

Why I Want to be a Posthuman When I Grow Up

Nick Bostrom
Future of Humanity Institute
Faculty of Philosophy & James Martin 21st Century School
Oxford University
www.nickbostrom.com

[Published in: *Medical Enhancement and Posthumanity*, eds. Bert Gordijn and Ruth Chadwick (Springer, 2008): pp. 107-137. First circulated: 2006]

I am apt to think, if we knew what it was to be an angel for one hour, we should return to this world, though it were to sit on the brightest throne in it, with vastly more loathing and reluctance than we would now descend into a loathsome dungeon or sepulchre.¹
Berkley (1685-1753)

Abstract

Extreme human enhancement could result in “posthuman” modes of being. After offering some definitions and conceptual clarification, I argue for two theses. First, some posthuman modes of being would be very worthwhile. Second, it could be very good for human beings to become posthuman.

1. Setting the stage

The term “posthuman” has been used in very different senses by different authors.² I am sympathetic to the view that the word often causes more confusion than clarity, and that we might be better off replacing it with some alternative vocabulary. However, as the purpose of this paper is not to propose terminological reform but to argue for certain substantial normative theses (which one would naturally search for in the literature under the label “posthuman”), I will instead attempt to achieve intelligibility by clarifying the meaning that I shall assign to the word. Such terminological clarification is surely a minimum precondition for having a meaningful discussion about whether it might be good for us to become posthuman.

I shall define a *posthuman* as a being that has at least one posthuman capacity. By a *posthuman capacity*, I mean a general central capacity greatly exceeding the maximum attainable by any current human being without recourse to new technological means. I will use *general central capacity* to refer to the following:

- *healthspan* – the capacity to remain fully healthy, active, and productive, both mentally and physically
- *cognition* – general intellectual capacities, such as memory, deductive and analogical reasoning, and attention, as well as special faculties such as the capacity to understand and appreciate music, humor, eroticism, narration, spirituality, mathematics, etc.

¹ (Berkeley, Sampson et al. 1897), p. 172.

² The definition used here follows in the spirit of (Bostrom 2003). A completely different concept of “posthuman” is used in e.g. (Hayles 1999).

- *emotion* – the capacity to enjoy life and to respond with appropriate affect to life situations and other people

In limiting my list of general central capacities to these three, I do not mean to imply that no other capacity is of fundamental importance to human or posthuman beings. Nor do I claim that the three capacities in the list are sharply distinct or independent. Aspects of emotion and cognition, for instance, clearly overlap. But this short list may give at least a rough idea of what I mean when I speak of posthumans, adequate for present purposes.

In this paper, I will be advancing two main theses. The first is that some possible posthuman modes of being would be very good. I emphasize that the claim is not that *all* possible posthuman modes of being would be good. Just as some possible human modes of being are wretched and horrible, so too are some of the posthuman possibilities. Yet it would be of interest if we can show that there are some posthuman possibilities that would be very good. We might then, for example, specifically aim to realize those possibilities.

The second thesis is that it could be very good *for us* to become posthuman. It is possible to think that it could be good to be posthuman without it being good *for us* to become posthuman. This second thesis thus goes beyond the first. When I say “good for us”, I do not mean to insist that for every single current human individual there is some posthuman mode of being such that it would be good for that individual to become posthuman in that way. I confine myself to making a weaker claim that allows for exceptions. The claim is that for *most* current human beings, there are possible posthuman modes of being such that it could be good for these humans to become posthuman in one of those ways.

It might be worth locating the theses and arguments to be presented here within a broader discourse about the desirability of posthumanity. Opponents of posthumanity argue that we should not seek enhancements of a type that could make us, or our descendants, posthuman. We can distinguish at least five different “levels” on which objections against posthumanity could be launched:

Table 1: Levels of objection to posthumanity

<p>Level 0. “It can’t be done” Objections based on empirical claims to the effect that it is, and will remain, impossible or infeasible to create posthumans.</p> <p>Level 1. “It is too difficult/costly” Objections based on empirical claims that attempts to transform humans into posthumans, or to create new posthuman beings, would be too risky, or too expensive, or too psychologically distracting. Concerns about medical side-effects fall into this category, as do concerns that resources devoted to the requisite research and treatment would be taken away from more important areas.</p> <p>Level 2. “It would be too bad for society” Objections based on empirical claims about social consequences that would follow from the successful creation of posthuman beings, for example concerns about social inequality, discrimination, or conflicts between humans and posthumans.</p> <p>Level 3. “Posthuman lives would be worse than human lives” Objections based on normative claims about the value of posthuman lives compared to human lives.</p> <p>Level 4. “We couldn’t benefit” Objections based on agent-relative reasons against human beings transforming themselves into posthuman beings or against humans bringing new posthuman beings into existence. Although posthuman lives might be as good as or better than human lives, it would be bad <i>for us</i> to become posthuman or to create posthumans.</p>
--

This paper focuses on levels 3 and 4. I am thus setting aside issues of feasibility, costs, risks, side-effects, and social consequences. While those issues are obviously important when considering what we have most reason to do all things considered, they will not be addressed here.

Some further terminological specifications are in order. By *a mode of being* I mean a set of capacities and other general parameters of life. A posthuman mode of being is one that includes at least one posthuman capacity.

I shall speak of the value of particular modes of being. One might hold that primary value-bearers are some entities other than modes of being; e.g. mental states, subjective experiences, activities, preference-satisfactions, achievements, or particular lives. Such views are consistent with this paper. The position I seek to defend is consistent with a wide variety of formal and substantive theories of value. I shall speak of the value of modes of being for the sake of simplicity and convenience, but in doing so I do not mean to express a commitment to any particular controversial theory of value.

We might interpret “the values” of modes of beings as proxies for values that would be realized by particular lives instantiating the mode of being in question. If we proceed in this way, we create some indeterminacy. It is possible for a mode of being (and even more so for a *class* of modes of being) to be instantiated in a range of different possible lives, and for some of these lives to be good and others to be bad. In such a case, how could one assign a value to the mode of being itself?

Another way of expressing this concern is by saying that the value of instantiating a particular mode of being is context-dependent. In one context, the value might be high; in another, it might be negative. Nevertheless, it is useful to be able to speak of values of items other than those we accord basic intrinsic value. We might for example say that it is valuable to be in

good health and to have some money. Yet neither having good health nor having some money is guaranteed to make a positive difference to the value of your life. There are contexts in which the opposite is true. For instance, it could be the case that because you had some money you got robbed and murdered, or that because you were always in rude health you lacked a particular (short, mild) disease experience that would have transformed your mediocre novel into an immortal masterpiece. Even so, we can say that health and money are good things without thereby implying that they are intrinsically valuable or that they add value in all possible contexts. When we say that they are valuable we might merely mean that these things would *normally* make a positive contribution to the value of your life; they would add value in a very wide range of plausible contexts. This mundane meaning is what I have in mind when I speak of modes of being having a value: i.e., in a very wide range of plausible contexts, lives instantiating that mode of being would tend to contain that value.³

A life might be good or bad because of its causal consequences for other people, or for the contribution it makes to the overall value of a society or a world. But here I shall focus on the value that a life has for the person whose life it is: how good (or bad) it is for the subject to have this life. The term “well-being” is often used in this sense.⁴

When I speak of the value of a life here, I do not refer to the moral status of the person whose life it is. It is a separate question what the moral status would be of human and posthuman beings. We can assume for present purposes that human and posthuman persons would have the same moral status. The value of a life refers, rather, to how well a life goes for its subject. Different human lives go differently well, and in this sense their lives have different values. The life of a person who dies from a painful illness at age 15 after having lived in extreme poverty and social isolation is typically worse and has less value than that of a person who has an 80-year-long life full of joy, creativity, worthwhile achievements, friendships, and love. Whatever terminology we use to describe the difference, it is plain that the latter kind of life is more worth having. One way to express this platitude is by saying that the latter life is more valuable than the former.⁵ This is consistent with assigning equal moral status to the two different persons whose lives are being compared.

Some pairs of possible lives are so different that it is difficult – arguably impossible – to compare their value. We can leave aside the question of whether, for every pair of possible lives, it is true either that one is better than the other, or that they are equally good; that is, whether all pairs of possible lives have commensurable value. We shall only assume that at least for some pairs of possible lives, one is definitely better than the other.

To supply our minds with a slightly more concrete image of what becoming posthuman might be like, let us consider a vignette of how such a process could unfold.

³ Compare this take on “mundane values” with the notion of mid-level principles in applied ethics. The principle of respecting patient autonomy is important in medical ethics. One might accept this if one holds that respect for patient autonomy is an implication of some fundamental ethical principle. But equally, one might accept patient autonomy as an important mid-level principle even if one merely holds that this is a way of expressing a useful rule of thumb, a sound policy rule, or a derived ethical rule that is true in a world like ours because of various empirical facts even though it is not necessarily true in all possible worlds. For the role of mid-level principles in applied ethics, see e.g. (Beauchamp and Childress 2001).

⁴ I am thus not concerned here with global evaluations into which individuals’ well-being might enter as a factor, e.g. evaluations involving values of diversity, equality, or comparative fairness.

⁵ I do not assume that the value of a life, or well-being, supervenes on the mental experiences of a person, nor that it supervenes on a thin time-slice of a person’s life. It could represent a wider and more global evaluation of how well a person’s life is going.

2. Becoming posthuman

Let us suppose that you were to develop into a being that has posthuman healthspan and posthuman cognitive and emotional capacities. At the early steps of this process, you enjoy your enhanced capacities. You cherish your improved health: you feel stronger, more energetic, and more balanced. Your skin looks younger and is more elastic. A minor ailment in your knee is cured. You also discover a greater clarity of mind. You can concentrate on difficult material more easily and it begins making sense to you. You start seeing connections that eluded you before. You are astounded to realize how many beliefs you had been holding without ever really thinking about them or considering whether the evidence supports them. You can follow lines of thinking and intricate argumentation farther without losing your foothold. Your mind is able to recall facts, names, and concepts just when you need them. You are able to sprinkle your conversation with witty remarks and poignant anecdotes. Your friends remark on how much more fun you are to be around. Your experiences seem more vivid. When you listen to music you perceive layers of structure and a kind of musical logic to which you were previously oblivious; this gives you great joy. You continue to find the gossip magazines you used to read amusing, albeit in a different way than before; but you discover that you can get more out of reading Proust and *Nature*. You begin to treasure almost every moment of life; you go about your business with zest; and you feel a deeper warmth and affection for those you love, but you can still be upset and even angry on occasions where upset or anger is truly justified and constructive.

As you yourself are changing you may also begin to change the way you spend your time. Instead of spending four hours each day watching television, you may now prefer to play the saxophone in a jazz band and to have fun working on your first novel. Instead of spending the weekends hanging out in the pub with your old buddies talking about football, you acquire new friends with whom you can discuss things that now seem to you to be of greater significance than sport. Together with some of these new friends, you set up a local chapter of an international non-profit to help draw attention to the plight of political prisoners.

By any reasonable criteria, your life improves as you take these initial steps towards becoming posthuman. But thus far your capacities have improved only within the natural human range. You can still partake in human culture and find company to engage you in meaningful conversation. Consider now a more advanced stage in the transformation process...

You have just celebrated your 170th birthday and you feel stronger than ever. Each day is a joy. You have invented entirely new art forms, which exploit the new kinds of cognitive capacities and sensibilities you have developed. You still listen to music – music that is to Mozart what Mozart is to bad Muzak. You are communicating with your contemporaries using a language that has grown out of English over the past century and that has a vocabulary and expressive power that enables you to share and discuss thoughts and feelings that unaugmented humans could not even think or experience. You play a certain new kind of game which combines VR-mediated artistic expression, dance, humor, interpersonal dynamics, and various novel faculties and the emergent phenomena they make possible, and which is more fun than anything you ever did during the first hundred years of your existence. When you are playing this game with your friends, you feel how every fiber of your body and mind is stretched to its limit in the most creative and imaginative way, and you are creating new realms of abstract and concrete beauty that humans could never (concretely) dream of. You are always ready to feel with those who suffer misfortunes, and to work hard to help them get back on their feet. You are also involved in a large voluntary organization that works to reduce suffering of animals in their natural environment in ways that permit ecologies to continue to function in traditional ways; this involves political efforts combined with advanced science and information processing services. Things are getting better, but already each day is fantastic.

As we seek to peer farther into posthumanity, our ability to concretely imagine what it might be like trails off. If, aside from extended healthspans, the essence of posthumanity is to be

able to have thoughts and experiences that we cannot readily think or experience with our current capacities, then it is not surprising that our ability to imagine what posthuman life might be like is very limited. Yet we can at least perceive the outlines of some of the nearer shores of posthumanity, as we did in the imaginary scenario above. Hopefully such thought experiments are already enough to give plausibility to the claim that becoming posthuman could be good for us.

In the next three sections we will look in a little more detail at each of the three general central capacities that I listed in the introduction section. I hope to show that the claim that it could be very good to be posthuman is not as radical as it might appear to some. In fact, we will find that individuals and society already in some ways seem to be implicitly placing a very high value on posthuman capacities – or at least, there are strong and widely accepted tendencies pointing that way. I therefore do not regard my claim as in any strong sense revisionary. On the contrary, I believe that the denial of my claim would be strongly revisionary in that it would force us to reject many commonly accepted ethical beliefs and approved behaviors. I see my position as a conservative extension of traditional ethics and values to accommodate the possibility of human enhancement through technological means.

3. Healthspan

It seems to me fairly obvious why one might have reason to desire to become a posthuman in the sense of having a greatly enhanced capacity to stay alive and stay healthy.⁶ I suspect that the majority of humankind already has such a desire implicitly.

People seek to extend their healthspan, i.e. to remain healthy, active, and productive. This is one reason why we install air bags in cars. It may also explain why we go to the doctor when we are sick, why higher salaries need to be paid to get workers to do physically dangerous work, and why governments and charities give money to medical research.⁷ Instances of individuals sacrificing their lives for the sake of some other goal, whether suicide bombers, martyrs, or drug addicts, attract our attention precisely because their behavior is unusual. Heroic rescue workers who endanger their lives on a dangerous mission are admired because we assume that they are putting at risk something that most people would be very reluctant to risk, their own survival.

For some three decades, economists have attempted to estimate individuals' preferences over mortality and morbidity risk in labor and product markets. While the tradeoff estimates vary considerably between studies, one recent meta-analysis puts the median value of the value of a statistical life for prime-aged workers to about \$7 million in the United States.⁸ A study by the EU's Environment Directorates-General recommends the use of a value in the interval €0.9 to €3.5 million.⁹ Recent studies by health economists indicate that improvements in the health status of the U.S. population over the 20th century have made as large a contribution to raising the standards of living as all other forms of consumption growth combined.¹⁰ While the exact numbers are debatable, there is little doubt that most people place a very high value on their continued existence in a healthy state.

Admittedly, a desire to extend one's healthspan is not necessarily a desire to become posthuman. To become posthuman by virtue of healthspan extension, one would need to achieve

⁶ Having such a capacity is compatible with also having the capacity to die at any desired age. One might thus desire a capacity for greatly extended healthspan even if one doubts that one would wish to live for more than, say, 80 years. A posthuman healthspan capacity would give one the option of much longer and healthier life, but one could at any point decide no longer to exercise the capacity.

⁷ Although on the last item, see (Hanson 2000) for an alternative view.

⁸ (Viscusi and Aldy 2003).

⁹ (Johansson 2002).

¹⁰ (Murphy and Topel 2003; Nordhaus 2003).

the capacity for a healthspan that greatly exceeds the maximum attainable by any current human being without recourse to new technological means. Since at least some human beings already manage to remain quite healthy, active, and productive until the age of 70, one would need to desire that one's healthspan were extended greatly beyond this age in order that it would count as having a desire to become posthuman.¹¹

Many people will, if asked about how long they would wish their lives to be, name a figure between 85 and 90 years.¹² In many cases, no doubt, this is because they assume that a life significantly longer than that would be marred by deteriorating health – a factor from which we must abstract when considering the desirability of healthspan extension. People's stated willingness to pay to extend their life by a certain amount does in fact depend strongly on the health status and quality of that extra life.¹³ Since life beyond 85 is very often beset by deteriorating health, it is possible that this figure substantially underestimates how long most people would wish to live if they could be guaranteed perfect health.

It is also possible that a stated preference for a certain lifespan is hypocritical. Estimates based on revealed preferences in actual market choices, such as fatality risk premiums in labor markets or willingness to pay for health care and other forms of fatality risk reduction might be more reliable. It would be interesting to know what fraction of those who claim to have no desire for healthspan extension would change their tune if they were ever actually handed a pill that would reliably achieve this effect. My conjecture would be that when presented with a real-world choice, most would choose the path of prolonged life, health, and youthful vigor over the default route of aging, disease, and death.

One survey asked: "Based on your own expectations of what old age is like, if it were up to you, how long would you personally like to live – to what age?" Only 27% of respondents said they would like to live to 100 or older.¹⁴ A later question in the same survey asked: "Imagine you could live to 100 or older, but you'd have to be very careful about your diet, exercise regularly, not smoke, avoid alcohol, and avoid stress. Would it be worth it, or not?" To this, 64% answered in the affirmative! Why should *more* people want to live beyond 100 when restrictions on activity are imposed? Is it because it frames the question more as if it were a real practical choice rather than as an idle mind game? Perhaps when the question is framed as a mind game, respondents tend to answer in ways which they believe expresses culturally approved attitudes, or which they think signal socially desirable personal traits (such as having "come to terms" with one's own mortality), while this tendency is diminished when the framing suggests a practical choice with real consequences. We do not know for sure, but this kind of anomaly suggests that we should not take people's stated "preferences" about how long they would wish to live too seriously, and that revealed preferences might be a more reliable index of their guiding values.

It is also worth noting that only a small fraction of us commit suicide, suggesting that our desire to live is almost always stronger than our desire to die.¹⁵ Our desire to live, *conditional on our being able to enjoy full health*, is even stronger. This presumption in favor of life is in fact so strong that if somebody wishes to die soon, even though they are seemingly fully healthy, with a long remaining healthy life expectancy, and if their external circumstances in life are not catastrophically wretched, we would often tend suspect that they might be suffering from

¹¹ At least one human, Jeanne Calment, lived to 122. But although she remained in relatively fair health until close to her death, she clearly suffered substantial decline in her physical (and presumably mental) vigor compared to when she was in her twenties. She did not retain the capacity to be *fully* healthy, active, and productive for 122 years.

¹² (Cohen and Langer 2005).

¹³ (Johnson, Desvousges et al. 1998).

¹⁴ (Cohen and Langer 2005).

¹⁵ For some, the reluctance to commit suicide might reflect a desire not to kill oneself rather than a desire not to die, or alternatively a fear of death rather than an authentic preference not to die.

depression or other mental pathology. Suicidal ideation is listed as a diagnostic symptom of depression by the American Psychiatric Association.¹⁶

Even if a stated preference against healthspan extension were sincere, we would need to question how well-considered and informed it is. It is of relevance that those who know most about the situation and are most directly affected by the choice, namely the elderly, usually prefer life to death. They usually do so when their health is poor, and overwhelmingly choose life when their health is at least fair. Now one can argue that a mentally intact 90-year-old is in a better position to judge how their life would be affected by living for another year than she was when she was 20, or 40. If most healthy and mentally intact 90-year-olds prefer to live for another year (at least if they could be guaranteed that this extra year would be one of full health and vigor), this would be evidence against the claim that it would be better for these people that their lives end at 90.¹⁷ Similarly, of course, for people of even older age.

One can compare this situation with the different case of somebody becoming paraplegic. Many able-bodied people believe that their lives would not be worth living if they became paraplegic. They claim that they would prefer to die rather than continuing life in a paraplegic state. Most people who have actually become paraplegic, however, find that their lives are worth living.¹⁸ People who are paraplegic are typically better judges of whether paraplegic lives are worth continuing than are people who have never experienced what it is like to be paraplegic. Similarly, people who are 90 years old are in a better position to judge whether their lives are worth continuing than are younger people (including themselves at any earlier point in their lives).¹⁹

One study assessed the will to live among 414 hospitalized patients aged 80 to 98 years, presumably representing the frailer end of the distribution of the “old old”. 40.8% of respondents were unwilling to exchange any time in their current state of health for a shorter life in excellent health, and 27.8% were willing to give up at most 1 month of 12 in return for excellent health.²⁰ (Patients who were still alive one year later were even less inclined to give up life for better health, but with continued large individual variations in preferences.) The study also found that patients were willing to trade significantly less time for a healthy life than their surrogates assumed they would.

Research shows that life-satisfaction remains relatively stable into old age. One survey of 60,000 adults from 40 nations discovered a slight upward trend in life-satisfaction from the 20s to the 80s in age.²¹ Life satisfaction showed this upward trend even though there was some loss of positive affect. Perhaps life-satisfaction would be even higher if positive affect were improved (a possibility we shall discuss in a later section). Another study, using a cross-sectional sample (age range 70-103 years), found that controlling for functional health constraints reversed the direction of the relationship between age and positive affect and produced a negative association between

¹⁶ DSM-IV (American Psychiatric Association. 2000).

¹⁷ This is a kind of Millian best-judge argument. However, if fear of death were irrational, one could argue that people who are closer to death are on average worse judges of the value for them of an extra year of life, because their judgments would tend to be more affected by irrational fear.

¹⁸ This basic result is reflected in many chronic disease conditions (Ubel, Loewenstein et al. 2003). The discrepancy of attitudes seems to be due to non-patient’s failure to realize the extent to which patients psychologically adapt to their condition (Damschroder, Zikmund-Fisher et al. 2005).

¹⁹ The analogy with paraplegia is imperfect in at least one respect: when the issue is healthspan extension, we are considering whether it would be worth living an extended life in perfect health and vigor. If anything, this discrepancy strengthens the conclusion, since it is more worth continuing living in perfect health than in poor health, not less worth it.

²⁰ (Tsevat, Dawson et al. 1998). See also (McShine, Lesser et al. 2000). For a methodological critique, see (Arnesen and Norheim 2003).

²¹ (Diener and Suh 1998).

age and negative affect.²² These findings suggest that some dimensions of subjective well-being, such as life-satisfaction, do not decline with age but might actually increase somewhat, and that the decline in another dimension of subjective well-being (positive affect) is not due to aging per se but to health constraints.

Most people reveal through their behavior that they desire continued life and health,²³ and most of those who are in the best position to judge the value of continued healthy life, at any age, judge that it is worth having. This constitutes prima facie support for the claim that extended life is worth having even when it is not fully healthy. The fact that this holds true at all currently realized ages suggests that it is not a strongly revisionary view to hold that it could be good for many people to become posthuman through healthspan extension. Such a view might already be implicitly endorsed by many.

4. Cognition

People also seem to be keen on improving cognition. Who wouldn't want to remember names and faces better, to be able more quickly to grasp difficult abstract ideas, and to be able to "see connections" better? Who would seriously object to being able to appreciate music at a deeper level? The value of optimal cognitive functioning is so obvious that to elaborate the point may be unnecessary.²⁴

This verdict is reflected in the vast resources that society allocates to education, which often explicitly aims not only to impart specific items of knowledge but also to improve general reasoning abilities, study skills, critical thinking, and problem solving capacity.²⁵ Many people are also keen to develop various particular talents that they may happen to have, for example musical or mathematical, or to develop other specific faculties such as aesthetic appreciation, narration, humor, eroticism, spirituality etc. We also reveal our desire for improving our cognitive functioning when take a cup of coffee to increase our alertness or when we regret our failure to obtain a full night's sleep because of the detrimental effects on our intellectual performance.

Again, the fact that there is a common desire for cognitive improvement does not imply that there is a common desire for becoming posthuman. To want to become posthuman through cognitive improvement, one would have to want a great deal of cognitive improvement. It is logically possible that each person would only want to become slightly more intelligent (or musical, or humorous) than he or she currently is and would not want any very large gain. I will offer two considerations regarding this possibility.

First, it seems to me (based on anecdotal evidence and personal observations) that people who are already endowed with above-average cognitive capacities are at least as eager, and, from what I can tell, actually *more* eager to obtain further improvements in these capacities than are people who are less talented in these regards. For instance, someone who is musically gifted is likely to spend more time and effort trying to further develop her musical capacities than is somebody who lacks a musical ear; and likewise for other kinds of cognitive gifts.

This phenomenon may in part reflect the external rewards that often accrue to those who excel in some particular domain. An extremely gifted musician might reap greater rewards in terms of money and esteem from a slight further improvement in her musicality than would

²² (Kunzmann, Little et al. 2000).

²³ This is fully consistent with the fact that many people knowingly engage in risky behaviors such as smoking. This might simply mean that they are unable to quit smoking, or that they desire the pleasure of smoking more than they desire a longer healthier life. It does not imply that they do not desire longer healthier life.

²⁴ One might even argue that a desire for cognitive improvement is a constitutive element of human rationality, but I will not explore that hypothesis here.

²⁵ U.S. *public* expenditure on education in 2003 was 5.7% of its GDP (World Bank. 2003).

somebody who is not musically gifted to begin with. That is, the difference in external rewards is sometimes greater for somebody who goes from very high capacity to outstandingly high capacity than it is for somebody who goes from average capacity to moderately high capacity. However, I would speculate that such differences in external rewards are only part of the explanation and that people who have high cognitive capacities are usually also more likely (or at least no less likely) to desire further increases in those capacities than are people of lower cognitive capacities even when only the intrinsic benefits of capacities are considered. Thus, if we imagine a group of people placed in solitary confinement for the remainder of their lives, but with access to books, musical instruments, paints and canvasses, and other prerequisites for the exercise of capacities, I would hypothesize that those with the highest pre-existing capacity in a given domain would be more likely (or at least not less likely) to work hard to further develop their capacities in that domain, for the sake of the intrinsic benefits that the possession and exercise of those capacities bestow, than would those with lower pre-existing capacities in the same domain.²⁶ While \$100 brings vastly less utility to a millionaire than to a pauper, the marginal utility of improved cognitive capacities does not seem to exhibit a similar decline.

These considerations suggest that there are continuing returns in the “intrinsic” (in the sense of non-instrumental, non-positional) utility of gains in cognitive capacities, at least within the range of capacity that we find instantiated within the current human population.²⁷ It would be implausible to suppose that the current range of human capacity, in all domains, is such that while increments of capacity within this range are intrinsically rewarding, yet any further increases outside the current human range would lack intrinsic value. Again, we have a prima facie reason for concluding that enhancement of cognitive capacity to the highest current human level, and probably beyond that, perhaps up to and including the posthuman level, would be intrinsically desirable for the enhanced individuals. We get this conclusion if we assume that those who have a certain high capacity are generally better judges of the value of having that capacity or of a further increment of that capacity than are those who do not possess the capacity in question to the same degree.

4. Emotion

It is straightforward to determine what would count as an enhancement of healthspan. We have a clear enough idea of what it means to be healthy, active, and productive, and the difference between this state and that of being sick, incapacitated, or dead. An enhancement of healthspan is simply an intervention that prolongs the duration of the former state. It is more difficult to define precisely what would count as a cognitive enhancement because the measure of cognitive functioning is more multifaceted, various cognitive capacities can interact in complex ways, and it is a more normatively complex problem to determine what combinations of particular cognitive competences are of value in different kinds of environments. For instance, it is not obvious what degree of tendency to forget certain kinds of facts and experiences is desirable. The answer might depend on a host of contextual factors. Nevertheless, we do have some general idea of how we might value various increments or decrements in many aspects of our cognitive functioning – a

²⁶ Complication: if high capacity were solely a result from having spent a lot of effort in developing that capacity, then the people with high capacity in some domain might be precisely those that started out having an unusually strong desire for having a strong capacity in that domain. It would then not be surprising that those with high capacity would have the strongest desire for further increases in capacity. Their stronger desire for higher capacity might then not be the result of more information and better acquaintance with what is at stake, but might instead simply reflect a prior inclination.

²⁷ It would be more difficult to determine whether the marginal intrinsic utility of gains in capacity are constant, or diminishing, or increasing at higher levels of capacity, and if so by what amount.

sufficiently clear idea, I suggest, to make it intelligible without much explanation what one might mean by phrases like “enhancing musical ability”, “enhancing abstract reasoning ability” etc.

It is considerably more difficult to characterize what would count as emotional enhancement. Some instances are relatively straightforward. Most would readily agree that helping a person who suffers from persistent suicidal depression as the result of a simple neurochemical imbalance so that she once again becomes capable of enjoyment and of taking an interest in life would be to help her improve her emotional capacities. Yet beyond cases involving therapeutic interventions to cure evident psychopathology it is less clear what would count as an enhancement. One’s assessment of such cases often depends sensitively on the exact nature of one’s normative beliefs about different kinds of possible emotional constitutions and personalities.

It is correspondingly difficult to say what would constitute a “posthuman” level of emotional capacity. Nevertheless, people often do strive to improve their emotional capacities and functionings. We may seek to reduce feelings of hate, contempt, or aggression when we consciously recognize that these feelings are prejudiced or unconstructive. We may take up meditation or physical exercise to achieve greater calm and composure. We may train ourselves to respond more sensitively and empathetically to those we deem deserving of our trust and affection. We may try to overcome fears and phobias that we recognize as irrational, or we may wrestle with appetites that threaten to distract us from what we value more. Many of us expend life-long effort to educate and ennoble our sentiments, to build our character, and to try to become better people. Through these strivings, we seek to achieve goals involving modifying and improving our emotional capacities.

An appropriate conception of emotional capacity would be one that incorporates or reflects these kinds of goal, while allowing perhaps for there being a wide range of different ways of instantiating “high emotional capacity”, that is to say, many different possible “characters” or combinations of propensities for feeling and reacting that could each count as excellent in its own way. If this is admitted, then we could make sense of emotional enhancement in a wide range of contexts, as being that which makes our emotional characters more excellent. A posthuman emotional capacity would be one which is much more excellent than that which any current human could achieve unaided by new technology.

One might perhaps question whether there are possible emotional capacities that would be *much* more excellent than those attainable now. Conceivably, there might be a maximum of possible excellence of emotional capacity, and those people who currently have the best emotional capacities might approach so closely to this ideal that there is not enough potential left for improvement to leave room for a posthuman realm of emotional capacity. I doubt this, because aside from the potential for fine-tuning and balancing the various emotional sensibilities we already have, I think there might also be entirely new psychological states and emotions that our species has not evolved the neurological machinery to experience, and some of these sensibilities might be ones we would recognize as extremely valuable if we became acquainted with them.

It is difficult intuitively to understand what such novel emotions and mental states might be like. This is unsurprising, since by assumption we currently lack the required neurological bases. It might help to consider a parallel case from within the normal range of human experience. The experience of romantic love is something that many of us place a high value on. Yet it is notoriously difficult for a child or a prepubescent teenager to comprehend the meaning of romantic love or why adults should make so much fuss about this experience. Perhaps we are all currently in the situation of children relative to the emotions, passions, and mental states that posthuman beings could experience. We may have no idea of what we are missing out on until we attain posthuman emotional capacities.

One dimension of emotional capacity that we can imagine enhanced is subjective well-being and its various flavors: joy, comfort, sensual pleasures, fun, positive interest and

excitement. Hedonists claim that pleasure is the only intrinsic good, but one need not be a hedonist to appreciate pleasure as one important component of the good. The difference between a bleak, cold, horrid painful world and one that is teeming with fun and exciting opportunities, full of delightful quirks and lovely sensations, is often simply a difference in the hedonic tone of the observer. Much depends on that one parameter.

It is an interesting question how much subjective well-being could be enhanced without sacrificing other capacities that we may value. For human beings as we are currently constituted, there is perhaps an upper limit to the degree of subjective well-being that we can experience without succumbing to mania or some other mental unbalance that would prevent us from fully engaging with the world if the state were indefinitely prolonged. But it might be possible for differently constituted minds to have experiences more blissful than those that humans are capable of without thereby impairing their ability to respond adequately to their surroundings. Maybe for such beings, gradients of pleasure could play a role analogous to that which the scale ranging between pleasure and pain has for us.²⁸ When thinking the possibility of *posthumanly happy* beings, and their psychological properties, one must abstract from contingent features of the human psyche. An experience that would consume us might perhaps be merely “spicy” to a posthuman mind.

It is not necessary here to take a firm stand on whether posthuman levels of pleasure are possible, or even on whether posthuman emotional capacities more generally are possible. But we can be confident that, at least, there is vast scope for improvements for most of individuals in these dimensions because even within the range instantiated by currently existing humans, there are levels of emotional capacities and degrees of subjective well-being that, for most of us, are practically unattainable to the point of exceeding our dreams. The fact that such improvements are eagerly sought by many suggests that if posthuman levels were possible, they too would be viewed as highly attractive.²⁹

5. Structure of the argument, and further supporting reasons

It might be useful to pause briefly to reflect on the structure of the argument presented so far. I began by listing three general central capacities (healthspan, cognition, and emotion), and I defined a posthuman being as one who has at least one of these capacities in a degree unattainable by any current human being unaided by new technology.

I offered some plausibility arguments suggesting that it could be highly desirable to have posthuman levels of these capacities. I did this partly by clarifying what having the capacities would encompass and by explaining how some possible objections would not apply because they rely on a misunderstanding of what is proposed. Furthermore, I tried to show that for each of the three capacities we find that many individuals actually desire to develop the capacities to higher levels and often undertake great effort and expense to achieve these aims. This desire is also reflected in social spending priorities, which devote significant resources to e.g. healthspan-extending medicine and cognition-improving education. Significantly, at least in the cases of healthspan extension and cognitive improvement, the persons best placed to judge the value and desirability of incremental improvements at the high end of the contemporary human capacity distribution seem to be especially likely to affirm the desirability of such additional improvements of capacity. For many cognitive faculties, it appears that the marginal utility of improvements *increases* with capacity levels. This suggests that improvements beyond the

²⁸ (Pearce 2004).

²⁹ The quest for subjective well-being, in particular, seems to be a powerful motivator for billions of people even though arguably none of the various means that have been attempted in this quest has yet proved very efficacious in securing the goal (Brickman and Campbell 1971).

current human range would also be viewed as desirable when evaluated by beings in a better position to judge than we currently are.

That people desire X does not imply that X is desirable. Nor does the fact that people find X desirable, even when this judgment is shared among those who are in the best position to judge the desirability of X, prove that X is desirable or valuable. Even if one were to assume some version of a dispositional theory of value, it does not follow from these premises that X is valuable. A dispositional theory of value might assert something like the following:

X is valuable for A if and only if A would value X if A were perfectly rational, perfectly well-informed, and perfectly acquainted with X.³⁰

The people currently best placed to judge the desirability for an individual of enhancement of her general central capacities are neither perfectly rational, nor perfectly well-informed, nor perfectly acquainted with the full meaning of such enhancements. If these people were more rational or obtained more information or became better acquainted with the enhancements in question, they would perhaps no longer value the enhancements. Even if everybody judged becoming posthuman as desirable, it is a logical possibility that becoming posthuman is not valuable, even given a theory of value that defines value in terms of valuing-dispositions.

The argument presented in the preceding sections is not meant to be deductive. Its ambition is more modest: to remind us of the plausibility of the view that (1) enhancements along the three dimensions discussed are possible in principle and of significant potential intrinsic value, and (2) enhancements along these dimensions large enough to produce posthuman beings could have very great intrinsic value. This argument is defeasible. One way in which it could be defeated would be by pointing to further information, rational reasoning, or forms of acquaintance, not accounted for by current opinion, and which would change current opinion if it were incorporated. Critics could for example try to point to some reasoning mistake that very old people commit when they judge that it would be good for them to live another year in perfect health. However, I think the considerations I have pointed to provide prima facie evidence for my conclusions.

There are other routes by which one could reach the position that I have advocated, which supports the above arguments. For instance, one might introspect one's own mind to determine whether being able to continue to live in good health longer, being able better to understand the world and other people, or being able more fully to enjoy life and to react with appropriate affect to life events would seem like worthwhile goals for oneself if they were obtainable.³¹ Alternatively, one might examine whether having these capacities to an enhanced or even posthuman degree could enable one to realize states and life paths that would have great value according to one's favorite theory of value. (To me, both these tests deliver affirmative verdicts on (1) and (2).)

Yet another route to making the foregoing conclusions plausible is by considering our current ignorance and the vastness of the as-yet unexplored terrain. Let S_H be the "space" of possible modes of being that could be instantiated by someone with current human capacities. Let S_P be the space of possible modes of being that could be instantiated by someone with posthuman capacities. In an intuitive sense, S_P is enormously much larger than S_H . There is a larger range of possible life courses that could be lived out during a posthuman lifespan than during a human lifespan. There are more thoughts that could be thought with posthuman cognitive capacities than with human capacities (and more musical structures that could be created and appreciated with posthuman musical capacities etc.). There are more mental states and emotions that could be experienced with posthuman emotional faculties than with human ones. So why, apart from a lack

³⁰ See e.g. (Lewis 1989).

³¹ See e.g. (Bostrom 2005).

of imagination, should anybody suppose that the S_H already contains all the most valuable and worthwhile modes of being?

An analogy: For as long as anybody remembers, a tribe has lived in a certain deep and narrow valley. They rarely think of what lies outside their village, and on the few occasions when they do, they think of it only as a mythical realm. One day a sage who has been living apart from the rest, on the mountainside, comes down to the village. He explains that he has climbed to the top of the mountain ridge and from there he could see the terrain stretching far away, all the way to the horizon. He saw plains, lakes, forests, winding rivers, mountains, and the sea. Would it not be reasonable, he says, in lieu of further exploration, to suppose that this vast space is likely to be home to natural resources of enormous value? – Similarly, the sheer size and diversity of S_P is in itself a *prima facie* reason for thinking that it is likely to contain some very great values.³²

6. Personal identity

Supposing the previous sections have succeeded in making it plausible that being a posthuman could be good, we can now turn to a further question: whether becoming posthuman could be good *for us*. It may be good to be Joseph Haydn. Let us suppose that Joseph Haydn had a better life than Joe Bloggs so that in some sense it is better to be Haydn and living the life that Haydn lived than to be Bloggs and living Bloggs' life. We may further suppose that this is so from Bloggs' evaluative standpoint. Bloggs might recognize that on all the objective criteria which he thinks makes for a better mode of being and a better life, Haydn's mode of being and life are better than his own. Yet it does not follow that it would be good for Bloggs to "become" Haydn (or to become some kind of future equivalent of Haydn) or to live Haydn's life (or a Haydn-like life). There are several possible reasons for this which we need to examine.

First, it might not be possible for Bloggs to become Haydn without ceasing to be Bloggs. While we can imagine a thought experiment in which Bloggs' body and mind are gradually transformed into those of Haydn (or of a Haydn-equivalent), it is not at all clear that personal identity could be preserved through such a transformation. If Bloggs' personal identity is essentially constituted by some core set of psychological features such as his memories and dispositions, then, since Haydn does not have these features, the person Bloggs could not become a Haydn-equivalent. Supposing that Bloggs has a life that is worth living, any transformation that causes the person Bloggs to cease to exist might be bad for Bloggs, including one that transforms him into Haydn.

Could a current human become posthuman while remaining the same person, or is the case like the one of Bloggs becoming Haydn, the person Bloggs necessarily ceasing to exist in the process? The case of becoming posthuman is different in an important respect. Bloggs would have to lose all the psychological characteristics that made him person Bloggs in order to become Haydn. In particular, he would have to lose all his memories, his goals, his unique skills, and his entire personality would be obliterated and replaced by that of Haydn. By contrast, a human being could retain her memories, her goals, her unique skills, and many important aspects of her personality even as she becomes posthuman. This could make it possible for personal identity to be preserved during the transformation into posthuman.³³

It is obvious that personal identity could be preserved, at least in the short run, if posthuman status is achieved through radical healthspan enhancement. Suppose that I learnt that tonight after I go to bed, a scientist will perform some kind of molecular therapy on my cells

³² (Bostrom 2004).

³³ See also (DeGrazia 2005). DeGrazia argues that identity-related challenges to human enhancement largely fails, both ones based on considerations of personal identity and ones based on narrative identity (authenticity), although he mainly discusses more moderate enhancements than those I focus on in this paper.

while I'm sleeping to permanently disable the aging processes in my body. I might worry that I would not wake up tomorrow because the surgery might go wrong. I would *not* worry that I might not wake up tomorrow because the surgery succeeded. Healthspan enhancement would help *preserve* my personal identity. (If the psychological shock of discovering that my life-expectancy had been extended to a thousand years were so tremendous that it would completely remold my psyche, it is possible that the new me would not be the same person as the old me. But this is not a necessary consequence.³⁴)

Walter Glannon has argued that a lifespan of 200 years or more would be undesirable because personal identity could not be persevered over such a long life.³⁵ Glannon's argument presupposes that personal identity (understood here as a determinant of our prudential concerns) depends on psychological connectedness. On this view, we now have prudential interests in a future time segment of our organism only if that future time segment is psychologically connected to the organism's present time segment through links of backward-looking memories and forward-looking projects and intentions. If a future time segment of my brain will not remember anything about what things are like for me now, and if I now have no projects or intentions that extend that far into the future, then that future time segment is not part of my person. Glannon asserts that these psychological connections that hold us together as persons could not extend over 200 years or so.

There are several problems with Glannon's argument, even if we accept his metaphysics of personal identity. There is no reason to think it impossible to have intentions and projects that range over more than 200 years. This would seem possible even with our current human capacities. For example, I can easily conceive of exciting intellectual and practical projects that may take me many hundreds of years to complete. It is also dubious to assume that a healthy future self several hundred years older than I am now might be unable to remember things from current life stage. Old people often remember their early adulthood quite well, and it is not clear that these memories always decline significantly over time. And of course, the concern about distant future stages being unable to remember their earlier stages disappears completely if we suppose that enhancements of memory capacity become available.³⁶ Furthermore, if Glannon was right, it would follow that it is "undesirable" for a small child to grow up, since adults do not remember what it was like to be a small child and since small children do not have projects or intentions that extend over time spans as long as decades. This implication would be counterintuitive. It is more plausible that it can be desirable for an agent to survive and continue to develop, rather than to die, even if psychological connections eventually become attenuated. In the same way, it could be desirable for us to acquire the capacity to have a posthuman healthy lifespan, even if we could not remain the same person over time scales of several centuries.

The case that personal identity could be preserved is perhaps less clear-cut with regard to radical cognitive or emotional enhancement. Could a person become radically smarter, more musical, or come to possess much greater emotional capacities without ceasing to exist? Here the answer might depend more sensitively on precisely which changes we are envisaging, how those changes would be implemented, and on how the enhanced capacities would be used. The case for thinking that both personal identity and narrative identity would be preserved is arguably

³⁴ It is not even a psychologically plausible consequence even within the limitations of current human psychology. Compare the case to that of a man on death row who has a remaining life-expectancy of 1 day. An unexpected pardon suddenly extends this to 40 years – an extension by a factor of 14,610! He might be delighted, stunned, or confused, but he does not cease to exist as a person. If he did, it would presumably be *bad* for him to be pardoned.

Even if one believed (erroneously in my view) that mortality or aging were somehow essential features of the persons we are, these features are consistent with vastly extended healthspan.

³⁵ (Glannon 2002).

³⁶ It is clear that in order for an *extremely* long life to not become either static or self-repeating, it would be necessary that mental growth continues.

strongest if we posit that (a) the changes are in the form of addition of new capacities or enhancement of old ones, without sacrifice of preexisting capacities; and (b) the changes are implemented gradually over an extended period of time; (c) each step of the transformation process is freely and competently chosen by the subject; and (d) the new capacities do not prevent the preexisting capacities from being periodically exercised; (e) the subject retains her old memories and many of her basic desires and dispositions; (f) the subject retains many of her old personal relationships and social connections; and (g) the transformation fits into the life narrative and self-conception of the subject. Posthuman cognitive and emotional capacities could in principle be acquired in such a way that these conditions are satisfied.

Even if not all the conditions (a)-(g) were fully satisfied in some particular transformation process, the normatively relevant elements of a person's (numerical or narrative) identity could still be *sufficiently* preserved to avoid raising any fundamental identity-based objection to the prudentiality of undergoing such a transformation. We should not use a stricter standard for technological self-transformation than for other kinds of human transformation, such as migration, career change, or religious conversion.

Consider again a familiar case of *radical* human transformation: maturation. You currently possess vastly greater cognitive capacities than you did as an infant. You have also lost some capacities, e.g. the ability to learn to speak a new language without an accent. Your emotional capacities have also changed and developed considerably since your babyhood. For each concept of identity which we might think has relevant normative significance – personal (numerical) identity, narrative identity, identity of personal character, or identity of core characteristics – we should ask whether identity in that sense has been preserved in this transformation.

The answer may depend on exactly how we understand these ideas of identity. For each of them, on a sufficiently generous conception of the identity criteria, identity was completely or in large part preserved through your maturation. But then we would expect that identity in that sense would also be preserved in many other transformations, including the ones that are *no more profound* as that of a child growing into an adult; and this would include transformations that would make you posthuman. Alternatively, we might adopt conceptions that impose more stringent criteria for the preservation of identity. On these conceptions, it might be impossible to become posthuman without wholly or in large part disrupting one form of identity or another. However, on such restrictive conceptions, identity would also be disrupted in the transformation of child into adult. Yet we do not think it is bad for a child to grow up. Disruptions of identity in those stringent senses form part of a normal life experience and they do not constitute a disaster, or a misfortune of any kind, for the individual concerned.

Why then should it be bad for a person to continue to develop so that she one day matures into a being with posthuman capacities? Surely it is the other way around. If this had been our usual path of development, we would have easily recognized the failure to develop into a posthuman as a misfortune, just as we now see it as a misfortune for a child to fail to develop normal adult capacities.

Many people who hold religious beliefs are already accustomed to the prospect of an extremely radical transformation into a kind of posthuman being, which is expected to take place after the termination of their current physical incarnation. Most of those who hold such a view also hold that the transformation *could* be very good for the person who is transformed.

7. Commitments

Apart from the concern about personal identity, there is a second kind of reason why it might be bad for a Bloggs to become a Haydn. Bloggs might be involved in various projects, relationships, and may have undertaken commitments that he could not or would not fulfill if he became Haydn. It would be bad for Bloggs to fail in these undertakings if they are important to him. For

example, suppose that Mr. Bloggs is deeply committed to Mrs. Bloggs. His commitment to Mrs. Bloggs is so strong that he would never want to do anything that contravenes any of Mrs. Bloggs' most central preferences, and one of her central preferences is that Mr. Bloggs not become posthuman. In this case, even though becoming posthuman might in some respects be good for Mr. Bloggs (it would enable him to understand more, or to stay healthy longer, etc.) it might nevertheless be bad for him all things considered as it would be incompatible with fulfilling one of the commitments that are most important to him.³⁷

This reason for thinking that it might be bad for a person to become posthuman relies on the assumption that it can be very bad for a person to forfeit on commitments that would be impossible to fulfill as a posthuman.³⁸ Even if we grant this assumption, it does not follow that becoming a posthuman would necessarily be bad for us. We do not generally have commitments that would be impossible to fulfill as posthumans. It may be impossible for Mr. Bloggs to become posthuman without violating his most important commitment (unless, of course, Mrs. Bloggs should change her mind), but his is a special case.

Some humans do not have any commitments of importance comparable to that of Mr. Bloggs to his wife. For such people the present concern does not apply. But even for many humans who do have such strong commitments, becoming posthuman could still be good for them. Their commitments could still be fulfilled if they became posthuman. This is perhaps clearest in regard to our commitments to projects and tasks: most of these we could complete – indeed we could complete them better and more reliably – if we obtained posthuman capacities. But even with regard to our specific commitments to people, it would often be possible to fulfill these even if we had much longer healthspans or greatly enhanced cognitive or emotional capacities.

8. Ways of life

In addition to concerns about personal identity and specific commitments to people or projects, there is a third kind of reason one might have for doubting that it could be good for us to become posthuman. This third kind of reason has to do with our interpersonal relations more broadly, and with the way that the good for a person can be tied to the general circumstances and conditions in which she lives. One might think that the very concept of a good life for a human being is inextricably wound up in the idea of flourishing within a “way of life” – a matrix of beliefs, relationships, social roles, obligations, habits, projects, and psychological attributes outside of which the idea of a “better” or “worse” life or mode of being does not make sense.

The reasoning may go something like this: It would not be good for a clover to grow into a rhododendron, nor for a fly to start looking and behaving like a raven. Neither would it, on this view, be good for a human to acquire posthuman capacities and start living a posthuman life. The criterion for how well a clover is doing is the extent to which it is succeeding in realizing its own particular nature and achieving the natural “telos” inherent in the clover kind; and the equivalent might be said of the fly. For humans, the case may be more complicated as there is a greater degree of relevant individual variation among humans than among other species. Different

³⁷ We may include under this rubric any “commitments to himself” that Mr. Bloggs might have. For example, if he has a firm and well-considered desire not to become posthuman, or if he has solemnly sworn to himself never to develop any posthuman capacities, then it could perhaps on grounds of these earlier desires or commitments be bad for Mr. Bloggs to become posthuman.

³⁸ One may also hold that a person in Mr. Bloggs' situation has additional reasons for not becoming posthuman that don't rely on it being worse for him to become posthuman. For instance, he might have moral reasons not to become posthuman even if it would be good for him to become one. Here I am concerned with the question whether it would necessarily be bad *for* Bloggs to become posthuman, so any moral reasons he might have for declining the transition would only be relevant insofar as they would make the outcome worse *for* Mr. Bloggs.

humans are pursuing different “ways of life”, so that what counts as flourishing for one human being might differ substantially from what counts as such for another. Nevertheless, as we are all currently pursuing human ways of life, and since what is good for us is defined by reference to our way of life, it is not the case for any human that it would be good for her to become posthuman. It might be good for posthumans to be posthumans, but it would not be good for humans to become posthuman.

This third concern seems to be a conglomerate of the two concerns we have already discussed. Why could it not be good for a human to become posthuman? One possible reason is if her personal identity could not be preserved through such a transformation. The comparison with the clover appears to hint at this concern. If a clover turned into a rhododendron, then the clover would presumably cease to exist in the process. If a fly started looking and behaving like a raven, would it still be a fly? So part of what is going on here seems to be that the assertion that the relevant form of identity could not be preserved in the transformations in question. But we have already addressed this concern insofar as it pertains to humans becoming posthuman.

There might be more at stake with this third concern than identity. The problem with a clover becoming a rhododendron is not just that the clover might cease to exist in the process, but that it seems a mistake to think that being a rhododendron is in any sense better than being a clover. There might be external criteria of evaluation (such as economic or aesthetic value to the human owner) according to which a rhododendron is better or more valuable than a clover. But aside from such extrinsic considerations, the two plants seem to be on a par: a thriving clover thrives just as much as a thriving rhododendron, so if the good for a plant is to thrive then neither kind is inherently better off or has a greater potential for realizing a good life than the other. Our challenger could claim that the same holds vis-à-vis a human and a posthuman.

I think the analogy is misleading. People are not plants, and the concept of a valuable mode of being for a person is fundamentally different from that of the state of flourishing for a plant. In a metaphorical sense we can ascribe interests to plants and other non-sentient objects: this clover “could use” some water; that clock “needs” winding up; the squeaky wheel “would benefit” from a few drops of oil. Defining interests relative to a functionalist basis might be the only way we can make sense of these attributions. The function of the clock is to indicate the time, and without being wound up the clock would fail to execute this function; thus it “needs” to be wound up. Yet sentient beings may have interests not only in a metaphorical sense, based on their function, but in a quite literal sense as well, based on what would be normatively good for them. A human being, for example, might have interests that are defined (partially) in terms of what she is actually interested in, or would be interested in given certain conditions.³⁹ So from the fact that we could not make sense of the claim that it would be good for a clover to become a rhododendron, it does not follow that we would similarly be unable to make sense of the claim that it would be good for a human to become a posthuman. Even if the successful execution of “the function” of a human were not facilitated by becoming posthuman, there would be other grounds on which one could sensibly attribute to a human an interest in becoming posthuman.

It is at any rate highly problematic that something as complex and autonomous as a human being has any kind of well-defined “function”. The problem remains even if we relativize the function to particular ways of life or particular individuals. We might say that the function of the farmer is to farm, and that of the singer is to sing, etc. But any particular farmer is a host of other things as well: e.g. a singer, a mother, a sister, a homeowner, a driver, a television watcher, and so forth *ad infinitum*. Once she might have been a hairdresser; in the future she might become a shopkeeper, a golfer, a person with a disability, a transsexual, or a posthuman. It is difficult to see how any strong normative conclusions could be drawn from the fact that she currently occupies a certain set of roles and serves a certain set of functions. At most we could conclude that when and insofar as she acts as a farmer, she ought to tend to her crops or livestock; but from

³⁹ Cmp. dispositional theories of value, discussed above.

the fact that she is a farmer, nothing follows about whether she ought to be or remain a farmer. Likewise, the most we could conclude from the fact that she is currently a human person is that she ought to do things that are good for humans – brush her teeth, sleep, eat, etc. – but only so long as she remains human. If she became a posthuman who did not need to sleep, she would no longer have any reason so sleep. And the fact that she currently has a reason to sleep is not a reason for her not to become a sleepless posthuman.

At this point, an objector could attempt an alternative line of argumentation. Maybe there are some crucial interests that we have that are not conditional on us occupying particular social roles or having particular personal characteristics or serving particular functions. These interests would be unlike our interest in sleep, which does not provide us with a reason not to change in such a way that we no longer need to sleep. Rather these unconditional (“categorical”) interests would be such as to give us reason not to change in ways that would make us no longer have those interests.

I have already admitted that individuals can have such interests, and in some cases this might make it the case for some possible individuals that it would not be good for them to become posthuman. I discussed this above as the “second concern”. This is not a problem for my position since it is compatible with it being true for other individuals (and perhaps for the overwhelming majority or even all actual human persons) that it could be good for them to become posthuman. But our hypothetical objector might argue that there are certain categorical interests we all have *qua* humans. These interests would somehow derive from human nature and from the natural ends and ideals of flourishing inherent in this essential nature. Might not the existence of such universally shared categorical human interests invalidate the thesis that it could be good for us to become posthuman?

9. Human nature

Let us consider two different candidate ideas of what a human “telos” might be.

If we seek a telos for human individuals within a naturalistic outlook, one salient candidate would be the maximization of that individual’s inclusive fitness. Arguably, the most natural way to apply a functional characterization of a human individual from an evolutionary perspective is as an inclusive fitness maximizer (tuned for life in our ancestral environment). From this perspective, our legs, our lungs, our sense of humor, our parental instincts, our sex drive and romantic propensities subserve the ultimate function of promoting the inclusive fitness of an individual. Now if we define the telos of a human individual in this way, as vehicle for the effective promulgation of her genes, then many of the seemingly most attractive posthuman possibilities would be inconsistent with our successfully realizing this alleged telos, in particular those possibilities that involve radical alteration of the human genome. (Replacing our genes with other genes does not seem to be an effective way to promulgate the genes we have.)

As a conception of the human good, however, the telos of maximizing inclusive fitness is singularly lacking in plausibility. I do not know of any moral philosopher who advocates such a view. It is too obvious that what is good for a person can, and usually does, diverge from what would maximize that person’s inclusive fitness.⁴⁰ Those who attempt to derive a theory of the human good from the telos inherent in a conception of human functioning will need to start from some conception of human functioning other than the evolutionary one.

One starting point that has had more appeal is the doctrine that a human being is essentially a rational animal and that the human telos is the development and exercise of our rational faculties. Views of this sort have a distinguished pedigree that can be traced back at least to Aristotle. Whatever the merits of this view, however, it is plainly not a promising objection to

⁴⁰ For example, for a contemporary man the life plan that would maximize inclusive fitness might be to simply donate as much sperm to fertility clinics as possible.

the claims I advance in this paper, since it would be perfectly possible for a posthuman to realize a telos of rationality as well as a human being could. In fact, if what is good for us is to develop and exercise our rational nature, this implies that it would be good for us to become posthumans with appropriately enhanced cognitive capacities (and preferably with extended healthspan too, so that we may have more time to develop and enjoy these rational faculties).

One sometimes hears it said that it is human nature to attempt to overcome every limit and to explore, invent, experiment, and use tools to improve the human condition.⁴¹ I don't know that this is true. The opposite tendency seems to be at least as strong. Many a great invention was widely resisted at the time of its introduction, and inventors have often been viciously persecuted. If one wished to be provocative, one might even say that humanity has advanced technologically *in spite of anti-technological tendencies in human nature*, and that technological advancement historically has been due more to the intrinsic utility of technological inventions and the competitive advantages they sometimes bestow on their users than to any native preference among the majority of mankind for pushing boundaries and welcoming innovation.⁴² Be that as it may; for even if it were "part of human nature" to push ever onward, forward, and upward, I do not see how anything follows from this regarding the desirability of becoming posthuman. There is too much that is thoroughly unrespectable in human nature (along with much that is admirable), for the mere fact that X is a part of human nature to constitute any reason, even a *prima facie* reason, for supposing that X is good.

11. Brief sketches of some objections and replies

Objection: One might think that it would be bad for a person to be the only posthuman being since a solitary posthuman would not have any equals to interact with.

Reply: It is not necessary that there be only one posthuman.

I have acknowledged that capacities may not have basic intrinsic value and that the contribution to well-being that having a capacity makes depends on the context. I suggested that it nevertheless makes sense to talk of the value of a capacity in a sense similar to that in which we commonly talk of the value of e.g. money or health. We can take such value ascriptions as assertions that the object or property *normally* makes a positive contribution to whatever has basic value. When evaluating posthuman attributes, the question arises what we should take to be the range of circumstances against which we assess whether something "normally" makes a positive contribution. As we do not have a concrete example in front of us of a posthuman civilization, there is a certain indeterminacy in any assertion about which things or attributes would "normally" make a positive contribution in a posthuman context. At this point, it may therefore be appropriate to specify some aspects of the posthuman context that I assume in my value-assertions. Let me here postulate that the intended context is one that includes a society of posthuman beings.

⁴¹ The quest for posthuman capacities is as old as recorded history. In the earliest preserved epic, the Sumerian *Epic of Gilgamesh* (approx. 1700 B.C.), a king sets out on a quest for immortality. In later times, explorers sought the Fountain of Youth, alchemists labored to concoct the Elixir of Life, and various schools of esoteric Taoism in China strove for physical immortality by way of control over or harmony with the forces of nature. This is in addition to the many and diverse religious traditions in which the hope for a supernatural posthuman existence is of paramount importance.

⁴² As J.B.S. Haldane wrote: "The chemical or physical inventor is always a Prometheus. There is no great invention, from fire to flying, which has not been hailed as an insult to some god. But if every physical and chemical invention is a blasphemy, every biological invention is a perversion. There is hardly one which, on first being brought to the notice of an observer from any nation which has not previously heard of their existence, would not appear to him as indecent and unnatural." (Haldane 1924).

What dialectical constraints are there on what I am allowed to stipulate about the posthuman context? The main cost to making such stipulations is that if I end up defining a gerrymandered “posthuman context”, which is also extremely unlikely ever to materialize, then the significance of any claims about what would normally be valuable in that context would tend to wane. It is simply not very interesting to know what would “normally” be valuable in some utterly bizarre context defined by a large number of arbitrary stipulations. I do not think that by postulating a society of posthumans I am significantly hollowing out my conclusions. I do, in fact, assume throughout this paper more generally that the postulated posthuman reference society is one that is adapted to its posthuman inhabitants in manners similar to the way current human society is adapted to its human inhabitants.⁴³ I also assume that this reference society would offer many affordances and opportunities to its posthuman inhabitants broadly analogous to those which contemporary society offers humans. I do not intend by this postulation to express any prediction that this is the kind of posthuman society that is most likely to form, nor do I mean to imply that being a posthuman could not be valuable even outside of the context of such a kind of society. The postulation is merely a way of delimiting the claims I am trying to defend in this paper.

Objection: The accumulated cultural treasures of humanity might lose their appeal to somebody whose capacities greatly exceeded those of the humans who produced them. More generally, challenges that seemed interesting to the person while she was still human might become trivial and therefore uninteresting to her when she acquires posthuman capacities. This could deprive posthumans of the good of meaningful achievements.

Reply: It is not clear why the ability to appreciate what is more complex or subtle should make it impossible to appreciate simpler things. Somebody who has learnt to appreciate Schoenberg may still delight in simple folk songs, even bird songs. A fan of Cézanne may still enjoy watching a sunrise.

Even if it were impossible for posthuman beings to appreciate some simple things, they could compensate by creating new cultural riches. I am assuming that the reference society would offer opportunities for doing this – see above.

If some challenges become too easy for posthumans, they could take on more difficult challenges. One might argue that an additional reason for developing posthuman cognitive capacities is that it would increase the range of interesting intellectual challenges open to us. At least within the human range of cognitive capacity, it seems that the greater one’s capacity, the more numerous and meaningful the intellectual projects that one can embark on. When one’s mind grows, not only does one get better at solving intellectual problems – entirely new possibilities of meaning and creative endeavor come into view.

Objection: A sense of vulnerability, dependence, and limitedness can sometimes add to the value of a life or help a human being grow as a person, especially along moral or spiritual dimensions.

Reply: A posthuman could be vulnerable, dependent, and limited.

A posthuman could also be able to grow as a person in moral and spiritual dimensions without those extrinsic spurs that are sometimes necessary to affect such growth in humans. The ability to spontaneously develop in these dimensions could be seen as an aspect of emotional capacity.

Objection: The very desire to overcome one’s limits by the use of technological means rather than through one’s own efforts and hard work could be seen as expressive of a failure to open

⁴³ But I do not assume that the reference society would only contain posthuman beings.

oneself to the unbidden, gifted nature of life, or as a failure to accept oneself as one is, or as self-hate.⁴⁴

Reply: This paper makes no claims about the expressive significance of a desire to become posthuman, or about whether having such a desire marks one as a worse person, whether necessarily or statistically. The concern here rather is about whether being posthuman could be good, and whether it could be good for us to become posthuman.

Objection: A capacity obtained through a technological shortcut would not have the same value as one obtained through self-discipline and sacrifice.

Reply: I have argued that the possession of posthuman capacities could be extremely valuable even were the capacities are effortlessly obtained. It is consistent with what I have said that achieving a capacity through a great expenditure of blood, sweat, and tears would further increase its value. I have not addressed what would be the *best* way of becoming posthuman. We may note, however, that is unlikely that we *could* in practice become posthuman other than via recourse to advanced technology.

Objection: The value of achieving a goal like winning a gold medal in the Olympics is reduced and perhaps annulled if the goal is achieved through inappropriate means (e.g. cheating). The value of possessing a capacity likewise depends on how the capacity was acquired. Even though having posthuman capacities might be extremely valuable if the capacities had been obtained by appropriate means, there are no humanly possible means that are appropriate. Any means by which humans could obtain posthuman capacities would negate the value of having such capacities.

Reply: The analogy with winning an Olympic medal is misleading. It is in the nature of sports competitions that the value of achievement is intimately connected with the process by which it was achieved. We may say that what is at stake in the analogy is not really the value of a medal, nor even the value of winning a medal, but rather (something like) winning the medal by certain specified means in a fair competition, in a non-fluke-like way, etc. Many other goods are not like this. When we visit the doctor in the hope of getting well, we do not usually think that the value of getting well is strongly dependent on the process by which health is achieved; health and the enjoyment of health are valuable in their own right, independently of how these states come about. Of course, we are concerned with the value of the means to getting well – the means themselves can have negative value (involving perhaps pain and inconvenience), and in evaluating the value of the consequences of an action, we take the value of the means into account as well as the value of the goal that they achieve. But usually, the fact that some means have negative value does not reduce the value of obtaining the goal state.

One possible exception to this is if the means are in a certain sense *immoral*. We might think that a goal becomes “tainted”, and its value reduced, if it was achieved through deeply immoral means. For example, some might hold that the value of medical findings obtained by Nazi doctors in concentration camps have reduced or no value because of the way the findings were produced. Yet this radical kind of “taint” is a rather special case.⁴⁵ Having to use bad means might be good reason not to pursue a goal, but *typically* this is not because the use of bad means

⁴⁴ Cmp. (Sandel 2004), although it is not clear that Sandel has an expressivist concern in mind.

⁴⁵ Even in the Nazi doctor example, it is plausibly the *achievement* of the doctors (and of Germany etc.) that is tainted, and the *achievement's* value that is reduced. The value of the *results* is arguably unaffected, although it might always be appropriate to feel uncomfortable when employing them, appropriate to painfully remember their source, regret the way we got them, and so forth.

would reduce the value of the attainment of the goal, but rather it is either because the means themselves have more negative value than the goal has positive value, or (on a non-consequentialist view) because it is morally impermissible to use certain means independently of the total value of the consequences.⁴⁶

The values that I have alleged could be derived from posthuman capacities are not like the value of an Olympic gold medal, but rather like the value of health. I am aware of no logical, metaphysical, or “in principle” reason why humans could not obtain posthuman capacities in ways that would avoid recourse to immoral means of the sort that would “taint” the outcome (much less that would taint the outcome to such a degree as to annul its extremely high surplus value). It is a further question to what extent it is *practically feasible* to work towards realizing posthuman capacities in ways that avoid such taint. This question lies outside the scope of the present paper. My conclusion may therefore be understood to implicitly contain the proviso that the posthuman capacities of which I speak have been obtained in ways that are non-Faustian.

Objection: Posthuman talent sets the stage for posthuman failure. Having great potential might make for a great life if the potential is realized and put to some worthwhile use, but it could equally make for a tragic life if the potential is wasted. It is better to live well with modest capacities than to life poorly with outstanding capacities.

Reply: We do not lament that a human is born talented on grounds that it is possible that she will waste her talent. It is not clear why posthuman capacity would be any more likely to be wasted than human capacity. I have stipulated that the posthuman reference society would offer affordances and opportunities to its posthuman inhabitants broadly analogous to those that contemporary society offers humans. If posthumans are more prone to waste their potential, it must therefore be for internal, psychological reasons. But posthumans need not be any worse than humans in regard to their readiness to make the most of their lives.⁴⁷

12. Conclusion

I have argued, first, that some posthuman modes of being would be extremely worthwhile; and, second, that it could be good for most human beings to become posthuman.

I have discussed three general central capacities – healthspan, cognition, and emotion – separately for most of this paper. However, some of my arguments are strengthened if one considers the possibility of combining these enhancements. A longer healthspan is more valuable when one has the cognitive capacity to find virtually inexhaustible sources meaning in creative

⁴⁶ It might help to reflect that we do not deny the value of our current human capacities on grounds of their evolutionary origin, even though this origin is (a) largely not a product of human achievement, and (b) fairly drenched in violence, deceit, and undeserved suffering. People who are alive today also owe their existence to several thousands of years of warfare, plunder, and rape; yet this does not entail that our capacities or our mode of existence is worthless.

Another possibility is that the result has positive value X, the way you get it has negative value Y, but the “organic whole” comprising both the result and the way it was obtained has an independent value of its own, Z, which also might be negative. On a Moorean view, the value of this situation “on the whole” would be $X+Y+Z$, and this might be negative even if X is larger than $(-Y)$ (Moore 1903). Alternatively, Z might be incommensurable with $X+(-Y)$. In either case, we have a different situation than the one described above in the text, since here X could invariant under different possible ways in which the result was obtained. However, I do not know of any reason to think that this evaluative situation, even if axiologically possible, would necessarily obtain in the sort of case we are discussing. (I’m indebted to Guy Kahane for this point.)

⁴⁷ If they have enhanced emotional capacity, they may be *more* motivated and more capable than most humans of realizing their potential in beautiful ways.

endeavors and intellectual growth. Both healthspan and cognition are more valuable when one has the emotional capacity to relish being alive and to take pleasure in mental activity.

It follows trivially from the definition of “posthuman” given in this paper that we are not posthuman at the time of writing. It does not follow, at least not in any obvious way, that a posthuman could not also remain a human being. Whether or not this is so depends on what meaning we assign to the word “human”. One might well take an expansive view of what it means to be human, in which case “posthuman” is to be understood as denoting a certain possible type of human mode of being – if I am right, an exceedingly worthwhile type.⁴⁸

References

- American Psychiatric Association. (2000). *Diagnostic criteria from DSM-IV-TR*. Washington, D.C., American Psychiatric Association.
- Arnesen, T. M. and O. F. Norheim (2003). "Quantifying quality of life for economic analysis: time out for time trade off." *Medical Humanities* 29(2): 81-86.
- Beauchamp, T. L. and J. F. Childress (2001). *Principles of biomedical ethics*. New York, N.Y., Oxford University Press.
- Berkeley, G., G. Sampson, et al. (1897). *The works of George Berkeley, D.D., bishop of Cloyne*. London,, G. Bell and Sons.
- Bostrom, N. (2003). "The Transhumanist FAQ: v 2.1." World Transhumanist Association. Available from <http://transhumanism.org/index.php/WTA/faq/>.
- Bostrom, N. (2004). Transhumanist Values. *Ethical Issues for the 21st Century*. F. Adams, Philosophical Documentation Center Press.
- Bostrom, N. (2005). "The Fable of the Dragon-Tyrant." *Journal of Medical Ethics* 31(5): 273-277.
- Brickman, P. and D. T. Campbell (1971). Hedonic relativism and planning the good society. *Adaptation-level theory: A symposium*. M. H. Apley. New York, Academic Press: 287-301.
- Cohen, J. and G. Langer (2005). "Most Wish for a Longer Life -- Despite Broad Aging Concerns." ABC News/USA Today Poll. Available from <http://abcnews.go.com/images/Politics/995a1Longevity.pdf>.
- Damschroder, L. J., B. J. Zikmund-Fisher, et al. (2005). "The impact of considering adaptation in health state valuation." *Social Science & Medicine* 61(2): 267-277.
- DeGrazia, D. (2005). "Enhancement Technologies and Human Identity." *Journal of Medicine and Philosophy* 30: 261-283.
- Diener, E. and E. M. Suh (1998). "Subjective well-being and age: An international analysis." *Annual Review of Gerontology and Geriatrics* 17: 304-324.
- Glannon, W. (2002). "Identity, prudential concern, and extended lives." *Bioethics* 16(3): 266-83.
- Haldane, J. B. S. (1924). *Daedalus; or, Science and the future*. London,, K. Paul, Trench, Trubner & co., ltd.
- Hanson, R. (2000). "Showing That You Care: The Evolution of Health Altruism." Available from <http://hanson.gmu.edu/showcare.pdf>.
- Hayles, N. K. (1999). *How we became posthuman: virtual bodies in cybernetics, literature, and informatics*. Chicago, Ill., University of Chicago Press.
- Johansson, P. O. (2002). "On the definition and age-dependency of the value of a statistical life." *Journal of Risk and Uncertainty* 25(3): 251-263.

⁴⁸ I am grateful to Ross Beaton, Bert Gordijn, Guy Kahane, Toby Ord, David Pearce, David Rodin, Anders Sandberg, Julian Savulescu, Harrosh Shlomit, and Elena Patigo Solana for helpful comments. Earlier versions of this paper were presented at the James Martin Advanced Research Seminar (Oxford, 30 January 2006), and at the Institute for Science, Innovation & Society (Nijmegen, 21 February 2006).

- Johnson, F. R., W. H. Desvousges, et al. (1998). "Eliciting stated health preferences: An application to willingness to pay for longevity." *Medical Decision Making* 18(2): S57-S67.
- Kunzmann, U., T. Little, et al. (2000). "Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study." *Psychology and Aging* 15(3): 511-426.
- Lewis, D. (1989). "Dispositional Theories of Value." *Proceedings of the Aristotelian Society, supp.* 63: 113-137.
- McShine, R., G. T. Lesser, et al. (2000). "Older Americans hold on to life dearly." *British Medical Journal* 320(7243): 1206-1207.
- Moore, G. E. (1903). *Principia Ethica*. Cambridge, Cambridge University Press.
- Murphy, K. and R. Topel (2003). *The Economic Value of Medical Research. Measuring the gains from medical research: an economic approach*. K. M. Murphy and R. H. Topel. Chicago, University of Chicago Press.
- Nordhaus, W. (2003). *The Health of Nations: The Contribution of Improved Health to Living Standards. Measuring the gains from medical research: an economic approach*. K. M. Murphy and R. H. Topel. Chicago, University of Chicago Press: vi, 263 p.
- Pearce, D. (2004). "The Hedonistic Imperative." Available from <http://www.hedweb.com/hedab.htm>.
- Sandel, M. (2004). "The Case Against Perfection." *The Atlantic Monthly* 293(3).
- Tsevat, J., N. V. Dawson, et al. (1998). "Health values of hospitalized patients 80 years or older." *Jama-Journal of the American Medical Association* 279(5): 371-375.
- Ubel, P. A., G. Loewenstein, et al. (2003). "Whose Quality of Life? A Commentary exploring discrepancies between health state evaluations of patients and the general public." *Quality of Life Research* 12(6): 599-607.
- Viscusi, W. K. and J. E. Aldy (2003). "The value of a statistical life: A critical review of market estimates throughout the world." *Journal of Risk and Uncertainty* 27(1): 5-76.
- World Bank. (2003). "EdStats - the World Bank's comprehensive Database of education statistics." 2006. Available from <http://www1.worldbank.org/education/edstats/index.html>.