In what way do future people count?

[draft. under review.]

William MacAskill tells us that future people count, but in what way do they count? I say that they count in just the same way that present people count and, as a result, that there's no route from their counting to any surprising claims about existential risk—contrary to how MacAskill would have it.

1. In *What We Owe The Future*, William MacAskill tells us that future people *count*. He says:

> The idea that future people count is common sense. Suppose that I drop a glass bottle while hiking. If I don't clean it up, a child might cut herself on the shards. Does it matter when the child will cut herself—a week, or a decade, or a century from now?

MacAskill says that it doesn't matter, and I agree. MacAskill concludes—as he's surely right to do—that future people therefore count.[1] But in what way do they count? MacAskill thinks they count in that

> (*) we have reasons to see to it that future people exist,

and, in virtue of them counting in that way, MacAskill says we should be dramatically concerned about the risk of human extinction, even at the expense of present-day concerns such as global poverty.

I disagree. I say that future people count in the same way that present people count, that their counting is accounted for by the same principles that account for how present people count, and, lastly, that there is therefore no route from their counting to any surprising claims about existential risk—certainly not one that is "common sense."

---

[1] The argument from this vignette features in the first pages of *What We Owe The Future*, but it's also MacAskill's opening gambit in various places: in these lectures https://globalprioritiesinstitute.org/will-macaskill-effective-altruism-a-better-way-to-lead-an-ethical-life/, https://forum.effectivealtruism.org/posts/AoHgbYvTjHnQw8kWX/what-we-owe-the-future-will-macaskill; in these podcasts https://80000hours.org/podcast/episodes/will-macaskill-what-we-owe-the-future/, https://www.econtalk.org/will-macaskill-on-longtermism-and-what-we-owe-the-future/; and so on. It also shows up throughout longtermism's online literature—e.g. https://longtermism.com/introduction, https://www.effectivealtruism.org/articles/longtermism.

2. I start by sharpening MacAskill's vignette. Suppose that the hiking trail is regulated in the following way: only one hiker is permitted on the trail at a time and when that hiker exits the trail, he or she spins a roulette wheel in order to select the next hiker—where each number on the roulette wheel corresponds to a number previously assigned to each of those waiting. (Why might the trail be regulated in this way? Perhaps its fragility is matched only by its popularity. Why do I stipulate these regulations? That will become clear presently.)

Now, suppose that it's Annie who shatters a glass bottle on the trail. (It will make things simpler to further suppose, for now, that the bottle contains some deadly toxin such that whoever steps on the glass won't just be wounded, but will certainly die.)

Why should Annie clean up the broken bottle? This is not a difficult question: Annie should clean up the bottle because, if she doesn't, there's a very good chance that someone will step on the glass and die. In fact, something stronger is true: it's not merely that, if Annie doesn't clean up the bottle, there's a very good chance that someone will step on the glass and die, but instead that, if she doesn't clean up the bottle, there's a very good chance that someone will step on the glass and die *and* Annie will have killed that person. (If I leave a bear trap on a path, then I those break whoever's leg is broken by the trap; if I leave a landmine on the path, then I maim whoever is maimed by the mine; and if I leave deadly toxins on a path, then I kill whomever is killed by stepping on those toxins.)

So why should Annie clean up the broken bottle? Because, if she doesn't, there's a very good chance that she will kill someone.

It's worth pausing over the meaning of "someone" in the previous sentence (and it is here that the regulations I stipulated above kick in). When I say that, if Annie doesn't clean up the bottle, then there's a good chance that Annie will kill *someone*, I do not mean that there is some particular person and there is a good chance that Annie will kill *him*. If Annie does clean up the bottle, it is not as if she will be able to point to a particular waiting hiker and truly say "if I hadn't cleaned up the bottle, then I would have killed you." There is no such person. After all, the person that she would have killed would have been fixed by the spin of the roulette wheel and there is no fact as to how that spin would have gone. Instead, all we can say is that if Annie doesn't clean up the bottle, then there's a good

chance that Annie will kill someone—while it remains *indeterminate* exactly who. (See Hare 2015 and REDACTED.)

Annie should clean up the broken bottle because, if she doesn't, there's a good chance that she will kill someone. And that fact—and surely it is a fact—follows from a general principle:

> (T1) if there's a good chance that person A will kill someone if she doesn't ϕ, then, other things equal, A should ϕ.[2]

This is simple stuff. Why would it have been thought complicated; and, in particular, why would a greater time span be thought complicate it?

Let's imagine a greater time span. Annie drops another glass bottle of deadly toxin on a second trail—so very clumsy!—but this trail won't be walked on for another 200 years. So long as the toxin in the bottle is as long-lasting as it is deadly, then it's plainly still the case that Annie should clean up the broken bottle. Why is that? I see no reason to think that our treatment of this second broken bottle should be any different to our treatment of the first—*viz*. Annie should clean it up because, if she doesn't, there's a good chance that she will kill someone. The greater time span is irrelevant and (T1) is all we need. Why would anyone have thought otherwise?

Perhaps they were cofounded by how Annie will be long-dead before anyone steps on her bottle and, in turn, by a puzzle that this gives rise to. Suppose it's 2222 and Ben has just stepped on the bottle and died. I have claimed that Annie thus killed Ben, but when did she do it? There is no obvious answer to this question: if she killed Ben in 2222 (when, say, the toxin stopped his heart), then she killed him when she was long-dead; but if she killed Ben in 2022 (when she dropped and failed to clean up the bottle), then she killed him long before he died—before he was born, even! (See Thomson 1971.)

Now, this is a lovely puzzle with a variety of possible solutions, but one of those solutions certainly isn't to conclude that Annie couldn't have killed Ben, after all. That a killer might die before her victim is plain and examples abound. Here's one: I lock the doors and set the house on fire, but succumb to the flames before the rest of the inhabitants. My early death has no bearing upon the obvious fact that I killed whosoever else died in the fire; just as Annie's early death has no bearing on

[2] Some readers might prefer to talk in terms of *causing to die* instead of *killing*. Elsewhere, I say that that's mistaken, but here it doesn't make any difference—so readers can proceed on that basis should they want.

the obvious fact that she kills whoever steps on her bottle. (I add, also, that a corresponding puzzle arises for geography. Agatha poisons Bert in London, who then travels to Patagonia, before succumbing to the poison. Agatha killed Bert, but where did she kill him? If she killed him in London (where she administered the poison), then she killed him somewhere he was never dead; but if she killed him in Patagonia (where the poison stopped his heart), then she killed him somewhere she never was. Another lovely puzzle! But it would be nonsense to respond to it by concluding that Agatha can't thus have killed Bert, after all.)

Or perhaps they were confounded by how Annie's potential victims don't yet exist (since no one alive today will be around to walk on the path in 200 years time). But why should that matter? If I bury a bomb in the foundations of an under-construction maternity ward, I plainly kill all those babies it obliterates upon exploding in, say, 2025.

Or perhaps they were confounded by the indeterminacy of how, if Annie does clean up the bottle, there is no specific person that she would have killed had she not. It's here that those regulations I stipulated earlier pay-off since we've already seen that that indeterminacy has no bearing on the fact that Annie kills the person who steps on the bottle.

Or perhaps they were confounded by some other thought. In any case, they shouldn't have been: what explains why Annie should clean up her broken bottle—regardless of whether there's a good chance it will be walked on tomorrow or in the distant future—is that, if she doesn't, there's a good chance that she will kill someone.

I earlier insisted that Annie's bottle contained a deadly toxin (and, in turn, that it would kill whomever stood on it), but I needn't have done so. That's because just as (T1) is true, so too is, for example:

> (T2) if there's a good chance that person A will dismember someone if she ϕs,
> then, other things equal, A shouldn't ϕ.

And that principle would explain why, for example, Annie shouldn't lay a landmine on the trail—regardless of whether it is likely to be triggered tomorrow or next century. And just as (T1) and (T2) are true, so too is (T3):

(T3) if there's a good chance that person A will give someone cancer if she ϕs, then, other things equal, A shouldn't ϕ.

And that principle would explain why Annie shouldn't bury dangerous quantities of radioactive waste on the trail—again regardless of when it is likely to be walked upon. And lastly:

(T4) if there's a good chance that person A will hurt someone if she doesn't ϕ, then, other things equal, A should ϕ.

And that principle explains why Annie—or MacAskill, or whoever—should clean up the broken bottle that they drop on a path. (In (T4), I use 'hurt' advisedly. The concept of *harm* has been tortured by philosophers into a term d'art, one which is best avoided in my opinion.)

I call these principles *T-principles* because they are timeless (T for 'timeless'), applying equally across time.

3. The aforementioned T-principles each concern us doing things to others that would be bad for them —killing them, maiming them, etc. Other T-principles concern us doing things to others that would be good for them. For example:

if there's a good chance that person A will save someone's life if she ϕs, then, other things equal, A should ϕ.

If Annie is hiking and there's a good chance that the next hiker to come along will be dying of thirst, and if Annie has water to spare, then Annie should cache some water for that next hiker. The preceding principle explains why—and it does so regardless of when that next hiker might come along. It's just one example, but the structure for creating others should be plain and I leave that to the reader.

The aforementioned T-principles also each concern determinate actions (killing, dismembering, etc.). Alongside each of them is its risk focused counterpart that governs those actions that merely increase the risk that bad (/good) things will happen to others. We shouldn't, for example, flood someone's house and alongside that principle is the following counterpart:

if there's a good chance that person A will increase the risk that someone's house will flood if she ϕs, then, other things equal, A shouldn't ϕ.

And that holds even when there's no risk that A herself will flood anyone's house. Our carbon emissions, for one, would fall under this principle's scope: the ten tons of CO_2 that I emitted flying my jet across the Atlantic increased the risk that, e.g., someone's house would flood, even though there was never any risk that I would flood someone's house. The preceding principle thus returns that, other things equal, I shouldn't fly my plane around—which is surely right—and it does so regardless of when the flooding is (more) likely to occur. That's just one more example, and I again leave it to reader to construct others.

These T-principles govern our behaviour as it concerns future people. I suggest that future people count simply to the extent that they fall under the scope of these principles—just like present people do. (One contingent difference is worth mentioning, if only to set it aside: while it is very easy to, e.g., kill a present person—so easy it can be done accidentally—it's much harder to kill a future person. As a result, present people affect what we should do, by the lights of these T-principles, much more frequently than future people do; but so too do my neighbours affect what I should do much more frequently than, e.g., those living in Patagonia.)

4. MacAskill thinks that future people count in a further way. He thinks they count in that

(*) we have reason to see to it that future people exist.

Because there are so many future people—trillions, in expectation—(*) has brave implications. It will suffice here to highlight just one. In a choice between (a) preventing the deaths of *n* (present) people (from, e.g., starvation) or (b) reducing the risk of human extinction by, say, one-thousandth of a percent, (*) returns that we should opt for (b) even when *n* is in the thousands—perhaps even when it's in the millions. Why? Because future people won't exist should humanity go extinct; yet those future people count; they count in that, as (*) says, we have reason to see to it that they exist; and given how many future people there are, those reasons quickly swamp local considerations—even ones as strong as preventing the deaths of, say, 10,000 people; and thus we should opt for (b).[3]

[3] One must be vague about these numbers because it's not clear what the right numbers are even when future people are excluded. How many people should one let starve in order to reduce the risk of 7 billion people dying by .001%? Whatever that number is, MacAskill's view is that the number is enormously bigger when extinction is at risk.

Do we have any reason to accept (*)? Not from anything said so far since there is no sensible route from T-principles to (*). Such a route would have it that if some future people fall under the scope of a given T-principle (and thus they count in virtue of doing so), then we have reason to see to it that those people exist. (For example, it would have it that since those people hiking on the second trail in 200 years fall under the scope of (T1), we therefore have reason to see to it that those hikers come to exist.) But this cannot be right.

Suppose that a maternity ward sits mothballed, only to be opened if and when the birth rate increases such that additional maternity capacity is required. Annie plainly shouldn't instal a bomb in that ward that is triggered by the first cry of a newborn. Why not? Because, by installing such a bomb, there's a good chance that Annie would kill someone—*viz*. the next baby born in the ward—and thus (T1) returns that Annie shouldn't instal the bomb, which is surely right.

Yet it can hardly be thought that the fact that Annie shouldn't bomb the mothballed maternity ward *entails* that Annie has reason to see to it that the birth rate increases such that babies are born in the ward (in virtue of those babies falling under the scope of (T1)). What a non-sequitur! As I say, there is no sensible route from the T-principles to (*). And since it is those T-principles that explain why it doesn't matter when someone will step on Annie's broken bottle, nor is there a sensible route from that datum to (*), either.

5. MacAskill suggests another route to (*)—again starting with a common sense datum, but this time going via climate ethics.

Here is a choice society faces concerning the storage of radioactive waste (from Parfit 1983). We can either adopt the *safe* policy which is expensive, but will keep the waste safely stored into the distant future; or we can adopt the *cheap* policy, which saves money by safely containing the waste for only 200 years, at which point it will leak out and radiate those living nearby—cancer, etc. Why shouldn't we adopt the cheap policy (it being common sense that we shouldn't)?

This is a thorny question because if we do adopt the cheap policy, then those radiated by the waste wouldn't have existed had we not adopted it. Why not? Because the sperm and egg of conception are

essential to one's identity, yet there's no way that the very same sperms and eggs would fuse in decades-time regardless of which policy is adopted. And if those radiated wouldn't have existed had we not adopted the cheap policy, they aren't worse off as a result of our adopting the policy; and if no one's worse off as a result of our adopting the cheap policy, yet it saves us money…why shouldn't we adopt it?

MacAskill says it must be because we have reason to see to it that future happy people exist; and if that's right, then (*) quickly follows (2022: ch8). Must it be right?

We've already seen that we shouldn't, other things equal, give people cancer. It's surely also true that we shouldn't give lots of people cancer, even if doing so will save us some money. Yet if we adopt the cheap policy, then we will give cancer to all those people whose cells are mutated by the radioactive waste when it leaks out in 200 years time (just as Annie would give cancer to anyone who steps on her radioactive waste in 200 years time), and we will have done it to save ourselves some money. So we shouldn't adopt that policy. Isn't that sufficient explanation? I think so.

It's simple too and it follows from the T-principles already introduced, again leaving no room for MacAskill's (*).[4]

Of course, there are other arguments for (*): it stands, for example, a short hop, skip and jump from total utilitarianism. But MacAskill takes his conclusions to follow from common sense premises —"the premises are simple, and I don't think they're controversial" (ch1)—and so it matters that (*) doesn't follow from common sense premises, after all. And it's little substitute that they might follow from premises as controversial as total utilitarianism instead.

6. Future people count. In what way do they count? They count simply to the extent that they fall under the scope of T-principles: other things equal, we shouldn't kill future people or break their arms or flood their homes or give them cancer, etc.; and, other things equal, we should help them when we

[4] Alas, it doesn't generalise to the non-identity problem in general; in particular, it can't explain why a mother shouldn't conceive a blind child this month, when she could instead conceive a sighted child next month (Parfit 1983). Other things equal, we shouldn't blind children, but that isn't what happens here: by conceiving immediately, the mother doesn't blind the child, but instead she just conceives a child that is blind. I add that this case doesn't support (*) since no one thinks the mother is required to conceive next month, but only that *if* she is going to conceive this month or next, *then* she should conceive next month. And that conditional claim is a long way from its antecedent and, in turn, from (*).

are able. Which is a long-winded way of saying that future people count in just the same way that present people do.

References

REDACTED.

Hare, Caspar. 'Obligation and Regret When There Is No Fact of the Matter About What Would Have Happened If You Had Not Done What You Did'. *Noûs* 45, no. 1 (March 2011): 190–206. https://doi.org/10.1111/j.1468-0068.2010.00806.x.

MacAskill, William. *What We Owe the Future*. New York: Basic Books, 2022.

Parfit, Derek. 'Energy Policy and the Further Future: The Identity Problem'. In *Energy and the Future*, edited by Peter G. Brown and Douglas MacLean. Maryland Studies in Public Philosophy. Totowa, N.J.: Rowman and Littlefield, 1983.

Thomson, Judith Jarvis. 'The Time of a Killing'. *The Journal of Philosophy* 68, no. 5 (1971): 115–32. https://doi.org/10.2307/2025335.