Are We Living At The Hinge Of History?

0. Introduction

In the final pages of *On What Matters, Volume II* (2011), Derek Parfit made the following comments:

> We live during the hinge of history. Given the scientific and technological discoveries of the last two centuries, the world has never changed as fast. We shall soon have even greater powers to transform, not only our surroundings, but ourselves and our successors. If we act wisely in the next few centuries, humanity will survive its most dangerous and decisive period. Our descendants could, if necessary, go elsewhere, spreading through this galaxy.[1]

These comments hark back to a statement he made twenty-seven years earlier in *Reasons and Persons* (1984):

> the part of our moral theory... that covers how we affect future generations... is the most important part of our moral theory, since the next few centuries will be the most important in human history.[2]

He also subsequently made the same claim in even stronger terms during a talk sponsored by Giving What We Can at the Oxford Union in June 2015:

> I think that we are living now at the most critical part of human history. The twentieth century I think was the best and worst of all centuries so far, but it now seems fairly likely that there are no intelligent beings anywhere else in the observable universe. Now, if that's true, we may be living in the most critical part of the history of the universe… [The reason] why this may be the critical period in the history of the universe is if we are the only rational intelligent beings, it's only we who might provide the origin of what would then become a galaxy-wide civilisation, which lasted for billions of years, and in which life was much better than it is for most human beings. Well, if you look at the scale there between human history so far and what could come about, it's enormous. And what's critical is that we could blow it, we could end it.[3]

---

[1] Parfit (2011), p. 616.

[2] Parfit (1984), p. 351. I thank Pablo Stafforini for reminding me that Parfit made this comment.

[3] "We are living in the most crucial moment in the history of the Universe - Derek Parfit - Oxford talk," https://youtu.be/j9Y26XUwtQQ

The claim that we live at the most important time in history is striking. But, despite the clear influence it had on his thought, in his written work Parfit simply asserts this claim, in the context of discussing other topics; he does not canvass arguments either for or against.[4]

In this article I try to make the hinge of history claim more precise, give arguments in favour and against, and assess whether it is true. Ultimately, I argue that the claim (as I construe it, which might be quite far from any claim Parfit would endorse) is quite unlikely to be true, and that this fact can serve as part of an argument for the conclusion that impartial altruists should generally be investing their resources, rather than trying to do good immediately.

In section 1 I give some background by sketching two worldviews that might motivate the claim that we live at the hinge of history. In section 2 I make the claim more precise, choosing to define my terms so that they are action-relevant, bearing on the question of whether to 'give now or give later'. In sections 3 and 4 I give two arguments against the hinge of history claim, and in section 5 I discuss two counter-arguments in favour. I conclude that there are some strong arguments for thinking that this century might be unusually influential, but that these are not strong enough to make the hinge of history claim likely.[5]

1. Two worldviews

I know of two worldviews that might motivate the idea that we live at the most important period in history. Both of these worldviews rely on a perspective that is impartial and longtermist:[6] they assess the importance of an event 'from the point of view of the universe,' rather than from our own parochial perspective; and they assume that, in expectation, the vast majority of value occurs in the very long-run future, appreciating that civilisation might persist for billions of years, spreading to the stars and potentitally settling trillions of solar systems.

---

[4] I say this with one caveat. In the talk he gives one argument, as follows: "There are many ways in which human history might be ended soon, probably nuclear war isn't one of them, but there are various others. The simplest is a really large asteroid. To guard against quite a lot of these dangers, we need to start colonising other parts of space. Need to put a few people on earthly planets and then go further. That's why, when we have spread out, it'll be less critical. That's why this is the most dangerous period." I'll come back to this argument in section 5.

[5] This article is indebted to many people: those who have been particularly influential include Nick Beckstead, Phil Trammell, Toby Ord, Aron Vallinder, Allan Dafoe, Matt Wage, and, especially, Holden Karnofsky and Carl Shulman. I also thank the many insightful commenters who responded to an early blog post version of this article, 'Are we living at the most influential time in history?' *Effective Altruism Forum*, Sep 2019, available at https://forum.effectivealtruism.org/posts/XXLf6FmWujkxna3E6/are-we-living-at-the-most-influential-time-in-history-1. As I mention in that blog post, these ideas have been discussed in the effective altruism community for some time and I don't claim originality for any of them, though their development and the resulting mistakes are my own.

[6] For discussion of the idea of 'longtermism' see my blog post 'Longtermism', *Effective Altruism Forum,* Jul 2019, https://forum.effectivealtruism.org/posts/qZyshHCNkjs3TvSem/longtermism. For a defense of 'strong longtermism' see Greaves and MacAskill (ms), The Case for Strong Longtermism.

The first worldview is the *Time of Perils* view: that we live at a period of unusually high risk of human extinction. The term comes from Carl Sagan, who I believe was an influence on Parfit:

> It might be a familiar progression, transpiring on many worlds—a planet, newly formed, placidly revolves around its star; life slowly forms; a kaleidoscopic procession of creatures evolves; intelligence emerges which, at least up to a point, confers enormous survival value; and then technology is invented. It dawns on them that there are such things as laws of Nature, that these laws can be revealed by experiment, and that knowledge of these laws can be made both to save and to take lives, both on unprecedented scales. Science, they recognize, grants immense powers. In a flash, they create world-altering contrivances. Some planetary civilizations see their way through, place limits on what may and what must not be done, and safely pass through the time of perils. Others, not so lucky or so prudent, perish.[7]

On this view, with the invention of nuclear weapons, we entered an era where we had the technological power to destroy ourselves. Similarly, as we improve technologies like synthetic biology, we will soon develop the ability to create novel pathogens that could infect and kill the entire world population. These technologies pose unprecedented risks to the continued survival of mankind. Within a few centuries, however, wisdom will have caught up with technological progress, and we'll take action to reduce the risks; alternatively, we will have spread out among the planets, and civilisation will not be so fragile.[8]

The 'unusual' is important on the Time of Perils view. Perhaps extinction risk is high at this time period, but will be even higher at some future times. In which case those future times might be even more important than today. Or perhaps extinction risk is high, but will stay high indefinitely, in which case in expectation we do not have a very long future ahead of us, and the grounds for thinking that extinction risk reduction is of enormous value fall away.[9]What's more, what's really crucial is not that we live at a period of unusually high extinction risk, but that we live at a period where we can do an unusual amount to reduce extinction risk. If extinction risk were high, but there was nothing we could do to reduce it, then the Time of Perils view would be of historical interest, but would not be of interest from the perspective of figuring out what we ought to do.

The second worldview that could motivate the hinge of history idea is what I'll call the *Value Lock-In* view: that we are coming to a point in time where we will invent a technology that will enable the agents alive at that time to maintain their values indefinitely into the future, controlling the broad sweep of the entire rest of the future of civilisation. The most prominent example of this worldview, most closely associated with the work of Nick Bostrom and Eliezer Yudkowsky, identifies greater-than-human-level artificial intelligence as the key

---

[7] Sagan (1994), p. 305-306. I thank Pablo Stafforini for this quote.

[8] Ord (2020) stakes out this viewpoint in considerable depth.

[9] For more on the importance of the exogenous risk of extinction to the value of the long-term future, see Tarsney (2019).

technology determining when value lock-in will happen. These authors are quite aware that they are, therefore, claiming that the invention of greater-than-human-level intelligence will be the most important event in history. In the Preface to *Superintelligence*, Bostrom describes his view as follows:

> In this book, I try to understand the challenge presented by the prospect of superintelligence, and how we might best respond. This is quite possibly the most important and most daunting challenge humanity has ever faced. And—whether we succeed or fail—it is probably the last challenge we will ever face.

Later in the book he expands on this idea:

> it may be reasonable to believe that human-level machine intelligence has a fairly sizeable chance of being developed by mid-century, and that it has a non-trivial chance of being developed considerably sooner or much later; that it might perhaps fairly soon thereafter result in superintelligence; and that a wide range of outcomes may have a significant chance of occurring, including extremely good outcomes and outcomes that are as bad as human extinction…"[10]

And later he summarises part his discussion on the powers of superintelligence as follows:

> the first superintelligence might well get a decisive strategic advantage. Its goals would then determine how humanity's cosmic endowment will be used.[11]

On this Bostrom-Yudkowsky view, in the coming decades or centuries, we will invent an artificially intelligent agent that has the power to improve its own intelligence, which then will give it greater powers to improve its own intelligence further. Through this process of recursive self-improvement, that agent might rapidly — perhaps over the course of days or weeks — develop intelligence greater than that of all of the rest of humanity combined. At that point, it will have the power to do what it wants with the human species, and will be able to spread to the stars and use the resources in the accessible universe in whatever way it wants. If, however, we are able to control this superintelligence, and align it with human values, then our preferences would determine how all of the resources in the accessible universe would be used. Either way, how this transition goes will determine the entire future of the universe.

On either of these views, we live at, or are approaching, the hinge of history. Let's now turn to making this claim more precise.

2. Making the Hinge of History claim precise

---

[10] Bostrom (2014), p. 21

[11] Bostrom (2014), p. 115. 'Cosmic endowment' refers to all accessible resources in the universe.

The claim that we are at the most 'important' or 'critical' time in human history is vague. There are various ways of making this idea more precise, and in this I will choose only one way of doing so. To be clear, I don't take myself to be undertaking exegesis of Parfit's views — this may or may not be the concept that Parfit had in mind, and other definitions of the concept could result in other interesting discussions.

The concept in this area that I will focus on is *how much expected good one can do with the direct expenditure (rather than investment) of a unit of resources at a given time*. I will call this the *influentialness* of a time. On my interpretation, then, the 'hinge of history' claim is that we live at the most influential time ever.

'Influentialness' is an interesting concept because it connects closely to an action-relevant issue: namely, whether as impartial altruists we should be trying to do good now, or whether we should be trying to invest resources in order that we (or people we pass our resources onto) can do more good at a later date. In particular, if we are longtermists — that is, we have a particular concern for ensuring that the long-run future goes well[12] — then there is a *prima facie* presumptive argument in favour of the idea that we should be investing[13] in order to have more impact at a later date. As Parfit notes, civilization might last for billions of years. Given this, if our aim is to influence the value of the long-term future, we have only lost a tiny proportion of that value if we delay the point at which we take action by a few centuries, passing on our philanthropic resources to younger people who share our values, who would then later do the same, passing those resources onto younger people who share their values. But over that time, those resources would have grown enormously. At a 5% real rate of return, over 200 years our invested financial resources would be 17,000 times as large. What's more, because the rate of return on investment exceeds the growth rate,[14] these resources would also be much larger as a proportion of the world economy: if the rate of return is one percentage point larger than the growth rate of the world economy, after 200 years the invested resources would be 7 times as large, when measured as a fraction of the world economy. Other things being equal, greater resources would allow us (or our inheritors) to do much more good. So there seems to be a strong *pro tanto* reason for impartial and longtermist altruists to invest their resources rather than donating now.

However, if now is a particularly influential time, then there is a potential response to this argument. If we have very unusually good opportunities for doing good now that we won't have in the future then, even though we would have greater resources in the future, nonetheless we might plausibly be able to have more of an impact now, with these very unusual opportunities. So assessing whether we are at a particularly influential time is crucial for assessing the decision of whether to try to have an impact now, or to invest and give later.

---

[12] MacAskill (2019).

[13] Here I use 'investment' to refer to both financial investment, and to using one's time to grow the number of people who are also impartial altruists. So the idea of investment, here, is not limited merely to money.

[14] Piketty (2013).

We can make this concept of 'influentialness' more precise in the context of Philip Trammell's work on the optimal timing of philanthropy.[15] In Trammell's basic model, the expected good that one does at a time is given by three factors. First, is simply the amount of philanthropic resources one uses at that time. Second, is how quickly philanthropic resources diminish in their returns.[16] Third is a scale factor: at different times, because of the opportunities available at the time, the same amount of resources, if well-spent, will generate more or less expected value. This scale factor is the idea of 'influentialness' that I refer to.[17]

However, for the purposes of true action-relevance, the influentialness of a time is not quite what we're looking for. Even if we assume that now is the most influential *time*, because there are available opportunities to safeguard the long-run future, a rural farmer in Central African Republic would simply not be able to access those opportunities, and so for that person the question of how influential the present time is is neither here nor there. So we can generalise Trammell's model slightly by talking about person-times rather than times: rather than being indexed to a particular time, each term in his model should be indexed to a particular person-time. This means the model could generate conclusions not just about *when* our resources should be used to generate impact, but also *who* we should give those resources to in order to generate impact.

Four features of the concept of influentialness are worth emphasising. The first is that one's influentialness is given by how much expected good one *can* do at a time. It is not given by how much (expected) good one *actually* does. So, hypothetically, someone who in 1910 knew what Hitler would go on to do, and had the opportunity to discourage him from ever moving into politics, but chose not to discourage him, would count, on this definition, as highly influential, even though as a matter of fact they did not change the course of human history in any way.

The second notable feature of this concept is that the influentialness of a person at a time is dependent on the level of knowledge and understanding of that person at that time. Imagine, for example, that some hunter-gatherer had an opportunity to shape the entire course of the future of the human race, but was not able to know, or be in a position to know, that this was possible. That hunter-gatherer would have been at what we might call an unusually 'pivotal' time, but they would not have been at a particularly *influential* time, because they would not have been able to turn resources into enormous amounts of (expected) value. Applying this to our own case: we could fail to be at a particularly influential time either because there are no unusually impactful opportunities to do good available to us, or because we lack the understanding necessary to take advantage of those opportunities. It may well be the case that, even if future opportunities are worse than they are today, future people will be more influential because they have much better scientific and moral knowledge than we have today.

---

[15] Trammell (ms).

[16] As is standard in economic theory, Trammell models this using an isoelastic utility function.

[17] Trammell himself refers to this as 'hingeyness', in reference to Parfit's claim. However, I worry that this term sounds too unserious, so I prefer 'influentialness'.

The third feature of this concept that I want to highlight is that the probability distribution that goes into the idea of 'expected value' in the definition of influentialness is our own. This can be confusing when we are considering people in the future who might have much better (or worse) evidence than us, and therefore have different probability distributions. But we can think about it using the standard analysis of the 'value of information'. In the standard analysis, the value of gaining information (including imperfect information, which might be misleading) is given by the expected value of making the best decision given the new information minus the expected value of making the best decision without that information. So, for example, if you currently believe you have a 60% chance of making the right decision about how to spend your resources, which would generate 1 unit of value if you make the right decision, but believe that if you gained some piece of new evidence you would have an 80% chance of making the right decision, then you should try to get that evidence, and indeed you should be willing to forego up to 0.2 units of value in order to get that evidence.[18]

Although value-of-information analysis is typically limited to a single decision-maker over time, we can use the same analysis to think about passing resources across multiple decision-makers over time. If you think you have a 60% chance of making the right decision about how to use a unit of resources, with value 1 if you do make the right decision and 0 otherwise, whereas some other future person has an 80% chance of making the right decision about how to use a unit of resources, with value 1 if they make the right decision and 0 otherwise, then you should pass your resources onto this other person. (Indeed, you should be willing to forego up to 0.2 units of value in order to pass on those resources.)

The same analysis applies, moreover, if one believes that the *option-set* might be different in the future. If your best option, with a unit of resources, generates 1 unit of value, and you believe that there's a 50% chance that a particular future person will have a new best option worth 2 units of value, and a 50% chance of a best option worth only 0.5 units of value, then you should pass resources on to that future person. Alternatively, if you thought this future person would have only a 10% chance of having a best option worth 2 units of value, and a 90% chance of a best option worth 0.5 units of value, then you should spend the resources yourself.[19]

In general, in this context we can usefully, though somewhat cold-heartedly, think of future people (including ourselves at later times), who we might pass resources on to, simply as machines for converting resources into good outcomes. If we think that such people will in expectation, and by our own lights, be better than us at converting resources into good outcomes then we should pass our resources on to those people.

The fourth feature of the concept I want to highlight is that it refers only to direct expenditure of resources, or what economists would call 'consumption,' rather than investment. This is

---

[18] Plugging these numbers into the formula: Expected value of best decision given new information – Expected value of best decision given no new information = (0.8*1 + 0.2*0) – (0.6*1 + 0.4*0) = 0.2

[19] For more discussion of the idea of 'option value', including in the application to moral learning over time, see MacAskill (ms).

crucial because the issue I am primarily addressing is whether we should spend our resources now, or invest them for a later date. This means that we need to be careful. If investment can reliably compound at a positive rate over time, then there is always an argument for thinking that earlier generations can do more good than later generations: if the later generation is more influential, then the earlier generation can invest the money and give a larger sum to the later generation. So when I later claim that previous generations were less influential than we are today, I am meaning that they could do less good with direct expenditure; I am not also making the claim that, had they invested the money, passing on those resrouces to today, they would still have done less good.

With all this on board, we can turn now to stating the Hinge of History claim. The version of the claim that I wish to assess is the claim we are at an enormously influential moment out of a vast future. That is, not merely are we among the most influential people ever, but we are among the most influential people ever out of a civilisation that will one day take to the stars. So I can state the claim as follows:

> HH: We are among the very most influential people ever, out of a truly astronomical number of people who will ever live.

In HH, the phrase 'among the very most influential' is vague. Nothing too significant rests on this vagueness; we could make it more precise by interpreting it as saying, for example, that you and I are among the million most influential people ever.

Parfit himself would not necessarily claim that HH is very likely. But what he would endorse, I think, is the conditional claim that *if* civilisation survives the next few centuries, then we are among the very most influential people ever, out of trillions upon trillions of people who will ever live. And, though to my knowledge Parfit does not state his views on whether we will survive the next few centuries, it seems hard to believe it's extremely likely that we'll destroy ourselves. If Parfit agreed with that, then he would have assigned quite a significant likelihood (greater than 10%) to HH being true.

Note that, in representing the claim the way I do, I am considering a somewhat bolder claim than the one that Parfit himself makes. He claims that the next few centuries are the most important ever, whereas I am considering primarily the claim that *we today* are among the most influential people who ever live. I do this, again, to make the discussion as action-relevant as possible. For the purposes of action today, it does not much matter whether the most influential time is one century or one thousand years away. Either way, we have an argument for saving. It's only if now — our own lifetimes — are far more influential than times after we die that we have, instead, an argument for trying to have an impact right away rather than growing our resources and passing them on to subsequent generations.

Having now defined the claim that I consider, I turn to assessing its plausibility. In the next two sections, I give two arguments against having a significant degree of belief in HH.

3. The base rates argument against HH

The first argument I want to marshall against HH is simply that our prior in HH should be very low, and the evidence we have in favour of it is not sufficiently strong to overcome this low prior.

A natural prior is given by what Bostrom called the *Self-Sampling Assumption*. I'll use the formulation of this assumption given by Thomas (ms):

> A rational agent's priors locate him uniformly at random within each possible world.[20]

An implication of this principle is that, for any property *F*, your prior should be such that if there are *n* people in a population, the probability that you are in the *m* most *F* people in that population is *m*/*n*. This principle seems compelling as a way of setting priors. The a priori probability that I am in the top 100 funniest people in Scotland today is 100 out of 5.4 million; the a priori probability that I am in the top 1000 strongest people in the UK today is 1000 out of 66.4 million. I believe that, among those who study anthropic reasoning, the self-sampling assumption (as stated in some form similar to Thomas's formulation) is widely accepted: the major question is whether to *also* accept further principles, like the 'self-indication assumption'.[21]

If we set our priors this way, we assign a very low prior probability to HH. If we don't go extinct in the next few centuries, then there are plausibly a vast number of people in the future. The Earth will remain habitable for something on the order of a billion years. Even if current population levels reduced to a tenth of what they are today (i.e. to about 1 billion people per century), that would mean that there would be ten thousand trillion people to come. If, as Parfit suggests, we would subsequently take to the stars, that number would get far higher: there are one hundred billion stars in the Milky Way; settling just 0.1% of them with the same population as on Earth would mean that there are a trillion trillion people to come. If we consider also the 8 billion other galaxies that we could access,[22] the numbers get correspondingly higher again.

---

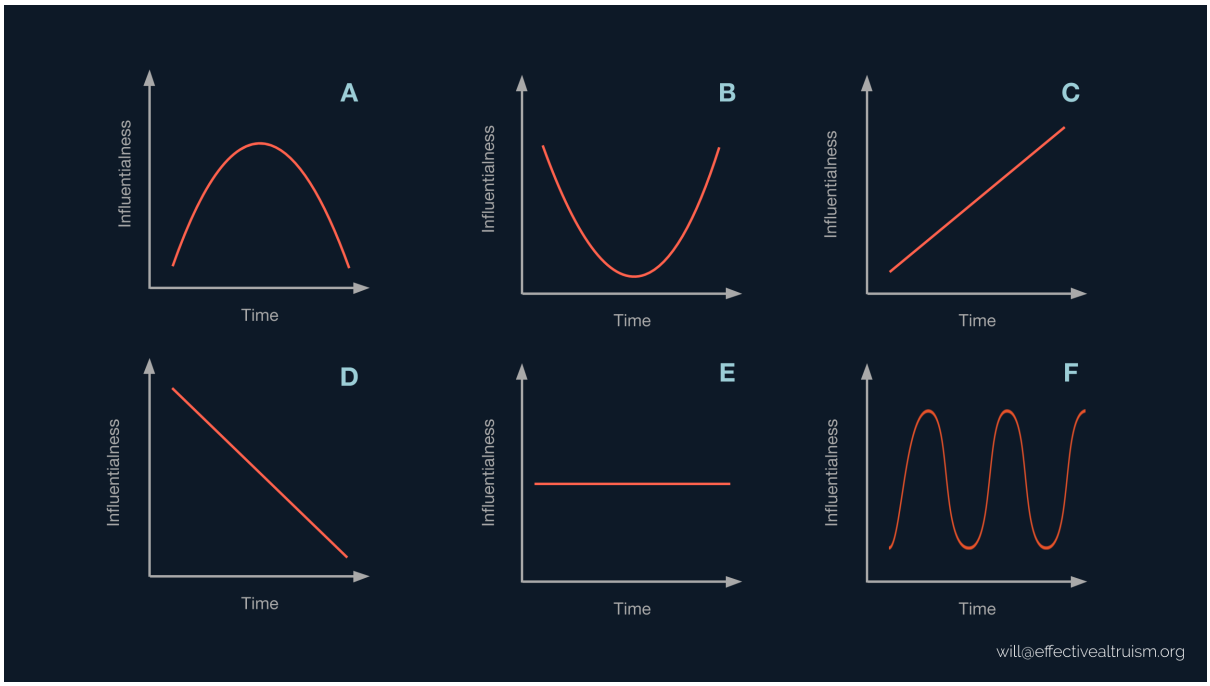
[20] Note that this is the 'very rough' formulation given by Thomas. His paper ultimately spells out a much more precise formulation, but this will not be necessary for our purposes. Bostrom (2002), p. 57, states the principle as follows: "(SSA) One should reason as if one were a random sample from the set of all observers in one's reference class."

[21] Thomas (ms) states the self-indication assumption (very roughly, before making it more precise) as: "A rational agent is proportionally more confident a priori in worlds with large populations than in worlds with small populations, all else equal." Bostrom (2002, p. 66) states the self-indication assumption as: "Given that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist.". Note, moreover, that the argument I make is not merely a version of the Doomsday argument. Even if we accept the self-indication assumption (as I believe we should), which neutralizes the Doomsday argument, there is a further question about whether we are among the most influential people ever. See Mogensen (2019) for an in-depth explanation of this.

[22] Ord (ms).

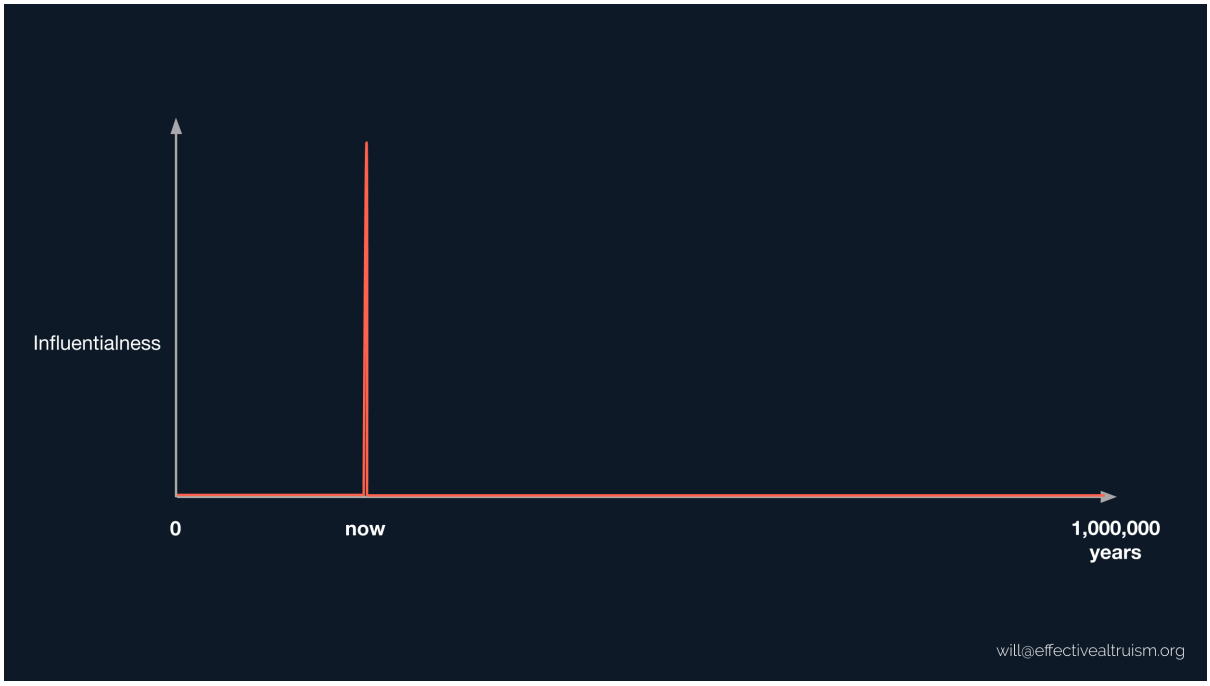
For the purposes of my argument, what matters is not these precise numbers, but that any of them are astronomical. If there are a trillion trillion people to come, then the a priori probability that we are among the million most influential people ever is one in a million trillion. This is about the same probability as dealing a Royal Flush frmo a well-shuffled pack of cards three times in a row. But even if we assume that there are only a hundred trillion people to come, the a priori probability of being among the million most influential people ever is still one in a hundred million — about as likely as winning the lottery.

An alternative, more visual, way of seeing the same argument is to think about all the ways that influentialness might vary over time. (Here, in order to be able to represent this on a two-dimensional graph, I'll just look at influentialness over time, rather than influentialness over person-times). Some ways in which we might a priori expect influentialness to vary are as follows:



However, at least on the Value Lock-In view, none of these represent how influentialness varies over time. Instead, the graph looks like this:[23]

[23] This, of course, is an approximation. Occasions prior to the lock-in event could still be very influential, if they gave the opportunity to prevent an extinction event, or if they gave the opportunity to shape the values of those who are alive during the lock-in event.



That is, on the Value Lock-In view, almost everything of importance that happens for the fate of the universe occurs over the course of just a few decades or a century — what is essentially a single point in time. What's more, on this view, that point in time is essentially now. This is an astonishing claim to make. It's not so clear that the Time of Perils view would have this same exact implication. But Parfit's comments that we might be living at the 'most critical part of the history of the universe,' suggests that the broad picture is roughly similar.

That we should have a very low prior in HH doesn't yet tell us that we should have a very low posterior in HH. Sometimes we should update from extraordinarily low priors to significant posteriors. For example, if I shuffle a pack of cards and then deal them out face up in a row, and there is a random-seeming sequence of cards in front of you, your posterior that the cards are in the sequence they seem to be in should be reasonably high. However, your prior that the cards would be in that sequence should have been astronomically low: 1 in 52!, or 1 in 10^68.

However, HH is not merely a priori extremely unlikely. It's also *fishy*. It's less like believing that I just dealt a random-seeming sequence of cards from a well-shuffled pack, and more like believing I dealt a sequence of cards in perfect order (2 to Ace of clubs, then 2 to Ace of diamonds, etc) from a well-shuffled pack. Being fishy is different than just being unlikely. The difference between unlikelihood and fishiness is the availability of alternative, not wildly improbable, hypotheses, on which the outcome or evidence is reasonably likely.[24] If I deal the random-seeming sequence of cards, I don't have reason to question my assumption that the deck was shuffled, because there's no alternative somewhat plausible background hypothesis on which the random-seeming sequence is a likely occurrence. If, however, I deal the deck of cards in perfect order, I do have reason to significantly update that the deck was not in fact

[24] Horwich (1982, p.94), though he talks about an event being 'surprising' rather than a claim being 'fishy'.

shuffled, because the probability of getting cards in perfect order if the cards were not shuffled is reasonably high. That is: P(cards not shuffled)P(cards in perfect order | cards not shuffled) >> P(cards shuffled)P(cards in perfect order | cards shuffled), even if my prior credence was that P(cards shuffled) > P(cards not shuffled). So I should update towards the cards having not been shuffled.

Similarly, if it seems to me that I'm among the most influential people who have ever or will ever live, this gives me good reason to suspect that the reasoning process that led me to this conclusion is flawed in some way, because P(I'm reasoning poorly)P(seems like I'm living at the hinge of history | I'm reasoning poorly) >> P(I'm reasoning correctly)P(seems like I'm living at the hinge of history | I'm reasoning correctly).

The strength of this argument depends in part on how confident we are of our own reasoning abilities in this domain. But it seems to me there's a strong risk of bias in our assessment of the evidence regarding how influential our time is. One reason for thinking this is *salience*. It's much easier to see the importance of what's happening around us now, which we can see, than it is to assess the importance of events in the future, involving technologies and institutions that are unknown to us today, or (to a lesser extent) the importance of events in the past, which we take for granted and involve unsalient and unfamiliar social settings.[25]

A second reason for thinking this is *confirmation bias*. For those of us, like myself and like Parfit, who would very much like for the world to be taking much stronger action on extinction risk mitigation (even if the probability of extinction is low), it would be a good outcome if people who do not have altruistic and longtermist values think that the risk of extinction is high, even if it's low. So we might be biased (subconsciously) to overstate the case in favour of taking action to mitigate extinction risk. And, in general, people have a tendency towards confirmation bias: once they have a conclusion ("we should take extinction risk a lot more seriously"), they tend to marshall arguments in favour of that conclusion, rather than carefully assess arguments on either side. Though we try our best to avoid such biases, it's hard to overcome them.

In general, if you accept that you should have a very low prior in HH, you need to be very confident that you're good at reasoning about the long-run significance of events (such as the magnitude of risk from some new technology like artificial intelligence or synthetic biology), and our ability to have leverage over them, in order to have a significant posterior credence in HH, rather than concluding, instead, that we're mistaken in some way. But we have no reason to believe that we're very reliable in our reasoning in these matters. We don't have a good track record of making predictions about the importance of historical events, and some track record of being badly wrong. So, if a chain of reasoning leads us to the conclusion that we're living at the most influential time ever, we should think it more likely that our reasoning has

---

[25] One example of this salience bias in action, with respect to the 'influentialness' claim, comes from Qin Dynasty philosopher Li Si: "With the might of Qin and the virtues of Your Highness, in one stroke, like sweeping off the dust from a kitchen stove, the feudal lords can be annihilated, imperial rule can be established, and unification of the world can be brought about. This is the one moment in ten thousand ages. If Your Highness allows it to slip away and does not press the advantage in haste, the feudal lords will revive their strength and organize themselves into an anti−Qin alliance. Then no one, even though he possess the virtues of the Yellow Emperor, would be able to annex their territories" (De Bary and Bloom, 1999, p. 208).

gone wrong than that the conclusion really is true. Given the low base rate, and given our unreliable tools for assessing the claim, the evidence in favour of HH is almost certainly a false positive.

Finally, we can assess the quality of the arguments given in favour of the Time of Perils or Value Lock-in views, to see whether, despite the a priori implausibility and fishiness of HH, the evidence is strong enough to give us a high posterior in HH. It would take us too far afield to discuss in sufficient depth the arguments made in *Superintelligence*, or *Pale Blue Dot*, or *The Precipice*. But it seems hard to see how these arguments could be strong enough to move us from a very low prior all the way to significant credence in HH. As a comparison, a randomised controlled trial with a p-value of 0.05, under certain reasonable assumptions, gives a Bayes factor[26] of around 3 in favour of the hypothesis;[27] a Bayes factor of 100 is regarded as 'decisive' evidence.[28] In order to move from a prior of 1 in 100 million to a posterior of 1 in 10, one would need a Bayes factor of 10 million — extraordinarily strong evidence.

But the evidence we currently have for either the Value Lock-In view or the Time of Perils view are merely informal arguments. They aren't based on data (because they generally concern future events) nor, in general, are they based on trend extrapolation, nor are they based on very well-understood underlying mechanisms, such as physical mechanisms. And the range of deep critical engagement with those informal arguments, especially from 'external' critics, has, so far, been limited. So it's hard to see why we should give them much more evidential weight than, say, a well-done randomised controlled trial with a p-value at 0.05, let alone assign them an evidential weight 3 million times that amount.

Of course, a full treatment of this would involve assessing at length the arguments that Bostrom and Ord and others give for their position, which it's not the purpose of this article to do. But it's hard to see how, even if the arguments in those texts seemed compelling, they could be strong enough to move us all the way from a tiny prior to a sufficiently large posterior.

I'll now consider two responses to the argument I have just made. The first response is to accept that we don't have good reasons for thinking that we're at the *most* influential time in history. Instead, we could just consider the idea that we're at an enormously influential time. And very little changes whether you think that we're at the most influential time ever, or merely at an enormously influential time.

However, I don't think this response is a good one, for two reasons. First, the Bostrom-Yudkowsky view on superintelligence is inconsistent with the idea that we're merely at an enormously influential time. On their view, the development of artificial general intelligence is the decisive moment for the entire rest of civilisation. But if you find the claim that we are

---

[26] Where the Bayes factor is P(hypothesis | evidence) / P(not-hypothesis | evidence).

[27] Benjamin et al (2018).

[28] E.g. Jeffreys (1962, p.432)

among the very most influential people ever hard to swallow, then you have, by modus tollens, to reject the Bostrom-Yudkowsky story of the development of superintelligence. So there is a material difference in the views we should hold, depending on whether we believe we're at the most influential time ever, or merely an enormously influential time.

Second, even if we're at some enormously influential time right now, if there's some future time that is even more influential, then an obvious strategy for longtermist altruists to pursue would be to send resources to that time in the future. So the question of whether we're at the most influential time, or a merely enormously influential time, is directly action-relevant.

A second counterargument that one could make is that, for the purposes of action-relevance, we do not need to consider how influential this time is compared to times in the past, or times in the distant future.[29] All that matters is the relative influentialness of now compared to any time we can (in expectation) pass resources on to, which might be the next thousand years or so, but not much longer than that.

In response, I'll note again that some views, such as the Bostrom-Yudkowsky view, are inconsistent with the claim that we're merely at the most important time in the next thousand years — on their view, we are at the most important time ever. And if, in considering the question of influentialness over time, we come to have less confidence in the Bostrom-Yudkowsky view, that could lead us to take very different actions than we would otherwise have taken. Moreover, understanding how influentialness may or may not have varied in the past is useful if we want to think about how it might vary in the future. (If, for example, we come to believe that the formation of the world religions were a particularly influential moment in the past, that might lead us to think that the points in time when new ideologies are formed in the future will also be particularly important.)

But the point that, ultimately, what we should care about is the action-relevant concept is well-taken. If some people and times in the past were enormously influential, that does not matter to us, today, for the purposes of action. Nor does it matter if people in a million years' time have enormous opportunities to have an impact, if we are certain that we cannot pass resources that far into the future.[30]

So one might instead defend a restricted claim:

> Restricted-HH: We are among the very most influential people, out of the very large number of people who will live over the coming thousand years.

---

[29] Greg Lewis raises this point in the comments section of my blog post, 'Are we living at the most influential time in history?'

[30] Though we should be careful about claiming that we are 'certain' that we cannot pass resources even further into the future than one thousand years' time. Given how enormous your influence would be if you were able to invest those resources over such large timespans (perhaps ending up with a significant fraction of global wealth), even a very low probability of attaining that outcome could have very great expected value.

In this statement, ‘very large’ might refer to the billions to hundreds of billions of people who are to come over the next thousand years. (Which is small compared to the trillions upon trillions of people who would live if we took to the stars.)

This claim is *much* weaker than the original HH, and is therefore much more plausible a priori. However, a similar line of argument can be made against Restricted-HH as can against HH. It’s a priori unlikely that, of all the people and times to come over the next thousand years, it is we, today, who can do the most good in expectation with a unit of resources.[31] Moreover, as we shall see in section 5, it’s much harder to come up with a convincing argument for the claim that we, now, are far more influential than people in the centuries to come, than it is to make arguments for the claim that those in the coming millennium are far more influential than those in the millenia that follow. The arguments that I’ll cover in section 5 — that we are unusually early on in history, on a single planet, and at a period of unusually high economic and technological growth — would plausibly support the idea that any time in the next few centuries is as influential as today is.

4. The inductive argument against HH

In addition to the base-rates argument against HH, which relies on priors and claims we shouldn’t move drastically far from those priors, there’s a positive argument against HH, which gives us evidence against HH, whatever our priors. This argument is based on induction from past times, as follows:

> P1. The influentialness of comparable people in the past has been increasing over time, with increasing knowledge and opportunities being the most important factor.
> P2. We should expect our knowledge and opportunities to continue into the future.
> C. So we should expect the influentialness of those future people who we can pass resources on to be greater, too.

Let’s begin with the first premise: that the influentialness of comparable people in the past has been increasing over time, with increasing knowledge and opportunities being the most important factor. I think it’s relatively clear, for example, that we should prefer that a well-educated European living in 1600 pass philanthropic resources to us, today, rather than attempting to directly do good with those resources. At least three considerations support this view. First, the opportunities available to this person in 1600 were in general less high-leverage than the opportunities available to us today.[32] In particular, they would have had few opportunities to shape the long-run future: most of the existential risks that someone faced in 1600, such as an asteroid collision or supervolcanic eruption, were not known of at the time,

---

[31] This claim also is not a reasonable interpretation of Parfit’s comments. Parfit’s claim was that the next few centuries are unusually important: if the claim was merely that they were unusually important out of the next thousand years, that would not be a very strong claim at all.

[32] An exception might have been the opportunity to shape the values of the time, which are plausibly persistent for a long time period, including via religious institutions.

and would have been impossible to do anything about even if they were known. Second, and even more importantly, was their impoverished scientific understanding. A well-educated European in 1600 still believed, for example, that witches could summon storms, that werewolves could be found in Belgium, that mice are spontaneously generated in piles of straw, that a murdered body will bleed in the presence of the murderer, and that the sun revolves around the Earth.[33] They did not have the modern scientific method, physics, biology, chemistry, or social science, and instead their worldview was theocentric. They could not have known about the vastness of the future, nor make reasonable guesses about how to positively influence the long-run future.

Finally, and most importantly of all, is moral progress. Those in 1600 believed that women and people of other races and religions are of lesser moral standing than European Christian men. Intense social hierarchy, inequality, and slavery were regarded as the natural and just way of things. Homosexuality and premarital sex were regarded as deeply immoral. The idea of liberalism had not been developed. Torture was commomplace and celebrated, as was cruel punishment and violence against heretics. In general, the moral beliefs that were widespread at the time were grounded in a narrow understanding of Christian doctrine that we would now deplore.[34] For these reasons, the altruistic priorities of someone in 1600 would have been radically different from what we would think today.

When we look over a shorter timespan — say, looking back to 1970, or to 1920 — the argument is not quite as clear-cut. In particular, possible existential risks from new technology were knowable at those times.[35] But even still, there is a good argument for thinking that we are in a much better position to have a positive impact today than we could during those times. Again, our opportunities are better today than they were before — there was little that one could do decades ago to work on risks from misaligned artificial intelligence or synthetic biology. Our scientific knowledge is considerably better, including our understanding of the nature of existential risks: the idea of a nuclear winter was only developed in the 1980s, and the scientific consensus regarding anthropogenic climate change was only developed over the 1970s to 1990s. We have only learned of the impressive success of deep learning as paradigm for progress in artificial intelligence (and therefore learned more detail on the shape that technical artificial intelligence safety work ought to take,[36] in the event that deep learning leads to artificial general intelligence) in the last decade. And moral progress has continued, too. Cosmopolitanism has continued to become more widespread, and rights for women, minorities and people of all sexual identities have been progressively secured. On the intellectual side of moral progress, most notable is that population ethics only became a serious field of inquiry after the publication of *Reasons and*

---

[33] These facts are taken from Wootton (2005, p.6)

[34] For more on moral change over time, see Morris (2015) and Pinker (2011, esp. Chapter 4).

[35] For example, possible risks from artificial intelligence were identified by the pioneers of computer science, such as Alan Turing and I.J. Goode.

[36] See Amodei et al (2016).

Persons in 1984. But without that work it's hard to believe that someone would have reliably prioritised existential risk reduction over other altruistic activities.

So, just as we concluded with longer timespans, it seems that our influentialness has increased since 1920 or 1970. And the primary driver of this increased influentialness is our increasing scientific and moral knowledge. But we should strongly expect this increase in knowledge to continue into the future. As a general matter, people in the future, who we could pass our resources to, will plausibly be far smarter and more informed than even the most brilliant minds of today: they may be the beneficiaries of enhancement technologies, more powerful intelligence-augmenting tools like computers and artificial intelligence, better educational methods and better nutrition. And they will very likely have a radically larger edifice of scientific knowledge to base their decisions on, with decades or centuries of further moral progress, including on the very particular question of how best to use resources to make the world better.[37]

But, as well as the general point, we also can identify specific, crucial gaps in our current understanding. On the empirical side, we still don't know how developments in synthetic biology and AI will play out; we have a very poor understanding of how resilient civilisation is, in terms of both how large a disaster would be required to kill everyone, or how likely civilisation would be to recover after a major but non-existential catastrophe; and we have very limited understanding of good forecasting practices beyond a few years. On the moral side: we have no good theoretical understanding of how to evaluate tiny probabilities of enormous amounts of value; nor do we have a compelling account of how to deal with the possibility of creating infinite amounts of value; there has been very little work trying to understand the expected value of the continuation of human civilisation; and we have very limited understanding of how to correctly make decisions in the face of normative uncertainty. Insights on these questions could all significantly change how we would choose to prioritise our altruistic efforts. And this list I've given is just a tiny subset of all the crucial questions that are still unanswered.[38] But if we should expect our knowledge and understanding to significantly increase over the coming decades and centuries, just as it has over the previous decades and centuries, and that knowledge and understanding is typically the dominant factor in terms of how much good an individual can do with a unit of resources, then we should think that future people will be more influential than we are.

---

[37] Might one worry that such reasoning would lead one *never* to use one's resources philanthropically? I don't think so. The pace at which we are improving our understanding of the world will inevitable slow, and may inded have already been slowing over the course of the last 50 years. Once we are reaching the plateau, we might well want to spend significant proportions of our resources at particularly pivotal moments in time. What's more, with every generation we wait, we have less future to be able to positively influence; this is a cost, and at some point in time, the cost of delay will outweigh any benefits thereby gained. This is explored more thoroughly in Trammell (ms).

[38] For a longer list, see Greaves et al (2019).

The claim that we're at the hinge of history (at least as I have defined this idea) is therefore in tension with another view of Parfit's: that we may be at just the beginning of intellectual and moral progress. In the final paragraph of *Reasons and Persons*, Parfit commented that:[39]

> There could clearly be higher achievements in the struggle for a wholly just world-wide community. And there could be higher achievements in all of the Arts and Sciences. But the progress could be greatest in what is now the least advanced of these Arts or Sciences. This, I have claimed, is Non-Religious Ethics. Belief in God, or in many gods, prevented the free development of moral reasoning. Disbelief in God, openly admitted by a majority, is a recent event, not yet completed. Because this event is so recent, Non-Religious Ethics is at a very early stage. We cannot yet predict whether, as in Mathematics, we will all reach agreement. Since we cannot know how Ethics will develop, it is not irrational to have high hopes.

I agree with Parfit's optimism here. But if there is a good chance that future generations will have discovered many ways in which we are misguided, scientifically and morally, then we have a strong argument for thinking that they will be able to spend resources in higher-value ways than we can, and are therefore more influential than we are. Indeed, if we are morally mistaken enough, perhaps even our best-intended efforts today could be doing harm.

In the discussion above, I looked at how the value of opportunities to shape the long-run future (in particular by reducing existential risks) have changed over time. I believe this is the most relevant question for an inductive argument, because I believe that opportunities that shape the long-run future tend to be the highest value opportunities.

But, as a sanity check, we could also ask how influentialness has varied over time if we just restrict ourselves to attempts for a person to make their own time better. This should seem to be a more favourable case for the idea that influentialness is going down over time, because the world has gotten so much richer over time, and we have made so much progress on so many of the social problems that affect the people of the day.[40] But, even so, I think that the opportunities we have to benefit individuals alive at the present time are far better than the opportunities that were available in 1970 to benefit people alive in 1970, those that were available in 1920 to benefit people alive in 1920, and similarly for 1600 and indeed for any other time in the past.

Again, the two principal factors that have caused this are increasing technology (allowing us to purchase a wider variety of goods) and increasing scientific knowledge. Since the 1970s we have a far greater understanding of how to improve the lives of those in poor countries, as a result of improvements in epidemiology, public health, economic theory and the randomista movement. And, especially since the 1920s, we have far more and better opportunities to benefit very poor people, especially via medical technology. In the 1920s, we simply did not have the technology to provide cheap lifesaving medicines to the poorest people in the world.

---

[39] Parfit (1984, p. 454).

[40] Pinker (2011)

In the 1970s we did, and in some cases the cost-effectiveness of the opportunities available (such as smallpox eradication) were enormous. But it's not clear at all that we could have identified these opportunities ex ante: we could not have known that global health would have been the enormous success that it was, and the long history of failures in aid spending suggests that altruistically minded individuals of the time were not able to reliably identify the actions that would turn out ex post to have enormous positive impact.

5. Some arguments for HH

So far I've given two arguments against having a significant degree of belief in HH. This section will consider two additional arguments that one might raise in favour of HH.[41] I think that these arguments are fairly strong, and they caveat my argument, giving us reason to think that now is quite influential. But they are not sufficiently strong to warrant giving significant credence to HH.

Living on a single planet.

We currently live in a civilisation that exists on a single planet. If, in the future, we take to the stars and form an interstellar civilisation, then the vast majority of people who ever live will be part of a multiplanetary civilisation. So this is a clear and objective way in which the present time is very unusual. Moreover, there are a number of reasons for thinking that times when we live on a single planet are unusually influential. First, as Parfit mentions in his talk at the Oxford Union, civilisation's period on a single planet might be one of unusually high existential risk. A single planet means a single point of failure. So, for example, a collision between Earth and a large asteroid could end human life on this planet, but it would not pose a risk to human life on other planets. Second, it means we are at a period of unusually low population and economic power (compared to a vast interstellar civilisation), so any

---

[41] One argument I won't consider here is that there is an annual risk of human extinction or lock-in, so earlier centuries are more likely to have people in them, and to be prior to some lock-in event. We of course need to take that into account, but in Trammell's model, that is taken into account in the 'δ' term, which represents our 'rate of pure time preference', rather than in the influentialness of a time.

I also won't discuss further the question of how to set fundamental priors in this context. For more discussion of that, see the comments from Toby Ord, and my replies, on my blog post 'Are we living at the most influential time in history?'. Ord proposes using a Jeffreys prior to model the chance that we are among the most influential people. I don't discuss Ord's proposal simply because it would involve a long digression into a proposal that I ultimately think is a red herring. I think there are reasons for thinking that people at earlier times are more likely to be influential, but these are given by the arguments I present in this section and shouldn't be built into one's fundamental prior. Moreover, Ord's proposal faces technical issues: on his account, the prior one chooses for being among the most influential people is highly sensitive to the reference class chosen; without further modification, it would generate inconsistent probability assignments to multiple hypotheses; it would have predicted that the most influential people were very likely in the past; and it involves treating the superlative 'most influential' very differently from other arbitrary superlatives, like 'most beautiful', 'funniest', and so on.

I also won't discuss the fact that we can affect the future but not the past, so those who live later on simply have fewer future lives that they can affect. So for this reason, we should expect earlier people to have more influence than later people. I don't discuss this in the body text because, though it is clearly true, it will not make a major difference to the argument: the first person who ever lives will only have twice as many lives ahead of her as the median person, who, on the scenarios we are considering, had many trillions of people before her. So taking this consideration into account would not make a significant difference to our assessment of HH.

resources we have are an unusually large fraction of total resources at the time, which might give us an unusual ability to influence the course of civilisation as a whole. Third, it means that any one person is able to communicate almost instantaneously with almost anyone else in civilisation. In contrast, once a civilisation is interstellar, communication with other solar systems will take many years: the closest solar system to our own is four light-years away, the galaxy is one hundred thousand light-years across, and the distances between galaxies is measured in the millions of light-years. Again, this gives individuals alive during the present time an unusual potential opportunity to influence civilisation as a whole.

This is an important argument — in particular in the form that emphasises how small civilisation is, today, compared to future civilisation — and I believe it should cause a major update in favour of HH, away from our prior.

However, its strength should not be overstated. First, the reduction in existential risk in virtue of being interplanetary may be relatively small. For example, even absent any technological intervention, the annual risk from an asteroid collision without any human action is about one in a hundred million. What's more, we have now detected all of the near-Earth asteroids over 10km in size, over 95% of near-Earth asteroids larger than 1km, and there are numerous methods by which we could deflect one if detected. Moreover, even if there were a major asteroid collision, it doesn't seem very likely that it would cause the extinction of the human race. Many mammal species survived the Chicxulub impactor (which caused the extinction of the dinosaurs), as did many reptiles and fish, and humans have an enormous population, with one hundred times the biomass of any large wild animal that's walked the Earth,[42] spread out across a wide diversity of environments, with the technological and scientific capability to weather a long period of global cooling following an asteroid impact.[43]

Instead, those who work on existential risk tend to believe that the most likely risks come from *omnicidal agents*, in particular from omnicidal superintelligence. In *The Precipice*, for example, Toby Ord gives an estimate of total existential risk this century as being at about one in six, with almost two-thirds of that coming from risks from misaligned superintelligence.[44] For this risk, there is not much additional benefit from being an

---

[42] Wilson (2002, p. 29).

[43] Might it be the case that the *existential* risk from asteroid collision is much higher than the extinction risk from asteroid collision, because of the possibility that an asteroid impact destroys advanced civilisation, and we never recover? I do think that the probability of unrecovered civilisational collapse from an asteroid impact is higher than the probability of extinction. But I think that, even if advanced civilisation were destroyed, it is very likely that we would recover. Agriculture was developed indepedently in several different locations within a short time period, suggesting that it was not a bottleneck; and it took merely thousands of years (which is a short period of time compared to the typical mammalian species lifespan of around half a million years) for us to move from agricultural to industrial civilisation. Over the course of the agricultural era we also saw sustained (though slow) economic growth and technological development, the rate which seemed to be continually accelating: for more discussion, see Roodman (2020). One might worry that we have used up so many fossil fuels that we could not rely on them to re-industrialise, but this is not true. For example, the 1.2 billion tons of recoverable coal in the US's North Antelope Rochelle mine alone is more than total global coal use between 1770 and 1830. This isssue is explored in much more depth by Rodriguez (ms).

[44] Ord (2020)

interplanetary civilisation: though it would be harder for a misaligned superintelligence to eliminate all human life across two planetary systems than one, it would not be much harder. Similar considerations would hold for other existential risks, such as those from perennial totalitarianism, convergence on the wrong moral view, and from sufficiently powerful doomsday cults.[45]

Second, the period where civilisation is close together enough that it is easy for one individual or group to influence the whole rest of civilisation may be quite prolonged. We do not know how hard it will be to become a truly interstellar civilisation. But when we think about future progress we should bear in mind that the last 250 years of rapid technological and economic progress is a historical anomaly, which could well slow into the future. Indeed, we have already some data that frontier growth is slowing,[46] and demographic changes predict a slowing or even negative growth rate by the end of the coming century.[47] We may well therefore be primarily Earthbound for many thousands or tens of thousands of years to come.

What's more, even when we start settling areas outside of Earth, we may well be primarily confined to the solar system for considerably longer again. And while we are still primarily limited to our solar system, it is still comparatively easy for one individual or group to communicate with and influence the rest of civilization: for example, it takes only one hour for light to traverse even the full diameter of the asteroid belt.[48]

Unusually fast economic and technological progress.[49]
The world growth rate is around 3.5% per year.[50] It is not plausible that we can sustain such a high growth rate indefinitely into the future. To see this, suppose that in the future the world economy will grow by (merely) 2% per year indefinitely. If so, then after 10,000 years there would be 10^19 times present-day GDP for every atom in the galaxy. This is not a plausible outcome.

I believe that this appeal to rapid economic and technological progress is the strongest argument in favour of thinking that we live at an unusually influential time. The present time is certainly highly *distinctive* in terms of its growth rate. And even if you only think it 10% likely that the most influential time is at a period of unusually high economic growth, then you should give at least a 10% credence to the idea that we are among the most influential 10,000 years. And there are positive arguments for thinking that we should expect the most

---

[45] Toby Ord discusses this more in *The Precipice*, chapter 7.

[46] See, for example, Gordon (2017) and Vollrath (2020).

[47] Jones (2020).

[48] Where the asteroid belt is around 3 Astronomical Units from the sun, with one AU being 8 light-minutes.

[49] This argument originally comes from Hanson (2009).

[50] See, for example, Roser (2020).

influential times to be those of unusually fast technological progress: in particular, if the fate of the future is determined by how we manage the invention and deployment of particular technologies (such as artificial intelligence, or particularly dangerous weapons), then at periods of unusually fast technological progress, we are moving faster through the space of all technological inventions, and are therefore more likely to discover one of the critical technologies.

However, there are still caveats that need to be made. First, crucially, though this argument indicates a way in which the present time is very unusual, and therefore potentially very unusually important, it doesn't give us grounds for thinking that the present time is the very most important time, rather than some future century over the coming few millennia. And that is the action-relevant question.

Second, there's an argument for expecting longtermist altruists to be *less* influential during periods of fast economic growth. In a very stable environment, it is easier to make and fund very long-term projects. And, in a world where most people only care about the short term (especially the period when they live), we should expect that projects that only have long-term payoffs will be the most neglected, and there would be low hanging fruit for longtermists to pick in this area. But if we live at a period of rapid change, that advantage that longtermists have is lost: it's much harder, or impossible, to have reliable long-term plans, because doing so would involve being able to predict inherently unpredictable changes in the technological landscape.[51]

Summing up

We have seen that there are some compelling arguments for thinking that the present time is unusually influential. In particular, we are growing very rapidly, and civilisation today is still small compared to its potential future size, so any given unit of resources is a comparatively large fraction of the whole. I believe these arguments give us reason to think that the most influential people may well live within the next few thousand years. But these arguments are far from watertight, and they do not give us very strong reasons for thinking that we, now, are among the most influential people ever, rather than people in the centuries or millennia to come. But we do have positive reasons — namely, our predictably increasing knowledge and opportunities, as canvassed in my inductive argument — for thinking that the most influential people are yet to be born.

---

[51] More fundamentally, there are also serious questions about how to make quantitative comparisons of economic power over such long timescales. There are very many things that the rich can buy today that people in the past could not, but there are also some things that people in the past could buy that those in the present could not. (Someone with an overwhelmingly strong preference for dodo meat would regard some people in the past as far richer than we are today.) So we could try to instead pick some objective indicator of economic growth, such as energy capture. But any such indicator seems to have problems. For example, over the last 20 years the US economy has grown by about 50%, but it has not increased its energy consumption, because it has become more energy efficient over time.

6. Conclusion

Because civilisation has such a long future ahead of it, and because resources grow over time (both in absolute terms, as a fraction of the world economy), there is a strong *prima facie* case for impartial altruists to invest their resources, passing them on to future people to use philanthropically. However, if we thought that the present time was exceptionally influential — or even the most influential time — this would be a strong counterargument. Parfit seemed to believe this, as he indicated in his comments in *Reasons and Persons*, *On What Matters: Volume 2*, and in a talk for Giving What We Can. Assessing whether this is true is crucially important, deserving of far more attention than I have been able to give it in this article. There are some good arguments for thinking that our time is very unusual, if we are at the start of a very long-lived civilisation: the fact that we are so early on, that we live on a single planet, and that we are at a period of rapid economic and technological progress, are all ways in which the current time is very distinctive, and therefore are reasons why we may be highly influential too. But the claim that we are among the *most* influential people is considerably stronger again, and does not seem warranted. I have given two arguments for scepticism about this stronger claim: first, that our prior on this claim should be very low, and that the evidence we have from moving away from this prior is not sufficiently strong; second, that if we look at how influentialness has been changing over time, we should expect it to continue into the future as our knowledge and understanding improves over time.

In *On What Matters: Volume 2*, Parfit comments that:

> Life can be wonderful as well as terrible, and we shall increasingly have the power to make life good. Since human history may be only just beginning, we can expect that future humans, or supra-humans, may achieve some great goods that we cannot now even imagine. In Nietzsche's words, there has never been such a new dawn and clear horizon, and such an open sea.

In Parfit's discussion, 'open sea' refers to the space of possible goods that future humans or supra-humans could enjoy. And, though it is undoubtedly true that life to date has sampled from only a tiny corner of the menu of possible experiences, the even more important 'open sea' that Nietszche himself referred to[52] is the newfound potential for knowledge given our liberation from a theocentric worldview.

We are only just starting out on this intellectual voyage. There is still far more to learn and understand. Over time we should expect to radically change our understanding of the good, and of how to promote it. Just as our powers to grow crops, to transmit information, to discover the laws of nature, and to explore the cosmos have all increased over time, so will our power to make the world better — our influentialness. And given how much there is still to understand, we should believe, and hope, that our descendents look back at us as we look back at those in the medieval era, marvelling at how we could have got it all so wrong.

---

[52] Nietzsche (2001), p. 199, aphorism 343.

Bibliography

Dario Amodei et al. Concrete problems in AI safety. arXiv preprint arXiv:1606.06565 (2016).

Daniel J. Benjamin et al (2018). Redefine statistical significance. Nature Human Behaviour.

Nick Bostrom (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge.

Nick Bostrom (2003). Astronomical Waste: The Opportunity Cost of Delayed Technological Development. *Utilitas* **15**(3):308–314.

Nick Bostrom (2014). *Superintelligence: Path, Dangers, Strategies.* Oxford: Oxford University Press.

WM. Theodore de Bary and Irene Bloom (1999). *Sources of Chinese Tradition, Vol 1. From Earliest Times to 1600.* New York: Columbia University Press.

Robert J. Gordon (2017). *The rise and fall of American growth: The US standard of living since the civil war.* Princeton University Press.

Hilary Greaves and William MacAskill (ms), The Case for Strong Longtermism.

Hilary Greaves, William MacAskill, Rossa O'Keeffe-O'Donovan and Philip Trammell (2019). A Research Agenda for the Global Priorities Institute. Available at https://globalprioritiesinstitute.org/wp-content/uploads/gpi-research-agenda.pdf

Robin Hanson (2009). Limits to growth. *Overcoming Bias*. Available at http://www.overcomingbias.com/2009/09/limits-to-growth.html

Paul Horwich (1982). *Probability and evidence*. Cambridge University Press.

Harold Jeffreys (1961). *The Theory of Probability.* Oxford, England.

Charles I. Jones (2016). Life and Growth. *Journal of Political Economy* **124**(2):539–578.

Charles I. Jones (2020). The End of Economic Growth? Unintended Consequences of a Declining Population. Available at https://web.stanford.edu/~chadj/emptyplanet.pdf

William MacAskill (2019). ‘Longtermism’. *Effective Altruism Forum.* Available at https://forum.effectivealtruism.org/posts/qZyshHCNkjs3TvSem/longtermism

William MacAskill (ms). Moral Option Value

Andreas Mogensen (2019). Doomsday rings twice. Available at https://globalprioritiesinstitute.org/wp-content/uploads/2019/Mogensen_doomsday_rings_twice.pdf

Ian Morris (2015). *Foragers, Farmers, and Fossil Fuels: How Human Values Evolve.* Princeton: Princeton University Press.

Friedrich Nietzsche (2001). *The Gay Science*. Edited by Bernard Williams. Cambridge: Cambridge University Press.

Toby Ord (ms). The Edges of our Universe.

Toby Ord (2020). *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury

Derek Parfit (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Derek Parfit (2011). *On What Matters*, *Volume 2.* Oxford: Oxford University Press.

Thomas Piketty (2013). *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Steven Pinker (2011). *The Better Angels of Our Nature.*

Luisa Rodriguez (ms). On Civilisational Collapse and Recovery.

David Roodman (2020). Modeling the Human Trajectory. OpenPhilanthropy.org. Available at https://www.openphilanthropy.org/blog/modeling-human-trajectory

Max Roser (2020). Economic Growth. OurWorldInData.org. Available at https://ourworldindata.org/economic-growth

Carl Sagan (1994). *Pale Blue Dot: A Vision of the Human Future in Space*. New York.

Christian J. Tarsney (2019). The Epistemic Challenge to Longtermism. Available at https://globalprioritiesinstitute.org/wp-content/uploads/2019/Tarsney_Epistemic_Challenge_to_Longtermism.pdf

Teru Thomas (ms). Self-location and Objective Chance.

Philip Trammell (ms). Discounting for Patient Philanthropists. Available at https://philiptrammell.com/static/discounting_for_patient_philanthropists.pdf

Dietrich Vollrath (2020). *Fully Grown: Why a Stagnant Economy Is a Sign of Success*. University of Chicago Press.

Edward O. Wilson (2002). *The Social Conquest of Earth*. New York: Liveright.

David Wootton (2015). *The Invention of Science: A New History of the Scientific Revolution*. Penguin UK.