significant philosophical attention due to the emergence of 'transhumanism', a loose coalition of philosophers, scientists, futurists and activists advocating for moving towards a posthuman future (see for instance the websites for Humanity+<sup>5</sup> and the Singularity University<sup>6</sup> or the recent very comprehensive 'Transhumanist reader'<sup>7</sup>). Transhumanists either see the posthuman future as a positive development in itself, or as the only way to avoid the extinction of intelligent life.

This paper will analyse some of the philosophical problems in evaluating whether a posthuman future is a good future that we ought to pursue. It will first briefly describe the transhumanist conception of the posthuman, and the different routes envisaged from the current human condition to the future posthuman condition. The second part will then consider some fundamental problems encountered when we try to assess whether and to what extent the posthuman future is good and/or desirable.

In the debates about the ethics of possible posthuman futures, considerations of rights and social justice have played a considerable role, e.g. considerations of whether the rights of any remaining humans would or should be protected.<sup>8,9</sup> These are important questions but they fall outside the scope of this paper.

### **Paths to the Posthuman Future**

The prominent Oxford philosopher and transhumanist Nick Bostrom defines the posthuman condition in the following way. We will have reached the posthuman state when any one of the following criteria has been reached:

- Population greater than 1 trillion persons.
- Life expectancy greater than 500 years.
- Large fraction of the population has cognitive capacities more than two standard deviations above the current human maximum.
- Near-complete control over the sensory input, for the majority of people for most of the time.
- Human psychological suffering becomes a rare occurrence.
- Any change of magnitude or profundity comparable to that of one of the above (Ref. 10, pp. 63–64).

There are several possible ways to distinguish between the two terms 'posthuman' and 'transhuman', and the literature is not uniform in its use of the terms. The term posthuman strictly speaking only connotes individuals that are in some way descendants of humans, but are no longer human, whereas transhuman connotes individuals who possess abilities that are beyond human abilities, but a transhuman might still be recognisably human. Most transhumanists believe that posthumans are necessarily enhanced compared with current humans. But, *pace* Bostrom, we can easily imagine posthumans such as Wells' Eloi and Morlocks, that are not enhanced versions of humans, but somewhat degraded versions.

There are several possible paths to the posthuman future.<sup>7</sup> One path is through gradual modification of human biology and possible integration of non-biological technologies to create genetically modified or cyborg descendants.<sup>7,11</sup> A second more radical path is giving up on the human body and uploading our minds into supercomputers.<sup>12,13</sup> And a third and even more radical path is through ‘The Singularity’. The Singularity is a hypothesised future event where the speed of technological development has reached a point that creates a radical break between the past and the future. The Singularity was formally proposed by the futurist Vernor Vinge in the 1980s in relation to developments of machine intelligence, and the idea has since been developed and popularised by Ray Kurzweil.<sup>14–17</sup> According to Vinge,

We will soon create intelligences greater than our own. When this happens, human history will have reached a kind of singularity, an intellectual transition as impenetrable as the knotted space-time at the center of a black hole, and the world will pass far beyond our understanding. This singularity, I believe, already haunts a number of science-fiction writers. It makes realistic extrapolation to an interstellar future impossible. To write a story set more than a century hence, one needs a nuclear war in between ... so that the world remains intelligible.<sup>14</sup>

When developed in the transhumanist literature all of the three paths usually also involve the idea that the spread of intelligent life beyond Earth will happen as part of the transition to the posthuman future.

### The Comparative Problem

Transhumanists positively advocate that we should pursue a posthuman future both by removing legal and practical restrictions that make it more difficult to pursue one of the paths to the posthuman condition, and by actively promoting the development, for instance through research funding into increased longevity, cognitive enhancement or the many other technologies that could play a role in helping us to reach the posthuman state. They advocate this because they think, and argue, that the posthuman state is individually preferable to the human state, and that the overall future posthuman state of the world is preferable to the current human state of the world. This means that there are two ways in which the posthuman is claimed to be better: (1) at the individual level, and (2) in aggregate.

How can we go about assessing these two claims? Let us first note that we are analysing arguments that are primarily conducted within a broadly consequentialist frame. The claim is not that posthuman entities are more worthy of respect in a Kantian sense, or more likely to display Aristotelian virtues. The claim is that individual posthumans will have better lives in terms of welfare or preference satisfaction when compared with humans living now. Given the consequentialist framing it is also clear that an answer to the individual question is prior to the answer to the aggregation question. The aggregate welfare in a given state of the world will be a, possibly complex, aggregative function of the welfare of the individuals existing in that world.

There are well known problems in both intra- and inter-individual comparisons of welfare, but let us for the sake of argument accept that we can make both intra- and

inter-individual comparisons of welfare between human beings. Can we make the same comparisons between human and posthuman individuals?

This question may initially seem to have a straightforward positive answer. When we make these comparative judgements now we look at various aspects of an individual's life, estimate how well they go and possibly compare this to how well they could go, and then reach a more global estimation of that individual's level of welfare. In the case of inter-individual comparisons we do the same for the other individual involved, and then reach a judgement about who has most welfare. Can't we just do the same when comparing human and posthumans? If a posthuman has, say a radically enhanced appreciation of music or the culinary arts compared with current humans we can say that in so far as the posthuman is able to pursue music etc. he or she<sup>18</sup> has more welfare than current humans because of the greater ability to appreciate these art forms. So, what we are doing is simply using the same method we would use in evaluating welfare in humans. And since posthumans are supposed to have enhanced capabilities it follows, almost by definition, that they must have more welfare and that the posthuman condition should be promoted. Savulescu, for instance argues that:

... we have a moral obligation to enhance human beings. It is argued that if one is committed to the moral obligation to treat and prevent disease, one is also committed to genetic and other enhancement in so far as this promotes human well-being. It is argued that this is not eugenic but expresses our fundamental human nature: to make rational decisions and to try to improve ourselves. To be human is to strive to be better. (Ref. 19, p. 36)

There is, however, a problem with this answer to the comparative question when it is applied, not to single enhancements, but to posthumans that are very different from current humans in a wide range of capabilities. Let us first consider a potentially analogous question, the inter-species comparison of welfare between lions and wildebeest. We are able to make intra-species comparisons, i.e. estimating with some degree of accuracy whether a particular lion or wildebeest is doing better or worse than some other lion or wildebeest. But there seems to be no obvious way of making inter-species comparisons between lions and wildebeest when both have normal levels of welfare for their species. Even if we can make plausible comparative welfare claims when we compare a happy and well-fed lion to a miserable and starving wildebeest, this does not entail that we can make comparative claims between happy lions and happy wildebeest. It might be objected that we can make other inter-species comparisons of welfare, e.g. between lions and mosquitos, where we would be fairly certain that a non-suffering lion had a higher level of welfare than even the best-fed mosquito. This is true, but the reason it is true is that most of us do not tend to think that mosquitos and other insects have a welfare in a subjective sense because their nervous systems are not sufficiently developed to sustain an inner life.

So, if posthumans are as different from us as lions are from wildebeest, there seem to be a problem about inter-individual comparisons of welfare. But perhaps that problem can be solved by imaginative projection. In order to assess the welfare of posthumans I imaginatively project myself into their condition and perform the assessment that way. But this is problematic for at least two distinct reasons. The first problem is the problem

identified and succinctly stated by Wittgenstein in *Philosophical Investigations*: ‘If a lion could talk, we wouldn’t be able to understand it’ (Ref. 20, p. 327). There are ways of life that are so different from ours that it is plausible that we cannot understand them, even if they are explained to us. The second problem arises if we reject the Wittgensteinian idea that imaginative projection may not be possible. We can only imaginatively project ourselves from where we are, and our projection will be shaped by extrapolations from what we (believe we) know about human nature and human welfare. Even with the most imaginative approach conceivable, we have to start from somewhere in our imaginative projection. We have to start from our implicit or explicit philosophical anthropology.<sup>21</sup>

Buchanan denies this:

It might be replied that the idea of human nature is nonetheless needed to flesh out an adequate conception of our well-being, to distinguish it from the well-being of other animals. For example, it could be argued that our well-being requires that we have the capacity for complex and self-conscious forms of sociability, for a degree of autonomy (the capacity to lead a life, not just to live), etc. All of that is no doubt true, but it can be said without invoking the idea of human nature, where this means a set of characteristics that is universal in all humans and unique to them. Presumably if there are intelligent extra-terrestrials, then the capacity for complex, self-conscious forms of sociability and for autonomy will be important for their well-being, too, even if they were not human beings. If we say that certain capacities that we believe (perhaps wrongly) are peculiar to human nature are important for our well-being, all the normative work is being done by the idea that they are important for well-being, not by the claim that they are part of our nature. (Ref. 22, p. 149)

Let us first note, *pace* Buchanan, that the idea of a ‘human nature’ or a worked-out philosophical anthropology does not necessarily mean a commitment to a claim that there is ‘a set of characteristics that is universal in all humans and unique to them’. Claims to universality and uniqueness are not an indispensable or necessary part of the idea of a ‘human nature’ that is at stake here. There is nothing which in principle prevents two biologically distinct species from having the same nature, if by nature we are meaning those features of our being that form the basis for an adequate conception of our well-being. The ‘nature’ that is at stake here is not a Platonic or Aristotelian essential nature that makes the thing the thing it is. In a *Planet of the Apes* scenario, where apes have acquired human-like cognitive abilities, human and ape natures are identical or near identical in the sense of nature as a base for assessments of welfare that is at stake here.<sup>23</sup> More importantly, Buchanan overlooks or ignores an important issue about the weighing of important interests. It is not enough to know that ‘the capacity for complex, self-conscious forms of sociability and for autonomy’ is important for the well-being of all cognitively advanced beings. In order to compare well-being we also need to be able to say how important these features are in isolation and combined, and in combination with the many other welfare-affecting features of a cognitively complex being. To use an animal analogy again, all mammals have complex forms of sociability, but how important this particular feature is in determining the overall welfare of a particular kind of mammal varies widely according to the particular form of life it leads. Lions and tigers are both big cats, but they lead very different lives with very different forms of sociability.

The second comparative problem is a problem of aggregation. Let us, for the sake of argument, assume that the well-known objections about the distinctness of persons raised against consequentialist aggregation can be overcome,<sup>24,25</sup> and that the goodness of a state of affairs can be expressed as a, potentially very complex, aggregation of the welfare score of individuals. In the posthuman future there will be many more different kinds of self-conscious individuals to take into account than there are now. There will not only be one kind of posthuman, but many radically different kinds. This means that our comparisons of welfare have to be scalable in order to perform the aggregation. We not only need to be able to say whether a particular kind of posthuman individual has more or less welfare than a human, we need to be able to say how much more or less welfare. This obviously complicates the comparative problem discussed above quite considerably. The same problem affects consequentialist aggregation that tries to take account of both human and animal welfare. This problem is usually ‘solved’ by restricting the scope of the aggregation to humans, but a solution by scope restriction will not work in relation to the posthuman problem since we have no plausible reasons to exclude humans or posthumans from the aggregation. And there is a further problem about numbers. Bostrom’s definition of posthuman quoted above illustrates that some of the projected posthuman scenarios operate with vastly increased population sizes, e.g. ‘greater than 1 trillion persons’. But, are we really sure how to handle aggregation at such a scale? Let us just consider a scenario where this population growth is the only thing that has happened, and that it has come about by copying the earth with all its current inhabitants 135 times. Is this new posthuman state of the world 135 times better (simple aggregation), 11.6 times better (aggregation by square root function) or not better at all, perhaps even worse? It is difficult to see how to get a convincing answer to this question. We do of course have theoretical answers, a classical hedonistic utilitarian would have to give the answer 135 times, but how certain can we be that the theoretical answers are right or even sensible?

### **The Singularity Problem**

The Singularity poses a particular problem when we are trying to evaluate whether we should pursue a posthuman future or not. One thing that, despite their other differences, unites the academics who predict the imminent advent of the Singularity is that the Singularity constitutes a radical break with the past, a break that is so radical that it is impossible to predict, analyse or talk about what comes after the Singularity. The Singularity is an impenetrable epistemic barrier. But we are told by some of the Singularity theorists that we should promote the technologies involved in reaching the Singularity and remove any obstacles in its way.<sup>16,17</sup>

This implies that what comes after the Singularity must be good, and better than what we have now, since we would not otherwise have reasons to hasten its coming.

But this lands the transhumanist advocate of the Singularity with a dilemma. Either the Singularity is epistemically opaque, we can’t say anything about whether the state of the world after the Singularity is good in itself or better than some other state, and we can give no reasons to promote it; or, the Singularity is epistemically

transparent, we can make the evaluations, and we may have reason to promote the Singularity. What the transhumanist cannot have is to have it both ways.

From the theorising about the Singularity it is evident that it is theorised as a fundamental and radical break with everything in the past, so the transhumanist is likely to have to choose the first horn of the dilemma. The Singularity is such a radical event that we cannot say whether it should be promoted or not, because we are epistemically blocked from any information that could form the basis for an evaluation of the goodness of the post-Singularity state.

In many ways this issue is similar to an issue faced by religious eschatologies that speak prophetically about a new and radically different future. To be attractive something has to be said about that future, e.g. a Christian eschatology has to say something about the conditions that the blessed will experience in heaven, but it cannot say too much or be too specific because the break with the past is prophesied to be, and has to be described as radical.<sup>26,27</sup>

So how are we to assess the Singularity in its radical otherness? One way is to say, we know what we have got, but not what we will get, so let us stick with what we have got. This way of arguing can, in some cases rightly, be accused of relying on ‘status quo bias’, i.e. the set of cognitive biases that lead decision makers to overestimate the value of their present state and underestimate the possibility that a change will be beneficial.<sup>28</sup> But the status quo bias objection does not apply to this particular choice situation if it is really true that nothing can be said about the state after the Singularity. We can neither underestimate, nor overestimate the value of the Singularity for the very simple reason that we cannot estimate it at all. And in that case it is not irrational to stick to what we have got.

Many theorists claim that the Singularity is inevitable. It will happen because it is produced by scientific and technological developments that are pursued for other reasons, such as artificial intelligence, cognitive enhancement or radical life extension. If we accept this inevitability we can still ask whether we ought to promote the Singularity, i.e. work to make it come earlier than it would otherwise have come, or whether we should try to slow down the developments leading to the Singularity or try to affect its outcomes. Vinge, for instance, presents some possible problematic outcomes of the Singularity in his 1993 paper, and discusses way they can possibly be prevented.<sup>15</sup> But these are, again, questions that cannot be answered if the Singularity really constitutes an epistemic barrier. If we can get no idea what comes after the Singularity, then we cannot try to affect these unforeseeable outcomes.

We are therefore at an impasse. If the Singularity happens it will be an event of a truly radical character, possibly as important in the grand scheme of things as the first emergence of life on earth, and this seems to indicate that it is an event that should play some action guiding role. It should in some way play a role in how we orient ourselves towards the future. But it cannot, because it is theorised to be so radical that it can play no role in rational decision making. Any arm of the decision tree terminating in the Singularity cannot be valued and could potentially have a value between plus or minus infinity. Any consideration of whether the Singularity should be promoted is therefore ultimately futile and pointless.

## Conclusion

The argument that we should promote developments towards a posthuman condition, because the latter is good for individuals and good overall, has been shown to be problematic. The more posthumans differ from humans the more difficult it becomes to assess and compare their individual welfare, and since an assessment of the individual welfares is a necessary component in assessing the overall goodness of the posthuman condition the more difficult it becomes to compare this state to our current state. The aggregation of welfare over a large posthuman population has also been shown to be problematic. Taken together these two issues entail that assessing the overall goodness of a posthuman future is close to impossible.

The argument has further shown that the epistemic consequences of the Singularity means that any consideration of whether or not it should be promoted is ultimately pointless.

## References and Notes

1. I will discuss the possible distinction between post- and transhumans below, but for ease of reference use the term ‘posthuman’ to cover both in this paper.
2. H. G. Wells (1895) *The Time Machine* (London: William Heinemann).
3. G. Bear (2000) *Darwin’s Radio* (London: Ballantine Books).
4. A. C. Clarke (1953) *Childhood’s End* (London: Ballantine Books).
5. <http://humanityplus.org/>.
6. <http://singularityu.org/>.
7. M. More and N. Vita-More (eds) (2013) *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future* (Chichester: Wiley).
8. D. Wikler (2009) Paternalism in the age of cognitive enhancement: do civil liberties presuppose roughly equal mental ability? In: J. Savulescu and N. Bostrom (eds), *Human Enhancement* (Oxford: Oxford University Press), pp. 341–355.
9. A. A. van Niekerk (2014) After humanity? Philosophical and moral perspectives on the idea of posthumanity. *Studies in Sociology of Science*, **5**(2), pp. 119–124.
10. N. Bostrom (2009) The future of humanity. *Geopolitics, History, and International Relations*, **1**(2), pp. 61–78.
11. N. Bostrom (2005) A history of transhumanist thought. *Journal of Evolution and Technology*, **14**(1), pp. 1–25.
12. R. F. Harle (2002) Cyborgs, uploading and immortality—Some serious concerns. *Sophia*, **41**(2), pp. 73–85.
13. R. C. Merkle (2013) Uploading. In: M. More and N. Vita-More (eds), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future* (Chichester: Wiley), pp. 157–164.
14. V. Vinge (1983) First word. *Omni*, **10** available at <https://www.flickr.com/photos/87913776@N00/7113330247/>
15. V. Vinge (1993) The Coming Technological Singularity: How to Survive in the Post-human Era. <http://ntrs.nasa.gov/search.jsp?R=19940022856>
16. R. Kurzweil (2005) *The Singularity is Near: When Humans Transcend Biology* (London: Penguin).
17. A. Sandberg (2013) An overview of models of technological singularity. In: M. More and N. Vita-More (eds), *The Transhumanist Reader: Classical and*

*Contemporary Essays on the Science, Technology, and Philosophy of the Human Future* (Chichester: Wiley), pp. 376–394.

18. We will here ignore questions about whether the gender binary will be perpetuated in the posthuman condition.
19. J. Savulescu (2005) New breeds of humans: the moral obligation to enhance. *Reproductive BioMedicine Online*, **10**, pp. 36–39.
20. L. Wittgenstein (2009) *Philosophical Investigations* (4th edn) (Oxford: Blackwell).
21. S. Holm (2007) Naturalness and anthropology in modern bioethics, with a special view to trans- and post-humanism. In: H. Kragh (ed.), *Theology and Science – Issues for Future Dialogue* (Aarhus: University of Aarhus), pp. 17–29.
22. A. Buchanan (2009) Human nature and enhancement. *Bioethics*, **23**(3), pp. 141–150.
23. P. Boule (1963) *Planet of the Apes* (New York: Vanguard Press).
24. J. J. C. Smart and B. Williams (1973) *Utilitarianism: For and Against* (Cambridge: Cambridge University Press).
25. J. Rawls (2009) *A Theory of Justice* (rev. edn) (Boston: Harvard University Press).
26. R. Cole-Turner (2012) The singularity and the rapture: transhumanist and popular Christian views of the future. *Zygon*, **47**(4), pp. 777–796.
27. M. Zimmerman (2008) The singularity: a crucial phase in divine self-actualization? *Cosmos and History: The Journal of Natural and Social Philosophy*, **4**(1–2), pp. 347–370.
28. N. Bostrom and T. Ord (2006) The reversal test: eliminating status quo bias in applied ethics. *Ethics*, **116**(4), pp. 656–679.

### About the Author

**Søren Holm** is a Danish medical doctor and philosopher. He is Professor of Bioethics at the University of Manchester and Professor of Medical Ethics (part-time) at the Universities of Oslo and Aalborg. He is a former President of the European Society for the Philosophy of Medicine and Health Care, and of the International Association of Bioethics. His main research interest is bioethics and especially the interaction between ethical theory and applied ethics.