

Philosophy as a Transhumanistic Discipline

Jon Garthoff
Northwestern University

1. Introduction¹

The emergence of biomedical nanotechnologies raises the prospect of a very different life for human beings in the not-so-distant future. It is no longer difficult to imagine a world in which these technologies are used to greatly increase human strength and intelligence, to greatly improve human psychological health, and to greatly extend the duration of human life. A loosely organized international movement known as “transhumanism” has formed in recognition of this prospect;² transhumanists advocate investing in these emerging nanotechnologies and doing whatever is fiscally and technologically feasible to make human life longer and more enjoyable and to make human beings more powerful and durable.³

I am broadly sympathetic with this transhumanist agenda. I see great potential and no inherent wrong in pursuing these technologies; and while the issue of investment is complicated by the opportunity costs of not investing elsewhere – in more conventional medical technologies like the development and propagation of vaccines, for example – it is plausible that significant investment in biomedical nanotechnologies is good social policy.⁴ One purpose of this essay is to defend the permissibility and advisability of the transhumanistic project. Another purpose of this essay is to explore the ways in which issues raised by the emergence of biomedical nanotechnologies illuminate the traditional topics of ethical theory: the nature of the good for human beings and what we should make of the basic facts of the human condition.

In order to clarify what it is to be a transhumanist, I will also distinguish and explicate several senses in which one might be a “humanist” and argue that there is no sense of “humanism” that we should endorse which would present a principled objection to transhumanism. Part of this effort is an attempt to rebut several consequences of the philosopher Bernard Williams’s account of human nature, which is perhaps most explicitly stated in his posthumously published collection of essays, *Philosophy as a Humanistic Discipline*. The view of Williams’s that I am most concerned to reject is his apology for what he labels “The Human Prejudice” in an essay of that title. The human prejudice is the claim that being human is a significant ethical consideration in itself, and not merely because human beings typically have capacities like sentience and rationality that entitle them to respect. I argue against the human prejudice that, in the context of a discussion concerning the possibility of altering human nature itself, the human prejudice is exposed as excessively conservative and parochial. I conclude by returning to the topic of the basic facts of the human condition in order to present a novel argument in favor of transhumanism; this argument relies crucially on the claim that working towards transhumanist goals makes human life less tragic.

2. The Human Prejudice, Part I: Human Nature

The first sense of “humanism” I want to introduce is the view that we should endorse the human prejudice. This is the view, introduced in the previous paragraph, that membership in the human species is an ethically significant category apart from the capacities, such as rationality and sentience, that human beings typically have. In Section 4 I present the considerations that Williams uses to support the human prejudice, but I

want first to consider an older route to this thesis. This route argues to the human prejudice from a different view that could be called “humanism”: the thesis that what is good for human beings is largely or entirely determined by their common human nature. If this view is correct, then it looks like there is no way to avoid the human prejudice, since membership in the human species not only is a relevant ethical category but also generates other significant ethical categories, such as what is good for human beings.

It is important in assessing the plausibility of this view, however, to distinguish between a weaker and a stronger formulation of it. A weak version of the thesis proceeds from the bottom up, so to speak, by noting that there are features of crucial significance to what is good for individual persons that are widely shared among members of the human species. Our shared capacities for sentience and rationality figure importantly in what is good for us, for example, since they explain why the experience of pleasure and success in our aims is good for us. Other typical human features include our lifespan, psychological constitution, and range of abilities such as strength and intelligence. The typical length of human life helps explain why it is good for us to have projects with the scope and duration of careers and our psychological constitution helps explain why it is good for us to have friends. No one should deny that human nature, understood in this way, significantly determines what is good and bad for us.

But this is a very weak notion of human nature, for it is compatible with the claim that human nature *as such* plays no essential role in determining what our good consists in. Human beings typically share a roughly equal lifespan and a roughly similar emotional constitution, and so typically should value careers and relationships of a certain sort; but if a person is an outlier with respect to human nature in these areas – if

she has an exceptionally short lifespan as a result of childhood leukemia, for example, or if she has an exceptionally distant emotional relationship to others as a result of severe Asperger's Syndrome – then what makes a career or friendship good for her will be different. What is good for human beings generally, therefore, may not be good for each particular individual. The bottom-up appeal to human nature fails to justify the human prejudice, for it makes no use of a specifically *human* nature – as opposed to a rational nature, a sentient nature, a particular sort of psychological constitution, or a particular set of physical characteristics – in accounting for what is good for individuals. The bottom-up appeal to human nature is thus no obstacle to transhumanism, since there is no reason why this sort of view would claim that it is inherently bad to be an outlier with respect to what is presently typical for human beings.

We have to be careful here, of course, because we need to vindicate the claim that having childhood leukemia or severe Asperger's is typically bad for people. But an account of why these diseases are bad need not make appeal to human nature. It could be that having childhood leukemia is bad because one's life is shorter, independently of what span of life is typical for human beings. And it could be that having more and closer friendships is good for people, regardless of what sorts of friendships human beings typically have. Since the transhumanist advocates expanding, and not contracting, our range of opportunities – enabling people to live longer and to have a greater variety of projects – here again we encounter no obstacle to the transhumanist agenda.

There is a stronger conception of human nature – what we might call a top-down conception – according to which outliers with respect to what is typical are for that reason worse off. Such a conception could potentially vindicate the human prejudice,

since it appeals to human nature itself in accounting for what is good for us, rather than appealing to independently articulated features that most humans share. This top-down conception could also potentially challenge arguments for transhumanism, insofar as the transhumanist strives to create conditions where more human beings are atypical by present standards.

This conception of human nature is difficult to believe, however, in light of the discoveries of modern science. Evolutionary theory exposes that human nature is contingent and that it changes significantly over time. In order to appeal to human nature to oppose transhumanism, then, a relevant difference must be articulated between changes in human nature that occur through the unplanned mechanisms of natural selection and those that occur through planned technological advances; and it is not clear what such a relevant difference could be.

3. Interlude: Secular Humanism

One consequence of the discoveries of modern science that render the strong conception of human nature implausible has been the rise of a different sort of “humanism.” This is the view sometimes known as “secular humanism,” and as Williams notes it is probably the most commonly used sense of the term “humanist” in ordinary discourse. A humanist of this sort is an atheist who is optimistic about human prospects, if human beings are able to free themselves from what the humanist regards as the shackles of religious modes of thought.

I will not address the merits of this brand of humanism here; but I think it is worthwhile to briefly digress on Williams’s discussion of secular humanism in the

opening paragraphs of “The Human Prejudice.” There Williams articulates a curious reservation about secular humanism. He writes that this view “encounters an immediate and very obvious paradox,” since

the outlook is stuck with the fact that on its own submission this evil, corrupting, and pervasive thing, religion, is itself a human invention: it certainly did not come from anywhere else. So humanists in this atheist sense should ask themselves: if humanity invented something as awful as they take religion to be, what should that tell them about humanity? In particular, can humanity really be expected to do much better without it?⁵

I say this reservation is curious because it risks entailing far too much to be an important problem for secular humanism. One way to see what I have in mind is to insert “slavery,” “monarchy,” or any other to-be-eliminated social practice everywhere in Williams’s comments where “religion” appears. Humanity produced all of those ills, but it scarcely follows that humanity could not be expected to do better without them.⁶

This feature of Williams’s thought is, moreover, connected to his reasons for endorsing the human prejudice, which I take up in the following section. Williams would not endorse the strong conception of human nature discussed earlier, but I believe that his reservations about secular humanism also result in part from his conceiving of human nature in static terms. We can infer that human beings will be no better off after abolishing a harmful institution of their own creation only if we assume that human beings are unable to reform their own natures by reforming themselves and their practices. The relevant changes here would be produced by social evolution, not biological evolution; but in each case it is the worry that such changes are impossible that fuels reservations about advocating them. I share with the transhumanist the opposing view that human nature can be improved, both through conscious social planning and through unplanned evolutionary and market mechanisms. I also share some of

Williams's pessimism about the likelihood of success in this effort, but I do not regard that as a conclusive reason not to try.

4. The Human Prejudice, Part Two: Cosmic Significance

Thus far we have considered three distinct senses of “humanism:” the thesis that membership in the human species is an ethically significant category apart from the capacities that human beings typically share, the thesis that the good of individual human beings is largely or entirely determined by their shared human nature, and the thesis that religion is a harmful institution that human beings would be better off without. I have raised doubts about the truth of the first thesis and about the truth of a strong version of the second thesis, and I have noted that the third thesis and a weak version of the second thesis pose no obstacle to transhumanism. I now want to introduce a fourth sense of “humanism,” which is the thesis that “in cosmic terms human beings have a definite measure of importance.”⁷ Williams articulates this thesis and identifies two different versions of it, which he labels “Petrarchan” and “Lutheran.”

The Petrarchan thesis maintains that human beings are cosmically important because of our excellence or because of our privileged position in a divine plan for the world;⁸ the Lutheran maintains that human beings are “hideously fallen” and hence are of special concern because of our uniquely atrocious nature.⁹ As Williams observes, both versions of the thesis are difficult to believe in light of modern scientific knowledge. Evolutionary theory exposes the contingency of the existence of human beings, the gross inefficiency of the processes that generated human beings, and the fact that alternative environmental conditions might have generated more excellent creatures than human

beings. These facts do not sit comfortably with the Petrarchan thesis. The Lutheran thesis fares little better, since the randomness and contingency of human nature appear incompatible with the claim that we are uniquely atrocious creatures.

Williams is thus not concerned to defend humanism of either the Petrarchan or the Lutheran variety. He believes, however, that many have drawn a mistaken inference from the falsity of these views regarding the cosmic significance of human beings. He cites the philosopher Bertrand Russell's comments concerning "the transitoriness of human beings, the tininess of the earth, the vast and pitiless expanses of the universe and so on" as a misunderstanding of the lessons of modern science; and he cites with approbation the mathematician Frank Ramsey's reply to Russell that "he himself was much less impressed than some of his friends ... by the size of the universe, perhaps because he weighed 240 pounds."¹⁰ The error in Russell's thinking, according to Williams, is that it involves "a muddle between thinking that our activities fail some test of cosmic significance, and ... recognizing that there is no such test."¹¹ The former thought evinces a feeling that human activities are absurd or meaningless, while the latter simply evinces a feeling that human activities are neither more nor less meaningful than they ordinarily seem to be. He endorses the latter view and claims that it is nicely captured by Friedrich Nietzsche when Nietzsche writes that "once upon a time there was a star in a corner of the universe, and a planet circling that star, and on it some clever creatures who invented knowledge; and then they died, and the star went out, and it was as though nothing had happened".¹²

I raise these issues not to assess this Nietzschean view of the meaningfulness of human life, but rather to show how this view is connected to Williams's endorsement of

the human prejudice. Williams correctly notes that, once we give up on the idea of a cosmic perspective from which we may assess human activities, we can assess human activities only from our ordinary human perspective. But he infers from this, in my view too hastily, that the threshold of justification for the human prejudice is establishing that membership in the human species is “important *to us*.”¹³ If our perspective is the only one that there is, Williams’s thought must be, then all that is involved in justifying the human prejudice is showing how it makes sense of human practices to human beings.¹⁴

There is an alternative understanding, however, of what it is to assess a thesis from a human perspective. On a view common among but not limited to contemporary followers of Immanuel Kant, which we might label “immanent transcendentalism”, the human perspective contains the capacity to criticize the judgments typical of the exercise of human reason. This view is “transcendental” because it claims that human reason can critique itself by calling into question the sorts of judgments that human beings are inclined to make; and it is “immanent” because it denies the claim that this transcendental capacity entails an extrahuman perspective that we mysteriously assume when we make transcendental assessments. If immanence is compatible with transcendentalism, then more is involved in assessing the human prejudice than simply whether such a prejudice makes sense of human practices to human beings; a further question, concerning the rationality of accepting those practices, can be fruitfully raised and considered.

The mere articulation of this alternative view does not, of course, constitute a defense of it; and I will not attempt to provide a defense here. But there is value simply in articulating the alternative, I think, since it shifts the burden of argument. Williams cannot infer his controversial claims about the limits of the human perspective simply

from the absence of a nonhuman perspective on the world; he must also explain why the human perspective is incapable of radically criticizing human practices.

The immanent transcendentalist has an argumentative burden of her own. She is committed to the possibility of an ethical theory, understandable by and justifiable to human beings, that explains what is good and bad for them. To the extent that she can sketch the contours of such a theory, she will be on firmer ground in her dispute with Williams. More specifically, this theory will need either to simultaneously vindicate our reluctance to eat humans and our willingness to eat animals or to explain why one or the other of these practices should be reformed. This theory will also need to either vindicate the idea that severely mentally impaired adult human beings are entitled to treatment that we would not extend to all sentient animals or explain why our practice of extending them such treatment should be reformed. One way to construe Williams's endorsement of the human prejudice, then, is as an impossibility postulate for ethical theory. No theory that fails to incorporate species-membership as an important ethical category can vindicate our typical judgments and practices; and on his view there is no court of appeal outside these judgments and practices from which we might justify wholesale reform of them, and so according to him the human prejudice is justified.

We should be wary of this route to endorsing the human prejudice, however, since future theoretical developments are exceedingly difficult to anticipate. It may be thought that the leading traditions of modern ethical theory – utilitarianism and Kantianism – are incapable of meeting Williams's challenge. Utilitarianism, which is the primary focus of Williams's arguments in "The Human Prejudice,"¹⁵ emphasizes the ethical importance of sentience, which is the capacity to experience pleasure and pain. Since we share this

capacity with higher animals, many prominent critiques of utilitarianism proceed by arguing that the view is unable to account for what is distinctively valuable about human beings and their projects.¹⁶ Kantianism, by contrast, emphasizes rationality – an interrelated set of capacities including the capacity to conceptualize and respond to reasons, the capacity to formulate plans, and the capacity to conceive of oneself as existing over extended periods of time – and so is often thought to be incapable of explaining the ethical status of animals, infants, and severely mentally impaired adults.

These approaches to ethical theory are living traditions, however, and it is premature to judge that neither is able to vindicate the entire range of our ethical commitments. It is also possible that a hybrid theory might be developed that is able to capture the virtues of each approach without too many of the vices. In addition to these possibilities, it is also likely that new ethical theories, or novel variations on old ethical theories, will emerge. These new theoretical approaches may emphasize capacities distinct from both sentience and rationality that are in some sense intermediate between them. Two plausible candidates for such intermediate capacities include the capacity for purposive nonrational agency and the capacity for social interaction.¹⁷ These capacities might be crucial for explaining just those cases that Williams identifies as problematic, since infants have the capacity for some purposive agency and all conscious human beings have some capacity for social interaction with others. A fully developed theory of the value of these intermediate capacities might, for example, provide us with a way of claiming that infants and the severely mentally disabled have a moral status short of well-functioning mature humans but beyond that of all but the most capable animals. If a plausible case can be made that these are indeed sources of value distinct from both

sentience and rationality, then there is considerable theoretical work to be done before we should accept the human prejudice as a brute consideration of ethical significance.¹⁸

5. The Rationality of Wanting What Is Humanly Impossible

Let us now assess transhumanism in a different way, by formulating a fifth sense of “humanism.” On this sense of “humanism”, a humanist is one who believes that human beings should not seek goods beyond those that have been typically possible for human beings. This sense of “humanism” is perhaps most relevant for formulating transhumanism as a doctrine: a transhumanist is one who advocates that we pursue goods whose achievement has previously been humanly impossible.

Assessing transhumanism in this way involves assessing what it is rational to want.¹⁹ The transhumanist claims that, insofar as we constrain what we want by the facts typical of the human condition at present, we forgo goods that we could achieve if only we were to invest in them. In addressing this claim, it is useful to distinguish several modalities of the claim that it might be rational to want what is presently impossible.

It cannot be rational to knowingly want what is logically impossible: there is something decidedly irrational about wanting – to use actual cases from two rather different periods of human history – pi to be equal to three or the square root of two to be expressible as a fraction of whole numbers, if one is aware that pi is a transcendental number or that the square root of two is an irrational number. Similarly, if there is such a thing as a metaphysical impossibility that is logically possible, it cannot be rational to knowingly want what is metaphysically impossible. If something cannot be, there is no sense in promoting or advocating it. It is difficult to formulate cases uncontroversially

both metaphysically impossible and logically possible, but the following illustrates what I have in mind: it would be irrational to want to paint an object – the ultimate Christmas ornament, say – simultaneously red all over and green all over.

If we move next to the modality of physical impossibility, matters become less clear. Some cases look similar to the cases articulated in the previous paragraph. It does not appear much more sensible to want to become free from the force of gravity than to attempt to produce the ultimate Christmas ornament. And to take a case of great relevance to transhumanists: if our present understanding of physics is at least roughly correct, then it is impossible for a human being to live forever. This is impossible not only because of our biological limitations or because of the impending explosion of the Sun, but also because the second law of thermodynamics entails that the universe will succumb to what is called “heat death.”

Heat death occurs when the distribution of energy in the universe is so even and constant that nothing interesting can happen. Interesting things, such as star formation and tennis matches, happen only when the distribution of energy is uneven and as a result energy moves from where it is more concentrated to where it is less concentrated. Over time the universe’s egalitarian policy of energy redistribution produces complete energy uniformity, and human life becomes unsustainable.²⁰ Since it is difficult to imagine what human life would be like – or indeed, how any kind of life would be possible – in a universe with no second law of thermodynamics, it does not make much sense to regard the inevitable heat death of the universe as a regrettable constraint on human beings.

But not all physical impossibilities are easily assimilated to these cases. Consider next what is sometimes called the “cold death” of the universe.²¹ Unless the universe is

dense enough to stop expanding and begin collapsing under the influence of its own gravity – and hence to bring back together energy that has been dispersed – the temperature of all regions of the universe will approach absolute zero. Eventually all regions of the universe will become so cold that nothing interesting, including the continued existence of a human being, can take place. If current estimates of the density of the universe are correct, then the cold death of the universe, which ironically enough is compatible with its simultaneously undergoing heat death, will make human life impossible.²²

That immortality is physically impossible in this sense seems less obviously something that we ought to accept as part of the ground rules of human life than, say, the inevitability of gravity's tug. The latter seems part of the ground rules of our universe: we would be seeking something incredibly different if we sought to rewrite or abolish the laws of gravitational attraction, since a tweak in one physical law will make others very different as well. The former seems less like a wholesale refiguring of the universe; all we would need to prevent the cold death of the universe, after all, is a little more stuff per unit volume.²³

Now we – understood as the human beings who are alive today – would of course be irrational to care a great deal about the cold death of the universe. This is not because it is inherently irrational to care about it, however, but rather because we have an excess of reasons not to expect cosmic longevity.

This, then, brings us to the next modality of impossibility. Can it be rational to want that which is presently technologically impossible? Here I take it that the answer, at least in some circumstances, is yes. After all, one of things it can be rational to want is to

change the state of the art of what is technologically possible. There is thus no principled obstacle to transhumanism in the current state of technology.

6. The Tension Between Rational Thought and Rational Prudence

Perhaps the central issues for assessing the ethics of pursuing transhumanistic goals have less to do with what is possible and more to do with what is feasible. If we strive for that which is possible but unlikely to be achieved, then there is a danger that we will be chronically frustrated. We will be frustrated in part because our attention will be focused on aspects of our condition that are lacking relative to a standard that is not likely to be achieved any time soon. We will be frustrated also with ourselves for our inability to achieve what is possible, and with those who oppose transhumanistic goals for presenting a political obstacle to achieving these goals.

Although he does not explicitly discuss transhumanism, I think something like this is what Williams has in mind when he closes “The Human Prejudice” as follows:

In many, more limited, connections hopes for self-improvement can lie dangerously close to the risk of self-hatred. When the hope is to improve humanity to the point at which every aspect of its hold on the world can be justified before a higher court, the result is likely to be either self-deception, if you think you have succeeded, or self-hatred and self-contempt when you recognize that you will always fail. The self-hatred, in this case, is a hatred of humanity.²⁴

To the extent that the humanity in question is one’s own, I am not sure I see the danger Williams has in mind. If I come to agree with the transhumanist that there are far greater possibilities for my life than I will ever achieve, this may well frustrate me. I do not see why, however, it should lead me to hate myself. I am not responsible for my humanity, and so I have no reason to treat myself as defective simply because I regard my humanity

as a constraint on what I am able to achieve. It is only if I regard my humanity as defining in a fundamental way who I am and what my life is about that frustration with my humanity should manifest as hatred of self. Although Williams believes that we ought to regard our humanity as fundamental in this way, that belief cannot serve as a premise in any argument against transhumanism without begging the question.

But to the extent that the humanity that constrains me is not my own but that of other people, hatred is a more serious concern. For if the problem is not simply that we were born too soon, or that our evolutionary past makes us prone to brain degradation, but instead that our fellow humans refuse to endorse improving the human condition because they mistakenly believe that such efforts are fruitless or impermissible, then our reactive attitudes can and should kick into gear. Getting upset at nature for ruining your wedding with bad weather may well be irrational, but if the wedding is ruined by other people's thoughtlessness or culpably ignorant beliefs it might well be irrational *not* to manifest some anger. There is thus a real danger that a commitment to transhumanism will in practice lead not to self-hatred but to misanthropy.

Not all misanthropy is, however, irrational. In closing, I would like to suggest a diagnosis of an important systematic irrationality among human beings that I believe figures prominently in opposition to transhumanism. I would suggest further that an unwillingness to practice or accommodate this irrationality is justified and is a mark of a mature view of the human condition.

The systematic irrationality I have in mind results from the often deep tension between what it is rational to believe as a responsible thinker and what it is rational to do as a prudentially self-interested agent. People tend to believe, for example, that they are

both more intelligent and more attractive than average; and it is plausible that people tend to believe this in part because it is better to have a high self-image than a low self-image and because the illusion of superior intelligence and attractiveness enables one to have a higher self-image. In this and similar cases, people make their lives genuinely better (they rationally promote their prudential interests) by artificially inflating their beliefs about themselves (by irrationally ignoring evidence that a rational thinker would heed). Further support for the hypothesis that there is a tension between rational thought and rational prudence is provided by psychological studies demonstrating lower self-reported happiness and higher incidence of clinical depression among persons with more accurate beliefs about themselves and their circumstances.

It appears that when the recommendations of these two different varieties of rationality – the rationality that consists in responsiveness to relevant epistemic considerations and the rationality that consists in effectively promoting one's own interests – conflict, human beings have a propensity to follow the recommendations of self-interest.²⁵ This human propensity likely underlies and helps explain the fact that so many people believe in an afterlife similar to ordinary life in the face of overwhelming evidence that the conditions necessary for ordinary consciousness are not present after death; this belief benefits many people by assuaging their fears about their future nonexistence.²⁶ And needless to say, belief in an afterlife with ordinary consciousness demotes the significance of the transhumanist agenda, since extending life in this world is much less urgent if death in this world is not the end of consciousness.²⁷

Why think that the rationality of thought takes precedence over the rationality of prudence when the recommendations of the two conflict?²⁸ One answer lies in the value

of authenticity. As Robert Nozick pointed out, it is not without good reason that we would be reluctant to hook ourselves up to a virtual reality machine that provides us with permanent artificial pleasure and illusory satisfaction.²⁹ It makes a difference that our understanding of our own lives matches up with what is actually taking place, and we have reason to be suspicious of values that can be realized only through illusion.

This is not the only reason, however, why the rationality of thought should take precedence over the rationality of prudence. For even if authenticity matters, it is still possible for one's circumstances to be so bad that an inauthentic life of pleasure and satisfaction would be preferable to an authentic life; under such circumstances I would maintain that epistemic rationality supersedes prudential rationality. I think it gets closer to the truth to say that it is a sign of maturity – of enlightenment, in Kant's sense – when we have the courage to dispense with our illusions even when they would make our lives better. Given the importance to human flourishing of social interaction, the widespread immature unwillingness of some to part with prudentially recommended but illusory beliefs will predictably interfere with the efforts of some individuals to be responsible and mature thinkers. A recent ridiculous example of this is the efforts of opponents of evolutionary theory to introduce creationism into public high school curricula in the United States; opposition to transhumanism will no doubt be pressed in a similar parochial spirit.

The tension between the rationality of thought and the rationality of prudence also helps explain why people have such mixed feelings about training in critical thinking. To be sure, some competence in critical thinking is indispensable to effectively securing one's interests; but with respect to some sorts of beliefs, including paradigmatically belief

in an afterlife with ordinary consciousness, people appear to prefer not to be subject to critical evaluation. (Sometimes this reluctance manifests as polite refusal to discuss these views, and sometimes it manifests as impolite assertion of these views without regard to criticism.) This is of course of particular concern to philosophers, since critical thinking is our stock in trade. Indeed, the founding moment – or perhaps myth – of Western philosophy is the trial of Socrates as recounted in Plato's *Apology*. The story is a passion play, with Socrates saving humanity by questioning prevailing beliefs about ethics and religion but earning an unjust death sentence for his efforts. This is no doubt an overly dramatic account of the relationship between philosophers and society, but I think there is considerable truth in it. There is truth in it because in an important sense philosophy is a transhumanistic discipline, seeking not only to make sense of human judgments and practices but also to unsettle them by subjecting them to no-holds-barred criticism.

In these highly abstract meta-philosophical observations can be located, of all things, a novel argument in favor of transhumanism. For if the tension between rational thinking and rational prudence is a source of deep tragedy for human beings, then any way in which we can narrow the gap between the recommendations of rational thought and the recommendations of prudential action will reduce this tension and hence reduce the tragedy. Immortality may be forever beyond the reach of human beings, but perhaps the authentic hope for a vastly extended lifespan, realized through investment in research and development in the service of a transhumanistic agenda, can make it a little easier for us to be good thinkers and hence more mature and enlightened persons.

¹ I would like to thank Kyla Ebels Duggan, Sam Fleischacker, Marta Flory, Richard Kraut, Melissa Yates, and Rachel Zuckert for their extremely helpful comments on an earlier draft of this essay.

² For an overview of the tenets and historical development of transhumanism, see Nick Bostrom (2005), "A History of Transhumanist Thought", *Journal of Evolution and Technology* 14:1, 2005. For further

information and references see Bostrum's webpage at www.nickbostrum.com and fellow transhumanist Anders Sandberg's site at www.aleph.se/Trans/.

³ It is worth stating explicitly that the transhumanist goal of escaping death, or postponing the event as long as possible, need not be based on the dubious claim that death is the worst thing that can happen to a person. Anyone who has had a migraine knows that claim to be false. (That is not to say, of course, that migraines are worse than death; it is to say, rather, that the experience of a migraine yields insight into one fate that would be worse than death. It is this sort of fate that Epicurus was presumably most concerned with when he famously argued, in his letter to Menoeceus, that death is not bad for us.) The transhumanist is not stuck with the view that death is the worst possible human fate because the transhumanist claims only that in ordinary circumstances we should seek to postpone death. Faced with the prospect of extreme and chronic pain (or extreme and chronic loneliness, as articulated in note 16 below), it may well be irrational to seek to postpone death. But these fates are not the typical alternatives; it is all too common for lives to end while they are without question still worth living.

⁴ The wisdom of investing in transhumanist technologies is also complicated by concerns about whether the goods these technologies produce will be justly distributed. Since I focus on whether transhumanism is inherently wrong, I do not engage the problem of distribution in this essay. I will note, however, that even though it seems likely that – at least over the near term – there will be injustice in the distribution of these goods, this may not be a strong consideration against investment in transhumanist technologies. Did the predictable distributive injustice produced by the technological developments of the industrial revolution render prior investment in those technologies unwise?

⁵ Bernard Williams (2006), "The Human Prejudice", in *Philosophy as a Humanistic Discipline* (Princeton, 2006), 135.

⁶ As Sam Fleischacker has pointed out to me, the pervasiveness of religion in human societies distinguishes it from slavery and monarchy, and may provide some support for the claim that human beings cannot get along well without it. This still strikes me as a relatively weak consideration against secular humanism, however, since even historically pervasive institutions have been fruitfully abolished. Slavery was at one time nearly as pervasive as religion, and in our time sexist family structures are arguably more pervasive than religious practices; but surely slavery and sexist family structures are institutions to be overcome and not institutions to be accepted.

⁷ Williams (2006), 137.

⁸ Petrarch (Francesco Petrarca) was an early exponent of yet another doctrine known as "humanism": the Renaissance humanism developed in late medieval Italy. I will not discuss Renaissance humanism here, in part because it involves several of the other humanisms I do articulate and discuss. Renaissance humanists often endorsed the Petrarchan humanism according to which human beings are uniquely excellent, they often endorsed the view that the good for human beings is determined by their shared human nature, and they often endorsed the human prejudice by claiming that membership in the human species is ethically significant. Renaissance humanism is important to the historical development of Western ethical theory because it popularized the view that moral status results from a shared human capacity. This view crucially entails that non-Christians have the same moral standing as Christians, and is an essential precursor to modern moral theories like utilitarianism and Kantianism. Nick Bostrum views Renaissance humanism as an intellectual precursor to transhumanism; see Bostrum (2005).

⁹ Williams (2006), 136-137.

¹⁰ Williams (2006), 137.

¹¹ Williams (2006), 137. On this question see also Thomas Nagel (1971), "The Absurd", reprinted in his collection of essays *Mortal Questions* (Cambridge, 1979).

¹² See Friedrich Nietzsche (1873), "On Truth and Lies in a Nonmoral Sense", in *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*, trans. and ed. Daniel Breazeale (Harvester Press, 1979). If this quotation seems familiar even though you have not read Nietzsche's early notebooks, this is perhaps due to Douglas Adams's allusion to the passage in the opening paragraphs of his popular 1979 novel *The Hitchhiker's Guide to the Galaxy*.

¹³ Williams (2006), 139. The emphasis is Williams's.

¹⁴ That this is the structure of Williams's view explains why there are others, influenced not by Nietzsche but by the later works of Ludwig Wittgenstein, who share it. See, for example, Cora Diamond (2004), "Eating Meat and Eating People", in *Animal Rights: Current Debates and New Directions*, eds. Cass R. Sunstein and Martha C. Nussbaum (Oxford, 2004).

¹⁵ See Williams (2006), especially 142-150. Williams cites Peter Singer (2002), *Unsanctifying Human Life: Essays on Ethics*, ed. Helga Kuhse (Oxford, 2002).

¹⁶ Prominent examples of these sorts of critiques of utilitarianism include Williams's essay in J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge, 1973); Samuel Scheffler (1982), *The Rejection of Consequentialism* (Oxford, 1982); and Michael Slote (1985), *Common-Sense Morality and Consequentialism* (Routledge, 1985).

¹⁷ This observation is not meant to imply that utilitarians or Kantians are unable to vindicate the value of social interaction or of nonrational purposive agency. The distinctive claim under consideration is the hypothesis that the value of these capacities is not to be assimilated in either direction to the value of sentience on the one hand (as ways that individuals may experience pleasure or pain) or rationality on the other (as ways that individuals may exercise partially rational or protorational capacities).

¹⁸ The recent dominance of utilitarian and Kantian approaches to ethical theory has been so great that it may be useful to articulate a common but controversial assumption that theorists of both traditions tend to accept. I call this assumption the "ethical continuum hypothesis", and it is the claim that there is no capacity intermediate between sentience and rationality that is a source of value. This label is adapted from what number theorists call the "continuum hypothesis", which is the claim that there is no number that is cardinally greater than the number of integers and cardinally smaller than the number of real numbers, the latter of which is sometimes referred to as "the continuum". (Or, to pun shamelessly on terms used with different senses in this paper, it is the claim that there is no number that is cardinally greater than the number of rational numbers but cardinally smaller than the number of transcendental numbers.) It has been proved that the leading contemporary number theories are compatible with the continuum hypothesis and with the negation of the continuum hypothesis; the conception of number implicit in those theories thus underdetermines whether the (number) continuum hypothesis is true. Leading ethical theories, on the other hand, are not similarly noncommittal with respect to the ethical continuum hypothesis; and I expect in the near future to see a proliferation of ethical theories that depart from standard versions of utilitarianism and Kantianism by denying this hypothesis. (I do not intend, with this prediction, to disparage either of these traditions of ethical thinking; I self-identify as a Kantian theorist, and I believe that a suitably revised Kantian theory can do justice to the entire range of ethical phenomena.) But even if no theory that denies the ethical continuum hypothesis is superior to a straightforward Kantianism, that would not suffice to justify the human prejudice. There are commonly held ethical judgments that sit uneasily with the human prejudice, since we do not typically think that membership in the human species necessarily entitles an individual to equal moral regard with all other members of the species. Consider the primacy that ought to be given to a mother's interests over a fetus's in the face of life-threatening pregnancy complications, or the primacy that ought to be given to a more capable individual over an individual in a persistent vegetative state when difficult decisions of medical triage must be made.

¹⁹ A few words about "want": the notion of wanting that I deploy here is intermediate between the notion of merely wishing that something were the case and the notion of actually pursuing the achievement of that thing. I will not attempt to explicate this intermediate notion; I expect that some such intermediate notion is indispensable to an adequate theory of thought and agency, and I hope that the reader has some intuitive idea of this notion. I think that our ordinary notion of wanting is such an intermediate notion.

²⁰ Or, to be more precise: it leads to energy uniformity as complete as is compatible with Heisenberg's uncertainty principle.

²¹ The cold death of the universe is also known as the "big freeze".

²² In note 1 above I alluded to one sort of fate that is worse than death; another such fate arises from the cold death scenario just presented. Suppose that you achieved immortality by assuming the form of an immaterial soul, but that you were unable to enable others to do the same. This raises the unsettling prospect of watching every other sentient being wiped out by the inevitable diffusion of energy through the universe, until eventually you are left entirely and eternally alone in an uninteresting universe. In another article, "The Makropulos Case: Reflections on the Tedium of Immortality", Williams argues that any infinite lifespan would be unbearably uninteresting. See Williams (1972), in Bernard Williams, *Problems of Self* (Cambridge, 1973). I will not address that position here explicitly, though I will note that I think it understates the creative potential of the rational mind. I will note also that it is essentially the inverse of Socrates's (possibly facetious) argument from Plato's *Apology* for the same conclusion, namely, that death is good for us. Since sleep is so enjoyable and death is the soundest possible sleep, Socrates's argument goes, it follows that death is good for us. This position also, I think, understates the availability and

importance of goods – including chiefly the company of other rational and sentient creatures – that are rich and complex enough to sustain our interest indefinitely.

²³ This comment comes too blithely, since apparently the alternative to cold death is the contraction of the universe to a space so small and hence so dense that it is too hot to accommodate human life (a possible future I am unable to resist calling the “heat death” of humanity).

²⁴ Williams (2006), 152.

²⁵ Because I am convinced of this deep tension between rational thinking and prudential agency, I find pragmatic arguments for belief in an afterlife – such as those found in William James (1902), *The Varieties of Religious Experience* (Modern Library, 1902) – to be among the most seductive, in both the positive and negative senses of the term. I do not have space here to explore those arguments.

²⁶ As Richard Kraut has pointed out to me, it is far from clear that it is typically in a person’s all-things-considered prudential interest to believe in an afterlife that involves ordinary consciousness. Suicide bombers provide a vivid example of people who waste their lives out of concern for the afterlife, and there are many more mundane examples of people who devote themselves to flourishing in the next world in a way that interferes with their flourishing in this one. I nonetheless believe that, at least in the contemporary United States, belief in an afterlife with ordinary consciousness is typically motivated in part by our interest in not worrying too much about our own mortality.

²⁷ I would speculate further that this propensity is a general feature of embodied rational beings, and not an accident of our apelike ways of thinking; or if it is an accident of our apelike ways of thinking, then this is part of the reason why apes and not (say) birds evolved full-blown rationality. If we encounter intelligent aliens I expect we will find them to be systematic confabulators and fantasizers, just as we are, and I expect we will find that they hold on to their hopes for the future well beyond the point of epistemic responsibility, just as we do.

²⁸ I am attracted to the Kantian idea, suggested to me by Sam Fleischacker, that the norms of thought have a status akin to the Categorical Imperative and that this explains why they take precedence over the norms of prudence when the two conflict. I cannot attempt to justify that view here.

²⁹ See Robert Nozick (1974), *Anarchy, State, and Utopia* (Basic Books, 1974), 42-45.