

Authenticity, Meaning and Alienation: Reasons to Care Less About Far Future People

Stefan Riedener

Abstract. The standard argument for longtermism assumes that we should care as much about far future people as about our contemporaries. I challenge this assumption. I first consider existing interpretations of ‘temporal discounting’, and argue that such discounting seems either unwarranted or insufficient to block the argument. I then offer two alternative reasons to care less about far future people: caring as much about them as about our contemporaries would make our lives less authentic and less meaningful. If I’m right, this undermines at least the strongest versions of longtermism. Plausibly, we should do much more for the far future than we’re currently doing. Still, we often ought to act for contemporaries’ sakes, even if longtermist actions would expectably do more good.

Keywords. longtermism · temporal discounting · authenticity · meaning of life · alienation · special relationships · fetishism · non-aggregationism

1. Introduction

What ought we to do today? Take the United States, and consider just two of their options. At the time of writing, Russia is attacking the Ukraine. The Ukrainians must defend themselves, accommodate refugees and start to rebuild their country. The US could help them by sending weapons or money or welcoming escapees. At the same time, AI technology is progressing. There’s a risk that an AI will go out of control, and cause enormous harm. The US could slightly reduce that risk by investing in AI safety. What ought they to do? According to longtermism (as I’ll use the concept), in situations like these, what we ought to do depends almost exclusively on the very long-term effects of our actions. What matters is how the US’s decision affects the next millions of years. It matters little, in itself, how it affects the coming year, century or even millennium.

The standard argument for this view builds on four simple premises.¹ First, the long-term future concerns the fate of an enormous number of people.² Humanity might colonize the universe, survive for billions of years, and live lives of unimaginable quality. Or we might soon go extinct, or remain a small group of miserable souls leading a wretched form of existence. The difference might

¹ Roughly this argument is given e.g. in MacAskill (2022: ch. 1), although MacAskill omits the fourth premise, and thus doesn’t draw a similarly strong conclusion. For a similar argument, with a similar conclusion, see e.g. Greaves et al. (2021) or Greaves and MacAskill (2021).

² I’ll follow much of the existing literature by making two simplifications. First, I’ll set aside non-human animals, although they clearly matter too. Second, I’ll call our far future descendants ‘people’, although they needn’t be humans (or homo sapiens) by any means.

concern, say, a septillion (10^{24}) people. Second, in general, far future people should matter to us just as much as contemporaries do. Consider spatial distance. If we can help someone, it shouldn't matter to us whether they're here or ten thousand kilometers away. The same is true for temporal distance. If we can do someone good, it shouldn't matter to us whether they live now or in a million years. Third, we can affect the existence and quality of life of these far future people. Of course, we can't predict long-term effects with any confidence. But in very many decision situations, at least, certain actions seem to have a *somewhat* higher probability than others of positively impacting the long-term future. And even just a one in a billion ($1/10^9$) chance to affect a septillion people, say, in expectation still affects a quadrillion (10^{15}). So even minuscule increments in probabilities *ex ante* affect *a lot*. Fourth, typically, if we can expectably affect the existence and life-quality of billions or quadrillions of people, we simply ought to do it. Our reasons to do so trump just about any other consideration. They're more important than our reasons to affect the existence and life-quality of fewer people—or to keep our promises, show loyalty to our friends, or even to refrain from harming.

Prima facie, the upshot of this simple reasoning is (what I'll call) longtermism: typically, we simply ought to do what's expectably best in the long term. All other considerations—in particular, our effects on the coming century or so—are largely irrelevant in themselves, or relevant only as tie-breakers when all of our options are *ex ante* equally good in the longer term.³ Take the US decision again. According to many experts, AI poses one of our greatest extinction risks. This risk can be reduced. Yet attempts at doing so are still very sparse. So investments in AI safety seem among the best longtermist interventions.⁴ By contrast, housing Ukrainian refugees or building up Ukrainian churches doesn't seem to affect the long-term prospect much. And while military aid might have long-term benefits in deterring aggressive warfare, it also involves a real risk of inducing a catastrophic escalation.⁵ So the long-term benefits of helping the Ukraine seem decidedly less clear. Presumably, according to longtermism, the US ought to invest in AI safety. And in any case, whatever the ultimate upshot, the 43 million Ukrainians themselves are practically irrelevant for the decision. The US's dominating concern should be to marginally increase the chance that in a million years, in remote corners of space, quintillions of our descendants enjoy fantastic lives.

³ There's currently no consensus about what exactly 'longtermism' means. The argument I've just sketched is very natural. I think it's not uncommon. And it's the conclusion of this argument that interests me here. So, I use 'longtermism' to denote this conclusion. Some people use it to denote the weaker view that the long-term future should be *a* key moral priority of our time; they use 'strong (deontic) longtermism' to denote the view I have in mind (see e.g. MacAskill 2022: 4; Greaves and MacAskill 2021). Depending on what it means to be 'a key priority', I think my arguments are compatible with this weaker view. So in this alternative terminology, I don't object to longtermism, but to strong (deontic) longtermism only.

⁴ See e.g. Ord (2020: chs. 5 and 6).

⁵ This has been a guiding concern of the White House in deciding what help to provide. See e.g. <https://www.nytimes.com/2022/09/17/us/politics/ukraine-biden-weapons.html> (last accessed Jan 11, 2023).

This is an extremely revisionist idea, to put it mildly. If the above assumptions are right, it would often imply we should take radically different actions than we're currently taking. And it would just about always imply we should base our actions on radically different considerations than we're currently doing.⁶ Notably, the just-mentioned argument isn't simply utilitarian. It applies on any view that gives relevantly nontrivial weight to promoting expected total welfare. Still, it involves a number of controversial assumptions. Some people might oppose population ethical presuppositions of the argument as I presented it—e.g., the idea that we have strong reasons to bring people into existence. Others might object to decision-theoretic assumptions—e.g., the idea that you ever ought to prefer such tiny-probability high-payoff options. Still others might oppose empirical premises—e.g., the idea that we aren't entirely clueless about our long-term effects. In this paper, I'll grant the longtermist all of these assumptions. For what it's worth, I tend to think they're either plausible, not ultimately essential to the argument, or not pointing to the fundamental problem about it. So instead, I focus on the second premise of the above argument—the thought that we should care just as much about far future people as about our contemporaries. This is a very revisionist idea itself. As a matter of fact, we care much more about people in the Ukraine, say, than about our distant descendants. I want to offer a philosophical justification of our unequal concern, and explore how this affects the longtermist argument.

To that end, I'll first consider existing understandings of 'temporal discounting', and argue that on these interpretations, our discounting seems either unwarranted or insufficient to block the argument (section 2). I'll then offer two alternative reasons to care less about far future people. The first consideration concerns our *reactive* relation to far future goods: the way in which we respond to, or appreciate, those goods. I'll argue that our caring equally much about far future people would be problematically inauthentic (section 3). The second consideration concerns our *proactive* relation to far future goods: the way in which we potentially bring those goods about. I'll argue that the way in which we could do good to far future people contributes much less meaning to our lives than the way in which we can benefit contemporaries (section 4). I'll end by discussing objections (section 5). If I'm right, my arguments are compatible with the view that we should be doing much more for the far future than we're currently doing. Still, they imply that in many situations, we ought to act for (roughly⁷) contemporaries' sakes, even if longtermist actions would expectably do much more good overall.

⁶ To illustrate this with our initial example: as of December 2023, US aid to the Ukraine totaled \$68 billion, and the White House has asked Congress for another \$37.7 billion; in contrast, recent estimates suggest that *global* annual spending on reducing existential risks from AI is between \$10 and \$50 *million*. See e.g. <https://www.csis.org/analysis/aid-ukraine-explained-six-charts> (last accessed Jan 11, 2023); Ord (2020: 58 fn. 55).

⁷ In what follows, by 'contemporaries', I mean people living now *or* in the nearer future (e.g., the next couple of hundred years or so). By 'far future people', I mean people living in the very distant future (e.g., ten or a hundred thousand years from now).

2. Preliminaries and Premises

The idea that we should care less about future people has been discussed a lot. The simplest interpretation of it is in terms of temporal discounting, in the form of a ‘pure time preference’. On this interpretation, temporal distance itself is fundamentally normatively relevant: if you can provide two goods x and y , but y would occur in the more distant future, this itself is a reason to prefer x . In the long term, such discounting—even by small factors—can have dramatic effects. If goods at any time are 99.99% as important as goods a year earlier, say, our interests are more than 10^{43} times as important as those of people in a million years. This would arguably undercut the far future’s overwhelming importance.

However, few people find such a pure time preference compelling. Temporal distance itself just doesn’t seem fundamentally normatively relevant. Here’s one way to see this. Suppose that, without anyone realizing anything, the universe will freeze for a millennium at midnight today. Subjectively, the night will feel as short as any. But actually, the macroscopic world will stand still for a thousand years. It doesn’t seem that you’d then have less reason today to try to benefit people ‘tomorrow’—e.g., to buy your partner a present for their upcoming birthday. Temporal distance, *in itself*, just doesn’t seem to fundamentally matter.⁸

If anything, it’s other properties which matter intrinsically, perhaps generally *correlate* with temporal proximity, and allow us to care less about future people. The most salient idea is personal relationships. Intuitively, you ought to care more about someone the closer your relationship is to them. So perhaps we generally ought to care less about far future people because we have a weaker relation to them than we have towards our contemporaries. Mogensen (2022) provides one way to flesh this out. Intuitively, you ought to care more about someone the more closely you’re *related* to them. Now consider the perspective of humanity’s current generation as a whole. Perhaps we’re less closely related to future generations the further away in time they are—just like you’re less closely related to your great-grandchildren than to your grandchildren, say. And so perhaps our current generation as a whole ought to care more about near future people than about people further away. Mogensen himself doesn’t apply this idea to longtermism, or claim that it undermines the above rationale. But one might think it does.

However, I’m skeptical. One might doubt that kinship is normatively important in the first place, that it makes sense to speak of entire generations as being related to each other, or that this collective perspective is sufficiently relevant to the overall longtermist agenda.⁹ But even setting these worries aside: the idea simply doesn’t seem to allow us to discount enough. Note that you’ll never be less closely related to any of your descendants than you are to a completely unrelated

⁸ For prominent critical discussions of pure time preference, see e.g. Sidgwick (1907: 381, 414), Ramsey (1928), Rawls (1999: §45), Parfit (1984: App. F), or Broome (1991, §8.3). For a helpful survey, see Greaves (2017).

⁹ Why focus on entire generations in the first place? The reason is that it’s not clear whether as an individual, you’re less related to a random person in ten thousand years, say, than to a random person of the present.

contemporary. So (as Mogensen emphasizes) in light of your kinship, you never ought to care less about future people than about such unrelated people of the present. But intuitively, we ought to care very considerably about the latter. We ought to care greatly about people in the Ukraine, say, even if we aren't related to them.¹⁰ The prospect of affecting a quadrillion such unrelated contemporaries would still seem more pressing than almost anything else. If we ought to care no less than this about far future people, they still seem dominantly important.

A similar issue arises with other salient accounts. Consider gratitude, compensation or friendship. Intuitively, you ought to care more about people who've benefitted you, people you've harmed, or people who are your friends, than about strangers with whom you've never interacted in any way. And plausibly, in general (though not with the same kind of necessity as in the case of kinship), we have stronger relationships of these kinds to contemporaries than to far future people. In light of the global economy, very many contemporary people have benefitted us in some way.¹¹ We have harmed very many people.¹² And societies like the US and the Ukraine might be regarded as friends of some kind. So plausibly, in light of these considerations, we generally ought to care more about contemporaries than about people in the far future. But again, you ought to care considerably about strangers, to whom you owe no obligations of gratitude, compensation or friendship. If you can affect a quadrillion such strangers, that seems to trump almost anything. These kinds of 'discounting' for special relationships don't seem to block our argument either.¹³

So we face a challenge. The simplest interpretation of the relevance of temporal distance is implausible, and more complex interpretations don't seem to get us near enough discounting to undermine the longtermist conclusion. The general problem is that considerations like kinship, or gratitude, compensation and friendship, give us reason to care *more* about some people than about completely unrelated contemporaries. But we should care considerably about the latter. So what we need, it seems, is features of far future people which mean we ought to care *less* about them than about such contemporary strangers. I think the two considerations I'll introduce in what follows have precisely this form. Neither of them correlates linearly, or with any metaphysical necessity, with temporal distance. But at least for contingent reasons, they're generally salient for far future people. So in the next two

¹⁰ Of course, given our common origins, you're somewhat related to every human being. To eliminate this point, suppose half of humanity descended from (perfectly human-like) aliens who arrived here long ago, and you aren't related to some contemporaries at all. You still ought to care considerably about them.

¹¹ For an illustration of this, see e.g. Jacobs (2018). I thank Adam Lovett for this reference.

¹² For a famous defense of this idea, see e.g. Pogge (2002).

¹³ Something similar applies to other possible justifications of discounting. Parfit discusses whether you may care less about your future selves insofar as you're less psychologically connected to them (1984: 313-317). We might turn this into an argument for discounting future generations, by considering the prudential perspective of humanity as a whole. However, your future self will plausibly never be less connected to you than a complete stranger. And you ought to care considerably about strangers. For other justifications of temporal discounting in terms of relationships, see e.g. De-Shalit (1994: 85) or Lloyd (2021).

sections, I'll suggest that these considerations undermine the second premise of the above argument: they imply we generally ought to care *less* about far future people than about contemporary strangers. In section 5, I'll then suggest they also undermine the longtermist conclusion: they imply we shouldn't generally just do what's expectably best in the long term.

To make my arguments, I'll rely on two assumptions: a normative and an empirical premise. I personally think they're true. But they're controversial, and thus constitute limitations of my argument. My first, normative assumption is an

Objective list theory of welfare: What makes someone's life good for them is (at least in part) its containing instances of a pluralist list of objective goods: friendships, achievements, virtues and so on.

With this assumption I'm ruling out, in particular, simple hedonist and desire satisfactionist theories. I'm assuming that what makes your life good for you isn't always just that it's pleasant for you or satisfies some desire of you.¹⁴ Instead, the grounds of your welfare aren't reducible to any single welfare-good, and at least partly independent of your attitudes towards them.¹⁵

My second, empirical assumption, concerns the nature of far future people's lives. It's the

Assumption of unknown differences: If far future people exist, their lives—e.g., the kinds of relationships, projects, art, understanding or character traits they'll have—will be radically different from ours. We're quite clueless of how exactly their lives will look, and our influence on it is enormously indirect at best.

The key thought here is simple. If cultural and technological development continues even roughly on the current trajectory, life in a million years won't be remotely like ours. This is something that advocates of longtermism have typically emphasized. For instance, Ord (2020: 22) says: 'beauty, understanding, culture, consciousness, freedom, adventure, discovery, art—our descendants would be able to take these so much further, perhaps even discovering entirely new categories of value, completely unknown to us.' Similarly, MacAskill claims that 'we can't even imagine the heights that human accomplishment might reach'.¹⁶ And Bostrom (2008: 2) describes far future life-quality in a similar tone: 'Beyond dreams. Beyond imagination.' This seems plausible. A million years is a very long time. Perhaps our coffee conversations won't concern the weather, the neighbors or children, but

¹⁴ For helpful overviews of hedonism and desire satisfactionism, see e.g. Gregory (2015) and Heathwood (2015) respectively.

¹⁵ For some relevant discussions of objective list theories, see e.g. Finnis (1980: ch. 3-4), Parfit (1984: App. I) or Fletcher (2013). Note that I'm only presupposing that objective goods are *part* of what grounds your welfare. This doesn't rule out that your subjective attitudes to these goods matter too. For an overview of such 'hybrid views', see e.g. Woodard (2015).

¹⁶ See <https://www.youtube.com/watch?v=WypRKhvGVYk>, at 7:50 (last accessed Jan 11, 2022).

hyperreal diophantine approximations and ultraviolet matter in nano-quark holes. Perhaps we'll be receptive to magnetic fields, and enjoy 'schmusic'-performances, in which magnets of different sizes are moved around us in different patterns. Maybe we'll have found ways of 'merging' with others, as the ultimate form of friendship or sex. Perhaps concepts like 'benevolence', 'courage' or 'compassion' will have become vain, as we're automatically reasonable and happy, and the core virtue will be 'post-humaneness'. Perhaps we'll have evolved into cyborgs or AI entirely, and spend our lives relishing the delights of the 'schmetaverse'. We have no clue, and at best an extremely indirect influence on it. Or so I assume. Let's now see how these assumptions affect our concern for far future people.

3. Authenticity

Suppose Leah is a Zen teacher in Melbourne. She guides meditations, lectures about Zen, and organizes retreats. She got into it through her former partner. But one day she admits to herself that it's all been a fancy façade. She's never really experienced that clarity of mind in meditation, which she so often extols. She generally finds that sitting and breathing tedious. She doesn't really believe, or even think she truly understands, the mysterious Zen-philosophy. If anything, it's during her modern dance classes, with her dream of becoming a professional dancer, or with an aesthetical ethics à la Nietzsche, that she feels truly at home. Perhaps it was all a sophisticated way of impressing her former partner, or indeed herself.

Intuitively, there's something problematic about Leah's life. She doesn't *really* seem to value the things around which she orients her outward actions. Her life seems inauthentic. To substantiate this thought, here's a way to define (at least a core aspect of)

Authenticity: It's inauthentic for one to respond to something's value in one's action (at least in part) to the extent that one doesn't genuinely appreciate that value.

Let me explain. What is it for you to 'genuinely appreciate' something's value? To begin with, it's for you to actually care about that value, with core parts of your person, beyond your intentional action. It's for you to not just act in accordance with it, but to also have appropriate affective, conative or cognitive responses to the value. Thus you don't genuinely appreciate the value of neighborhood if you never have the faintest desire to live in proximity to others. If you can witness sexual violence without a trace of aversion, or even find pleasure in it, you don't really appreciate its wrongness. And perhaps you don't genuinely appreciate the value of non-human animals if you're certain they have no moral status. In this sense, we might say, authenticity is about staying true to *you*—about aligning your actions with the things that actually matter to you deep down.

But to genuinely appreciate something's value, it's not enough for you to *have* these responses to it. You must also have them for the right reasons. To the

extent that something is valuable in virtue of certain valuable-making properties, you must care for the value in virtue of caring about these properties. This means you must have access to these grounding properties—in virtue of your intellectual understanding, physical capacities, life-experience or whatever—and must have the relevant affective, conative or cognitive responses to these grounding properties too. Thus, if you’ve lost your taste functions to COVID, you can’t genuinely appreciate the deliciousness of that pumpkin soup. You can’t really appreciate the brilliance of Parfit’s *Reasons and Persons* if you have no understanding of its theses and arguments, or if you think they’re all profoundly misguided, or if they all leave you entirely cold. And perhaps if you’ve never loved anyone, you can’t genuinely appreciate the full value of love. In this sense, we might say, authenticity is also about staying true to a value—about your actions flowing from a genuine understanding of it, rather than a superficial or accidental response to it.¹⁷

According to the above account, insofar as you don’t genuinely appreciate something’s value, it’s inauthentic for you to respond to it in your action. So, if you respond to these values in such cases—if you compliment the chef on that pumpkin soup (barred of your taste-functions), do activism against sexual violence (against a background of indifference about it), or publicly praise *Reasons and Persons* (while it’s actually all Greek to you)—it will be inauthentic.¹⁸ Now the account doesn’t say that it’s necessarily (even pro tanto) *morally* problematic to act inauthentically. Perhaps even if you don’t genuinely appreciate the wrongness of sexual violence, say, there needn’t be anything morally bad about protesting against it. Still, intuitively, the fact that an action would be inauthentic for you provides a reason against it, or mitigates your reasons for it. Other things equal, if one of your actions would be more authentic than another, you have more reason to perform the former.

This seems to capture the inauthenticity of Leah’s life. Leah doesn’t have full access to the grounds of the value of Zen: she doesn’t thoroughly know that clarity of mind, doesn’t really understand its philosophy. She lacks many of the relevant affective or conative responses to that value: she doesn’t unreservedly enjoy the Zen-practice, and feels no real admiration for it all. To that extent, the way in which she nonetheless adheres to that value in practice—her teaching, lecturing or extolling—is inauthentic. *Ceteris paribus*, Leah thus seems to have more reason to do modern dance rather than Zen.

If all of this is right, this ideal of authenticity will affect the longtermist premises. In light of the objective list theory, in caring about far future welfare, you should care about far future objective goods. But in light of our assumption of unknown differences, you can’t genuinely appreciate these goods’ value. Consider schmusic, say: you’re not receptive to magnetic fields, and in witnessing a ‘piece’, you’d presumably respond with bewildered amusement instead of appropriate awe.

¹⁷ Some people might agree that a practical concern for a higher-order property, without appreciation of its grounds, is problematic—but doubt that it’s a problem of ‘authenticity’. Not much hinges on this label for my purposes. What matters is that there’s something problematic about such concern.

¹⁸ For related ideas on authenticity, or ‘integrity’, and for defenses of its importance, see e.g. Taylor (1992), Varga (2011), Paul (2014), Archer (2017) or Bauer (2017).

Similarly, consider the value of partner-merging or being ‘post-humane’: you have no access to why these things are good, and would presumably be more appalled or left indifferent than appropriately impassioned by them. You can’t really value those values. So if you take longtermist actions with the goal of promoting schmusic, merging and post-humaneness, there’s something inauthentic about it. It’s a bit like praising a soup you haven’t tasted, or like Leah’s promoting Zen.

Actually, in fact, the inauthenticity of longtermist actions must cut considerably deeper. After all, it’s not that we *know* the constituents of far future welfare—like schmusic, partner-merging or the schmetaverse—but can’t appreciate these things properly. We’re largely clueless about what will make far future lives good. So it would be unreasonable to take longtermist actions with the intention to promote specific goods in particular: the probability of promoting any actually specifiable thing is overly tiny. More reasonably, we may take such actions with the intention to promote far future welfare in the abstract, whatever it will consist in: we may try to contribute to future people *existing*, perhaps under favorable conditions, and hope that over time our descendants will have improved their form of life to whatever specific effect. In other words, our motivation for longtermist action can’t be a *de re* concern with the grounds of far future welfare. It must be a concern with such welfare as such, or *de dicto*. Compare this with contemporaries. Of course even in donating to Oxfam, say, you can’t know how exactly you’ll help. But by and large, you know what makes present people’s lives good: our familiar relationships, achievements, virtues and so on. So you’re dealing with known unknowns: you know that if you donate to Oxfam, someone will get something like *that*. And in virtue of caring *de re* about every item from our objective list, you can care about the uncertain prospect of benefiting someone in some such way. But that’s not so with far future people. Here you’re dealing with unknown unknowns: things you currently can’t even imagine. So, you can’t care about the uncertain prospect of benefitting far future people *in virtue* of caring about some list of objective goods. If you care about this prospect, it will be because you care about the sheer higher-order fact that far future people live well. You’ll try to do them good, without any idea about what this goodness might consist in.

Now, some people might think that to care about welfare *de dicto* is ‘fetishistic’ in a strong sense: that welfare in the abstract is an altogether inappropriate object of concern, and you therefore have *no* reason whatsoever to care about it.¹⁹ I think this is too strong. I think you have some reason to care about welfare as such.²⁰ However, if my account above is right, acting out of such higher-order concern is deeply inauthentic. It takes the problem of inauthenticity—of orienting your actions around a value whose grounds you fail to appreciate—to the extreme. It’s a bit as if you celebrated the value of Parfit’s *Reasons and Persons* without

¹⁹ For a related skeptical view regarding *de dicto* concern about rightness, see e.g. Smith (1994) or Zhang (2021). This view is often motivated by a ‘buck-passing account’ of rightness (see e.g. Dancy 2000; Stratton-Lake 2002: 15). The skeptical view about welfare might be motivated by a parallel account of welfare (see e.g. Sidgwick 1907: 112; Darwall 2002; Skorupski 2007; or Fletcher 2012).

²⁰ For a related permissive view regarding *de dicto* concern about moral rightness, see e.g. Lillehammer (1997), Copp (1997), Svavarsdóttir (1999), Carbonell (2013) or Johnson King (2020).

even knowing whether it's a book of philosophy, a painting, or a pretentiously misnamed orphanage. It's a bit as if Leah played her well-memorized Zen-role without the faintest hunch of what it could all be about, just following the testimony of a reliable friend who told her such a life was good.²¹

Again, promoting the welfare of contemporaries won't be nearly as inauthentic. True, we don't all share the exact same goals, desires and values. But we contemporaries are still pretty similar, and can largely appreciate each other's good. So, if the inauthenticity of an action mitigates your reasons for it, the second premise of the above argument is false: *ceteris paribus*, you have stronger reason to care about the welfare of our contemporaries than about that of far future people. To what extent this undermines the longtermist conclusion is a more complicated question. I'll get back to it in section 5. But first, let me introduce a second reason to care less about far future people.

4. Meaning

Suppose Egan donates his semen to a sperm bank in Buenos Aires. Djamila receives it, and raises the resulting child Imani alone. With enormous dedication, mindful decisions and loving affection, she teaches Imani gratefulness and environmentalist virtues, supports her passion for the accordion, helps her through crises in her friendships and studies. So Imani lives a very good life. At the age of fifty, she wants to know her biological father, determines the identity and whereabouts of Egan, and meets him. He is delighted to see her. He says he's always thought the meaning of life was to make the world better. And fathering such a happy person contributes so much to that. Indeed, he proclaims in enthusiasm, the value of Imani's life contributes just as much meaning to his life as to Djamila's.

Intuitively, that last claim is wrong. Imani's welfare contributes more meaning to Djamila's life than to Egan's. Let's see why that might be so. Intuitively, Egan is right insofar as the meaningfulness of your life isn't *just* due to how meaningful you subjectively find it. If you dedicate your life to counting grass or

²¹ Thus, I think the intuitive problem of fetishism (as suggested e.g. by Smith 1994 or Zhang 2021) is one of authenticity. I hope to explore this idea in more depth in the future. For related ideas, see e.g. Hopkins (2007), Hills (2009), Mogensen (2017), Johnson King (2019; 2020) or Heering (forthcoming). One might think there's an intermediate level. Perhaps we can't care authentically about the highest-order property of people being well off, or about fully concrete instances of schmusic, merging or post-humaneness. But, one might think, these things will be instances of familiar *categories* of welfare-goods—forms of art, relationships or virtues—and we can care authentically about these intermediate general categories. But this seems dubious. Suppose you find Arnold Schönberg's music unnerving, but zealously promote it—simply because you care about 'art' in general, or about Frida Kahlo's paintings and Sophocles' tragedies. That seems inauthentic. Plausibly, the value of 'art' in general (insofar as it has any value at all) is grounded in the value of specific artworks, rather than vice-versa. So, if you appreciate specific artworks, you'll only have a very partial appreciation of the grounds of the value of art. And you can't appropriately care about some artwork just in virtue of its being art. Thus there'll be something inauthentic if you promote schmusic, merging or post-humaneness *simply* in virtue of their being art, relationships or virtues.

killing foreigners it isn't meaningful, even if you find it so. Promoting objectively good things is an important part of life's meaning. But how exactly is it? One might adopt a simple consequentialist account, on which all actual good consequences of your actions contribute equally (or just in proportion to their objective goodness) to the meaning of your life.²² This, we might suppose, is Egan's view. But it's overly simplistic. Imani's life was a consequence of both Egan's and Djamila's actions, but provides more meaning for Djamila. Alternatively, one might adopt an ex-ante, expectationalist interpretation, on which it's the expected value of your actions at the time of acting that provides the meaning. One might suggest that Egan couldn't have expected to produce a life like Imani's, while Djamila could. But that too seems false. Perhaps Egan had ample data on children from that sperm bank, and could reasonably expect to produce a life like Imani's. Still, Imani seems to provide more meaning for Djamila. More promisingly perhaps, one might say Djamila made a greater causal contribution to Imani's life than Egan. She not only contributed genetic material, but was also responsible for the specific virtues, relationships and projects that formed her daughter's life. And this difference in degrees of causal contribution, one might say, is why Djamila derives more meaning. But this still seems too crude. Suppose Egan influenced Imani's life deeply, but in an entirely fortuitous way. He once hailed a Taxi, which otherwise would have run over Imani further down the road. He caused the pub quarrel in which the parents of Imani's future husband first met. He once fired an employee, who then started a business and gave Imani the job of her life. Even if Egan thus affected Imani's life deeply, Djamila derives more meaning from it.

Intuitively, the problem in this last scenario is that Egan's contributions were too accidental. To capture this, I propose the

Manifestation account of meaning: One's life is meaningful (at least in part) to the extent that it promotes valuable things, such that the value of those things, or its grounds, are manifestations of one as a person.

At least as a rough approximation, we might say that x is a manifestation of y to the extent that it's an actualization of y 's dispositions. Suppose you drop your friend's vase, it shatters, and while fetching the flinders from under the sofa your friend then finds a long-lost bracelet. The shattering of that vase manifests its fragility: it's an actualization of it. The discovery of that bracelet doesn't manifest the fragility of the vase: it's not an actualization of that disposition, but an accidental consequence of it. So, if something is a manifestation of you as a person, it's not merely an accidental causal effect that you've had. It's an actualization—and thus an expression, emanation or reflection—of your values, talents or character.

To illustrate this idea, take Imani's life. If an objective list theory is correct, Imani's life is good in virtue of her specific virtues, relationships or projects. The fact that Imani had those things is very much a manifestation of Djamila as a person: an actualization of Djamila's dedication and mindful attention, her concern for

²² For such a view, see e.g. Smuts (2016).

friendships, music, environmentalist virtues and so on. Djamila is like the sculptor shaping the clay of her daughter's life. Perhaps to some degree, the fact that Imani had a good life is a manifestation of Egan too: an actualization of his concern for producing happy people, say. But the grounds of Imani's welfare weren't emanations from his person to anything like the degree to which they were for Djamila. It wasn't due to him that Imani was so environmentally conscious, played the accordion with such dedication, or maintained her friendships instead of ending them. The extent to which Imani's life is a manifestation of Egan—to which he can deem himself responsible for her welfare in the relevant sense—is much smaller. He's like the patron *funding* the sculptor shaping the clay of her life. And that's true even if objectively, or in expected value terms, or even in degrees of causal influence, his contribution was the same.

The manifestation account thus says that Imani's existence contributes less meaning to Egan's life than to Djamila's. More generally, it says that being non-accidentally responsible for good things contributes to the meaning of life. As I understand the account, it also implies the reverse. Being non-accidentally responsible for bad things, in this form, detracts importantly from meaning. If Imani lived a miserable life, as a direct manifestation of her mother's carelessness or incompetence, that reduces the meaningfulness of Djamila's existence—and much more so than it diminishes the meaningfulness of Egan's life. Now plausibly, meaning in life is not all about such manifestation. It's also about subjectively enjoying one's promoting good, or about overcoming challenges or passively appreciating existing value.²³ So the manifestation account aims to capture only part of the grounds of meaning. Still, non-accidentally promoting the good seems an important contributor to meaning.²⁴ And intuitively, you have reason to live a meaningful life. So if all of this is right, you have reason to promote good (or avoid promoting bad) things, such that their values are manifestations of you as a person. In particular, other things equal, you have more reason to promote good (or avoid promoting bad) things in such ways than through more accidental chains.

²³ For the first idea, see most prominently Wolf (2010). For something like the third idea, see e.g. Lovett and Riedener (2022b). Perhaps in particular, it can add meaning to your life if you appreciate the value of your own non-accidental works. Intuitively, say, there can be something tragic if your project reaches success only after your death—as with Gregor Mendel, Emily Dickinson or Franz Kafka. This might add another problem of meaning for longtermists: they won't ever see the fruits of their labor.

²⁴ One might doubt this. In particular, one might think other factors better explain our intuitions about Egan and Djamila. One might say Djamila's raising of Imani seems more meaningful (in large part) because it involves more of a challenge or achievement, or takes up more time, or because she takes more subjective enjoyment from Imani's life. But I doubt that these factors add very much, if the underlying causal connection is sufficiently thin. Suppose Egan's sperm bank is hidden somewhere in the Argentinian wilderness and enormously challenging to reach; he thus spends a full 18 years to get to it; and once Imani is born, he witnesses every step she does (perhaps through her social media presence) and enjoys it just as much as Djamila. Such a life still seems considerably less meaningful than Djamila's—perhaps even somewhat pitiable. The manifestation account thus seems to capture an important part of our initial intuitions.

This should suffice to elucidate the idea for now.²⁵ If it's correct, it impacts longtermist actions. Suppose you invest in AI safety today, and through a fortunate chain of events thus contribute to some marvelous life in a million years. Then perhaps to some minimal degree, the value of that future life manifests your forethought. You invested in AI safety precisely to allow far future people to exist. And perhaps it wasn't 100% fortuitous that you contributed to it. Still, *very* largely, the grounds of the value of this far future life are accidental from your perspective. Again, that's part of our assumptions. And as with authenticity, the distance here cuts deep. For a start, any details about why or how this life will be good can't possibly manifest you as a person. Insofar as the constituents of far future welfare are unknown unknowns from your perspective, you can't have intended to promote them in particular. So these details can't emanate from your specific concern. Moreover, and more simply, there's just a gargantuan array of intermediate causal factors, which you can't control or foresee, and which swamp your contribution almost completely. You're a tiny cog in the astronomical machine of the millennia grinding the effects of your actions. And that's true even if, in expected or objective value terms, your actions matter a lot. Your attempt to derive equal meaning from a far future individual would be a *bit* like Egan's attempt to derive meaning from Imani. More aptly perhaps, it would be like saying about one of our Neanderthal ancestors that our present-day benefits from enjoying classical music or displaying environmentalist virtues are just as significant for the meaningfulness of their lives as were their children's benefits from the parental warmth they gave them.

Again, that's quite different with contemporaries. True, even if you save someone through an Oxfam donation, say, many aspects about it are accidental from your perspective. But you know relatively well about the possible effects of your action, and can promote them comparatively directly. The degree to which you can understand this contemporary's life as a manifestation of your concern is much greater than with far future people. Conversely, if you do nothing and they die, the degree to which you should understand this as a manifestation of you is much greater than with the death or non-existence of far future individuals. So, if the meaningfulness of an action provides a reason for it, you again have stronger reasons to promote the welfare of contemporaries than that of far future people.²⁶

²⁵ For a related view of meaning, see Brogaard and Smith (2005). In a deeper exploration of the account, more would need to be said especially about what it is for you to be more or less manifest in something. For such details, and some more general motivation, see Lovett and Riedener (2022a; 2022b; 2022c). One point may be worth making here. As indicated, manifestation isn't simply a probabilistic notion. Even if Egan could have predicted with certainty how Djamila will raise Imani, Imani's life would be more a manifestation of her than of him. Still, probabilities are obviously relevant. If some state of affairs resulted from your actions, it's generally more a manifestation of you if you could have been certain that it will result than if it was highly unlikely to.

²⁶ These negative examples in particular suggest another possible argument against longtermism. One might take manifestation-relationships to underpin *moral responsibility*. For instance, one might say you're morally responsible for good or bad things to the extent that the grounds of their goodness or badness are manifestations of you as a person. So one might argue that we're generally less morally responsible for bad things in the present than for such things in the very far future, and

5. Objections and Clarifications

Let's take stock. I've introduced two reasons to care less about far future people than even about contemporaries you have no special positive relation to. Treating their welfare in just the same way would be problematically inauthentic and meaning-sapping. This suggests that, *ceteris paribus*, we should give more weight to the interests of our contemporaries than to those of far future people. My two assumptions from section 2 were relevant for both points. Suppose instead of an objective list theory, hedonism was correct. Then we could care authentically about far future pleasure. And since our contributions to far future pleasure might not be quite as contingent as our contributions specifically to schmusic, the schmetaverse or post-humaneness, we might derive a little more meaning from promoting *that*. Similarly, suppose we could reasonably assume that future lives will be similar to ours, and could have a more direct influence on them. Well, then we could authentically care about future friendships and virtues and so on. And our promoting them wouldn't be as accidental. So suppose my assumptions are true and my arguments sound.²⁷ What exactly does that mean for the longtermist conclusion? One might worry it doesn't mean much. So let me now clarify the nature, scope and limitations of my arguments by answering three salient worries.

First, one might doubt that our two considerations should have any normative weight at all, or that I've even shown we should give *any* priority to our contemporaries. One might think concerns about authenticity and meaningfulness are simply self-indulgent. Ideally, you should care about others: you should try to do them as much good as you can, regardless of how you personally appreciate the

that we thus have stronger reasons to prevent the former. This line of reasoning is perhaps more controversial, but I think it's promising, and I intend to explore it in future work.

²⁷ There's a worry that comes from questioning my second assumption. One might say it isn't *certain* that far future lives will be radically different. And suppose there's even just a one in ten thousand chance that, conditional on our survival, our descendants will live largely as we do. Then longtermist actions might expectably still affect 100 billion (10^{11}) people like us. And that, one might say, still trumps almost everything. So doesn't the longtermist argument hold if we focus on these conservative scenarios only? In reply, first, this would constitute an interesting shift in focus. As indicated, longtermists have typically emphasized how different the far future could be. If we rely on conservative scenarios alone, this is an important dialectical reorientation. Second, it's not clear that the argument will actually work—i.e., that the relevant probabilities will be high enough—if we just rely on these cases. Even relatively small changes in our way of life, our capacities or values, can impact the nature of our welfare deeply. And our status quo is marked, on the one hand, by countless imperfections and problems (unlike perhaps the state of gods) and on the other hand, by ample knowledge and technology to address them (unlike perhaps the state of non-human animals or our distant ancestors). It just seems *very* unlikely that our form of life will stay largely intact for hundreds of thousands of years. Third, this worry anyway doesn't seem to affect the second point about meaning. Even if we do survive billions of years largely unchanged, that would still be almost entirely due to forces outside of our present actions.

way they live, or how your efforts contribute to *your* life. You should not—not even *ceteris paribus*—give any weight to such egocentric concerns.²⁸

Let me say two things in reply. First, even granting that our considerations are largely self-centered, at a minimum, they seem to ground prerogatives. They set limits to what morality can demand of us. In particular, among other things, there's something *alienating* about actions that lack such authenticity and meaning. If you act in response to a value you fail to appreciate, there's a distance between your outward action and your deeper concerns. It's like a lack of 'integrity' à la Williams (1973; 1981b): a form of ignoring what actually moves you, discounting your attachments, dissolving your own person under the dictates of the impersonal good. Similarly, if you promote some good in highly accidental manners, there's a distance between you and that good. It's reminiscent of '*Entfremdung*' à la Marx: a way of doing things without intrinsic or immediate meaning, of reducing your own agency to a cog in a vast causal machine, without relevant responsibility or connection to the ultimate product.²⁹ Intuitively, morality can't demand of us to ignore these concerns entirely. So at least sometimes, it seems permissible to give *some* priority to our contemporaries.

Second and less concessively, however, it just needn't be self-centered to live in accordance with our two considerations. More precisely, it needn't be a concern for yourself which makes you act in authentic and meaningful ways. It can simply be a concern for the values you care about. And for this reason (if not for others too³⁰) it needn't be a permissible defect, but can be a realization of the ethical ideal. To see what I mean, suppose Kaya is a rising politician from Eritrea. Having experienced how a lack of education can hamper, she committed herself to improving the country's schools. And she's got an exceptional talent for it: courage, dedication and charisma, and plenty of creative and effective ideas. Now she thinks she could do *slightly* more good by going into consulting and donating ample money to improve conditions on Chinese factory farms. (Let's say other salient options don't occur to her for now.) But she has no relationship whatsoever to China, let alone to its animals, and feels rather appalled by the world of consulting. Kaya may then decide for a career as a politician, not out of concern for *herself*, but simply out of commitment to her fellow Eritreans, or out of concern for the value of literacy or female education. This will orient her life around values she appreciates, and make it more manifest in the good she effects. And intuitively (at least if the

²⁸ The worry here is similar to the worry that nonconsequentialist morality is 'self-indulgent'. For discussions of this worry, see e.g. Williams (1981a) or Blustein (1991).

²⁹ At one point, Marx describes the opposite of non-alienated production in words strikingly resembling the manifestation account: 'Let us suppose that we had carried out production as human beings. [...] In my *production* I would have objectified my *individuality, its specific character*, and therefore enjoyed not only an individual *manifestation of my life* during the activity, but also when looking at the object I would have the individual pleasure of knowing my personality to be *objective*.' (Marx 1975: 227; emphasis in original).

³⁰ A simpler reason for why it can be admirable to live in accordance with our considerations might be that some amount of self-centeredness is just part of the well-lived life. For instance, perhaps it can manifest a dubious lack of self-respect to ignore our considerations entirely.

difference in promoted good isn't too vast), such a choice may well be admirable: it may seem disloyal towards her co-citizens, or noncommittal about her own values, if she applied at Boston Consulting. Thus, considerations of authenticity and meaning don't just ground prerogatives for permissible but suboptimal egotism. They're not categorically less worthy than concerns of impartial beneficence. They're a genuine part of the well-lived life.

But there's an obvious subsequent worry. Perhaps it can be permissible, or even admirable, to accord with our principles when you could otherwise do *slightly* more good. But the longtermist rationale suggests that, in many situations, you could do *vastly* more good by taking some longtermist action—save trillions more people in expectation. And surely, one might say, considerations of authenticity and meaning can't have enough weight to outbalance *that*. More concretely, perhaps you should give the interests of contemporaries n times more weight than those of far future people, for some n greater than 1. But for any reasonable such factor, the longtermist argument will still go through. Even if we count far future people a hundred thousand times less, say, in light of their astronomical number they should still dominate our concern. In sum, our considerations can't really threaten the longtermist conclusion—or so one might object.

In response, it's worth noting first that such a factor-interpretation doesn't capture our initial intuitions. For any such factor, the longtermist argument stands, provided our actions affect sufficiently many future people. So on this reading, the problem with the argument in section 1 was, at most, that the septillion (10^{24}) of future people I imagined were still too few. But that surely wasn't the intuitive problem. The argument wouldn't have been any more common sensical if I had spoken of an octillion (10^{27}) or a decillion (10^{33}) to start with. Intuitively, something stronger is true. *No matter* how many far future people our actions expectably affect, our contemporaries aren't generally practically irrelevant. In other words, at least sometimes, a certain form of non-aggregationism or lexical priority applies.³¹

That's our initial intuitions. But is it plausible that we should give lexical priority to our contemporaries, in light of the above concerns? I very much doubt such priority is always warranted, or that far future people are generally practically irrelevant. I've suggested that the well-lived life isn't simply or primarily about impartial beneficence. But surely it isn't simply or primarily about authenticity or meaning either. Plausibly, the relative weight of different considerations must ultimately be determined in the context of entire lives. Currently, most agents *never* take actions with explicitly longtermist goals. And it's not plausible that our lives overall become problematically inauthentic and meaningless if we take some—or indeed quite considerable—longtermist action. I thus think our considerations of authenticity and meaning are absolutely compatible with the plausible view that we

³¹ For related kinds of non-aggregationism, see e.g. Scanlon (1998: 235) or Voorhoeve (2014).

should spend *much* more resources on the long-term future than we currently do, and partly in light of the sheer number of possible people.³²

However, I do think they cast doubt on the more radical idea that our contemporaries are *generally* practically irrelevant, or relevant only as tie-breakers. Suppose we literally always make our choices depend on longtermist considerations, and follow shorter-term goals only when these considerations are exactly tied. *Then* our lives would be devoid of authenticity and meaning to an extent that seems problematic. Consider Kaya again, and suppose she realizes that various longtermist actions are expectably *much* better than improving either education in Eritrea or animal welfare in China. So she tries to follow the longtermist rationale as well as she can: in deciding which job to take, what to do with her money and spare time, whether to prioritize the climate crisis in her action or present-day racism or education in Eritrea, which politicians to elect or policies to support, whether or not to engage in comparatively minor lies or acts of defraud or betrayals of trust, and so on. She often does things with little short-term benefits at all. And when her actions do yield short-term goods, she regards them as merely instrumentally valuable, or as insignificant tie-breakers. *Then*, it seems, the grand orientation of Kaya's life would be worryingly alienated from anything whose goodness or badness she can genuinely value or have any direct influence on. At a minimum, that's not a life that can be demanded of her. More strongly, it's not the life she has most reason to live. And that's true regardless of how many far future people are at stake. At some point, in some contexts, a lexical priority seems right.³³

When exactly that's so—between total neglect and total prioritization of longtermist concerns—is naturally difficult to pinpoint. Perhaps we're permitted, and often have most reason, to live our lives *by and large* in response to goods we

³² As indicated in footnote 3, by 'longtermism', some people mean the view that the long-term future should be *a* key moral priority of our time. Depending on what it means to be 'a key priority', I think my arguments are thus compatible with this view.

³³ One might counter that longtermism wouldn't be as radically revisionist and alienating in practice. One might say we're almost totally clueless about the long-term effects of our actions. This means that in most situations, all of our options are expectably equally good in the longer term, or that the expected longer-term stakes (or value-differences among options) are generally small. So, one might argue, in most situations we can use near-term effects as tie-breakers, live largely as we live now, and thus avoid any serious alienation. In response, first, this isn't the view that longtermists typically take. They typically emphasize that their view is highly revisionist in practice (see e.g. MacAskill 2022: esp. ch. 10; Greaves and MacAskill 2021). Second, that longtermism is highly revisionist just seems very plausible, once you accept that *some* actions are expectably exceptionally good in the longer term. Suppose AI safety investments are generally exceptionally good. Well, you always have the option of trying to earn as much money as you can, and donate it for AI safety. And thus the long-term stakes almost always seem very high. Third, plausibly, even if it implied that our outward lives could remain largely as they are, longtermism would still have highly revisionist implications about the *attitudes* we should have towards our lives. It would imply we should see all of our standard concerns as insignificant tie-breakers—as *almost* totally and absurdly trivial in comparison to what really matters (see e.g. Nagel 1971). Plausibly, there'd still be something alienating about this. We cannot authentically see our own actions as insignificant in this sense. And doing so seems to threaten our sense of meaning. I'd love to explore these implications of longtermism for our attitudes on another occasion.

can appreciate relatively well and affect relatively directly. Or conversely, we're permitted, and often have most reason, to live our lives *by and large* without causing or allowing bad things we can appreciate rather well and affect rather directly. In any case, I think these considerations of authenticity and meaning shape our normative landscape importantly. For all the longtermist action we should plausibly undertake, we often have most overall reason to be guided by nearer-term concerns. One point may need emphasis at this juncture. My argument isn't simply a demandingness worry against strong altruistic obligations. It's compatible with very wide-ranging duties of beneficence. Specifically, suppose we could somehow rather directly save quadrillions of contemporaries. Plausibly, this would put extremely wide-ranging demands on us: it would outweigh almost any competing moral or non-moral concerns. My argument is perfectly compatible with this. It only implies that such demands aren't equally (or nearly) as weighty if these 'people' are profoundly different and distant.

Still, there's a third worry. *My* conclusions might just seem deeply counter-intuitive. I've claimed that a life oriented too much around longtermist concerns must be inauthentic and meaning-deprived. But this, one might say, just seems false. Quite to the contrary, it seems that *few* things could be of more genuine concern to us than our long-term trajectory. Few things could be as meaningful as trying to safeguard our long-term future. Perhaps Kaya currently cares a lot about the impact of menstruation on girl's absenteeism in the schools of Asmara, and derives meaning from addressing that. And perhaps we all focus on such parochial issues now. But couldn't we *possibly* come to care just as—or indeed much more—authentically and meaningfully about the long-term future of life in the universe? Taking a step back from my arguments, it may seem absurd to deny this.³⁴

I think there's something to this intuition. But it doesn't vindicate the longtermist conclusion, or undermine my points. I've addressed one argument for longtermism only: the argument that grounds the importance of longtermist actions on the sheer number of far future people. And here I stick to my claims. We just cannot care as authentically about far future individuals, or derive as much meaning from doing them good, as with contemporaries. If we think we can, we must be underestimating their difference or distance from us, or employing some incorrect theories of authenticity and meaning. Now importantly, there could be *other* arguments for caring about the long-term future, which don't hinge on total future welfare. To take just one example, consider the long-term-flourishing of humanity as a species. Perhaps the fulfilment of our species-potential has a value in itself, which doesn't even partly reduce to the value of our total individual welfare.³⁵ And so perhaps our long-term species-flourishing can be a goal whose value we can genuinely appreciate, and in whose realization we can be manifest rather distinctly. All of this is compatible with my argument. However, it seems implausible that such alternative rationales can make longtermist actions nearly as important as the standard argument *prima facie* did. Our long-term species-flourishing may be good.

³⁴ I thank Jacob Barrett for helping me see this.

³⁵ For similar thoughts, see e.g. Rolston (1985), Frick (2017) or Ord (2020: 52-53).

But it doesn't seem *so* valuable as to dominate all nearer-term concerns. Why should it be? What made the long-term future seem *so* important, on the standard argument, was simply that it contained so many people. To the extent that our concern for the far future becomes independent of the number of future individuals, it becomes less overwhelmingly significant.

6. Conclusion

Let me conclude. I've focused on one assumption of the standard longtermist argument: the idea that we should care as much about far future people as about our contemporaries. I've argued against this premise: considerations of authenticity and meaning suggest we generally ought to care less about them. And I've claimed that this means, at least, that the far future shouldn't be as dominating a concern in practice as the longtermist argument initially suggests. We should plausibly take the far future much more seriously than we currently do. But at the same time, in many situations, even if longtermist actions would expectably do more good, we ought to act for contemporaries' sakes. Consider the United States again. Plausibly, the US should ramp up investments in AI safety. At the same time, it seems right for them to make it one of their priorities now to help the Ukraine, simply because of the Ukrainians, and even if it isn't optimistic in the very long term.

Acknowledgments. For very helpful inputs or comments on this material, I thank Sarah Arrenbrecht, Jacob Barrett, Samuel Hughes, Harvey Lederman, Harry Lloyd, Adam Lovett, Andreas Mogensen, Lukas Naegeli, Peter Schaber, Christian Tarsney and especially Hilary Greaves. I also thank the Global Priorities Institute for hosting me and providing a very inspiring environment while I was working on this paper.

Bibliography

- Archer, A. (2017), 'Integrity and the Value of an Integrated Self', in *Journal of Value Inquiry* 51/3: 435-454.
- Bauer, K. (2017), 'To be or Not to be Authentic. In Defence of Authenticity as an Ethical Ideal', in *Ethical Theory and Moral Practice* 20/3: 567-580.
- Blustein, J. (1991), 'Integrity and Self-Indulgence', in J. Blustein, *Care and Commitment: Taking the Personal Point of View* (Oxford University Press), 82-89.
- Bostrom, N. (2008), 'Letter From Utopia', in *Law and Ethics of Human Rights* 2/1: 1-7.
- Brogaard, B., and Smith, B. (2005), 'On Luck, Responsibility and the Meaning of Life', in *Philosophical Papers* 34/3: 443-458.
- Broome, J. (1991), *Weighing Goods* (Blackwell).

- Carbonell, V. (2013), 'De Dicto Desires and Morality as Fetish', in *Philosophical Studies* 163/2:459-477.
- Copp, D. (1997), 'Belief, Reason, and Motivation: Michael Smith's "The Moral Problem"', in *Ethics* 108/1: 33-54.
- Dancy, J. (2000), 'Should We Pass the Buck?', in *Royal Institute of Philosophy Supplement* 47: 159-173.
- Darwall, S. (2002), *Welfare and Rational Care* (Princeton University Press).
- De-Shalit, A. (1994), *Why Posterity Matters: Environmental Policies and Future Generations* (Routledge).
- Finnis, J. (1980), *Natural Law and Natural Rights* (Oxford University Press).
- Fletcher, G. (2012), 'Resisting Buck-Passing Accounts of Prudential Value', in *Philosophical Studies* 157/1: 77-91.
- Fletcher, G. (2013), 'A Fresh Start for the Objective-List Theory of Well-Being', in *Utilitas* 25/2: 206-220.
- Frick, J. (2017), 'On the Survival of Humanity', in *Canadian Journal of Philosophy* 47/2-3: 344-367.
- Greaves, H. (2017), 'Discounting for Public Policy: A Survey' in *Economics and Philosophy* 33/3: 391-439.
- Greaves, H., and MacAskill, W. (2021), 'The Case for Strong Longtermism', GPI Working Paper No. 5-2021.
- Greaves, H., MacAskill, W., and Thornley, E. (2021), 'The Moral Case for Longterm Thinking', in N. Cargill and T. M. John (eds.), *The Long View: Essays on Policy, Philanthropy, and the Long-Term Future* (FIRST), 19-28.
- Gregory, A. (2015), 'Hedonism', in G. Fletcher (ed.), *The Routledge Handbook of Philosophy of Wellbeing* (Routledge), 113-123.
- Heathwood, C. (2015), 'Desire-Fulfillment Theory', in G. Fletcher (ed.), *The Routledge Handbook of Philosophy of Well-Being* (Routledge), 135-147.
- Heering, D. (forthcoming), 'Why and When is Pure Moral Motivation Defective', in *Erkenntnis*.
- Hills, A. (2009), 'Moral Testimony and Moral Epistemology', in *Ethics* 120/1: 94-127.
- Hopkins, R. (2007), 'What Is Wrong With Moral Testimony?', in *Philosophy and Phenomenological Research* 74/3: 611-634.
- Jacobs, A. (2018), *Thanks a Thousand* (TED Books).
- Johnson King, Z. (2019), 'We Can Have Our Buck and Pass It, Too' in *Oxford Studies in Metaethics* 14: 167-188.
- Johnson King, Z. (2020), 'Praiseworthy Motivations', in *Noûs* 54/2: 408-430.
- Lillehammer, H. (1997), 'Smith on Moral Fetishism', in *Analysis* 57/3: 187-195.
- Lloyd, H. (2021), 'Time Discounting, Consistency, and Special Obligations: A Defence of Robust Temporalism', GPI Working Paper No. 11-2021.
- Lovett, A., and Riedener, S. (2022a), 'Touching the Good', manuscript.
- Lovett, A., and Riedener, S. (2022b), 'The Good Life as the Life in Touch With the Good', manuscript.
- Lovett, A., and Riedener, S. (2022c), 'Commonsense Morality and Contact with Value', manuscript.

- MacAskill, W. (2022), *What We Owe the Future* (Basic Books).
- Marx, K., and Engels, F. (1975), *Collected Works, Vol. 3, 1843-44* (Intl Pub Co).
- Mogensen, A. L. (2017), 'Moral Testimony Pessimism and the Uncertain Value of Authenticity', in *Philosophy and Phenomenological Research* 95/2: 261-284.
- Mogensen, A. L. (2022), 'The Only Ethical Argument for Positive δ ? Partiality and Pure Time Preference', in *Philosophical Studies* 179/9: 2731-2750.
- Nagel, T. (1971), 'The Absurd', in *Journal of Philosophy* 68/20: 716-727.
- Ord, T. (2020), *The Precipice: Existential Risk and the Future of Humanity* (Bloomsbury).
- Parfit, D. (1984), *Reasons and Persons* (Oxford University Press).
- Paul, L. A. (2014), *Transformative Experience* (Oxford University Press).
- Pogge, T. (2002), *World Poverty and Human Rights* (Polity Press).
- Ramsey, F. P. (1928), 'A Mathematical Theory of Saving', in *Economic Journal* 38/3: 543-555.
- Rawls, J. (1999), *A Theory of Justice* (Belknap Press).
- Rolston, H. (1985), 'Duties to Endangered Species', in *BioScience* 35/11: 718-726.
- Scanlon, T. M. (1998), *What We Owe to Each Other* (Harvard University Press).
- Sidgwick, H. (1907), *The Methods of Ethics* (Macmillan).
- Smith, M. (1994), *The Moral Problem* (Blackwell).
- Smuts, A. (2016), *Welfare, Meaning, and Worth* (Routledge).
- Skorupski, J. (2007), 'Buck-Passing About Goodness', in T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson (eds.), *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*.
- Stratton-Lake, P. (2002), 'Introduction', in P. Stratton-Lake (ed.), *Ethical Intuitionism: Re-Evaluations* (Oxford University Press), 1-28.
- Svavarsdóttir, S. (1999), 'Moral Cognitivism and Motivation' in *The Philosophical Review* 108/2: 161-219.
- Taylor, C. (1992), *The Ethics of Authenticity* (Harvard University Press).
- Varga, S. (2011), *Authenticity as an Ethical Ideal* (Routledge).
- Voorhoeve, A. (2014), 'How Should We Aggregate Competing Claims?', in *Ethics* 125/1: 64-87.
- Williams, B. (1973), 'A Critique of Utilitarianism', in J.J.C. Smart and B. Williams, *Utilitarianism: For and Against* (Cambridge University Press), 77-150.
- Williams, B. (1981a), 'Utilitarianism and Moral Self-Indulgence', in B. Williams, *Moral Luck* (Cambridge University Press), 40-53.
- Williams, B. (1981b), 'Persons, Character and Morality', in B. Williams, *Moral Luck* (Cambridge University Press), 1-19.
- Wolf, S. (2010), *Meaning in Life and Why It Matters* (Princeton University Press).
- Woodard, C. (2015), 'Hybrid Theories', in G. Fletcher (ed.), *The Routledge Handbook of Philosophy of Well-Being* (Routledge), 161-174.
- Zhang, X. (2021), 'Why De Dicto Desires Are Fetishistic', in *Ratio* 34/4: 303-311.