How Effective Altruism Lost the Plot

Utilitarianism is an ethical theory with a long and distinguished history. It is a species of consequentialism, the theory that tells us we should seek to act in ways that will maximize the goodness or value of the consequences of our actions.  As G. E. Moore dramatically put it, “Our ‘duty’… can only be defined as that action, which will cause more good to exist in the Universe than any possible alternative.” What makes utilitarianism distinctive is its understanding of what goodness or value consists in, namely happiness or utility.

Though many of the philosophers I most esteem and respect are utilitarians, I am not. There are several good reasons not to be a utilitarian. One, I think, fatal, problem is that a theory that tells us to perform at any given time “that action, which will cause more good to exist in the Universe than any possible alternative” is a theory that fails spectacularly to do what we want an ethical theory to do: offer some practical guidance in life. The Universe is just way, way too big, the future ramifications of at least many of your actions way too vast for us to have even the faintest idea what actions will cause more good to exist than any other, not just proximally but in the very very long term, from now to the heat death of the Universe.

Utilitarianism is often criticised for demanding too much of us, imperiously robbing us of any autonomy by seeking to control and direct every aspect of our lives. Really it has the opposite problem. It demands *nothing* of us. Entirely clueless as we are about the long-term consequences of our actions, any choice we make makes as much sense as any other.  Utilitarianism is a fast track to nihilism.  It's not that happiness is not important. But the Universe is above our pay grade.

Recently a sect of academic philosophers and others embracing the core utilitarian ideal of maximizing goodnesss, flying under the flag of ‘Effective Altruism’, have established themselves as a loosely organised movement. This movement originated with some young people in the late noughties determined to dedicate themselves to doing good in effective ways, many publicly pledging to give away a large percentage of their income over their lifetimes to charitable causes. It was all, at least at first, about mosquito nets and parasitic worm infections, advocating and promoting costs-effective charities doing good things for public health in some of the poorest countries in the world.

Which is, of course, wonderful. Generosity is an important virtue and people exercising it on a large scale deserve our admiration and respect. They deserve our admiration and respect no less when they draw inspiration from some narrative we may not share like, for many, Christianity or, at least for me, utilitarianism. To criticise them for that reason would be meanness, a real and obnoxious vice.

Increasingly, however, focus on mosquito nets seem to be getting rather crowded out by different, far more dubious agendas. Increasingly dominant among these is something calling itself *Longtermism*. This has its origins in 2003 when Nick Bostrom published a paper called “Astronomical Waste: The Opportunity Cost of Delayed Technological Development”, a paper so astronomically potty I fear when I read it first I thought it was perhaps a parody. Certainly the argument if sound would constitute a devastatingly effective *reductio ad absurdum* of its premises. But Bostrom’s paper, and subsequent writings, proved highly influential in some circles and many of the leading figures of EA now make furthering his Longtermist ideas their central preoccupation.

The total resources of energy and useful forms of matter to be found in the entire universe are, to put it very mildly, a very great deal more than is needed to sustain the lives of such living creatures that there are at the minute. And that, you might think, is no bad thing. What can be nicer than having more than we need? But that is not Bostrom sees it. For Bostrom accepts the *total utilitarian* doctrine that not only is it important that people should be happy, it also of urgent importance that happy people should be *numerous*, effectively assigning moral weight to nonexistent sentient beings. So for all that energy to be going unused is for Bostrom a colossal moral scandal, one we make it an overriding priority to address. If it’s there, why aren’t we using it all? Just think how many sentient beings we could then accommodate. Longtermists like to throw eye-wateringly large numbers around on the basis of some decidedly wild reasoning but, as Bostrom himself tells us, “What matters for present purposes is not the exact numbers but the fact that they are huge.” (AW,309) So let us just say LOTS.

This, as I said, is a scandal and something must be done. And by something I mean we, or our descendants – I use the word loosely as will become clear - need to *colonise space*. It is an unfortunate sounding rallying cry. Colonising stuff is an old idea with an unfortunate history. Actually, it is worth remembering here, it is an old idea whose justificatory narrative in relatively recent times was in considerable measure supplied by the utilitarian philosophers of 19th century England. Most of the planets out there are, it seems a safe bet, not very hospitable to life. The ones that are so hospitable may very well turn out to be occupied already by creatures that may very well not be pleased to see us.

Though perhaps by then what we presently think of as hospitality to life may not matter so much. Hospitality to life is hospitality to familiar, natural, carbon-based biological life and the future for these folks is very much *postbiological*, peopled by synthetic digital creatures with which we will have replaced ourselves. And everything else.

You should not be so foolish and sentimental as to register that as a loss. I think it helps to get the hang of the mindset here to take a look at another area that has garnered a lot of recent attention with the EA movement.

\*\*\*

Moralists have long been preoccupied by the condition of domestic farm animals but lately there has been more focus on *wild* animals. The thing about wild animals, we are told, is that they have horrible lives being preyed on by other wild animals. Some EA thinkers advocate we should keep them around but contrive to reengineer the wild animal biosphere to make it nicer, less disagreeable for its denizens. Others go further and cheerfully welcome what so many others deplore, the steady human encroachment on wild animal habitats that is driving may species to extinction at an alarming rate. This, it turns out, is a good thing and it would be no bad thing if it continued. Thus William MacAskill writes in his *What We Owe The Future*:

It’s very natural and intuitive to think of humans’ impact on wild animal life as a great moral loss. But if we assess the lives of wild animals as being worse than nothing on average, which I think is plausible (though uncertain), then we arrive at the dizzying conclusion that from the perspective of the wild animals themselves, the enormous growth and expansion of *Homo sapiens* has been a good thing.(213)

In due course. I guess, plants can go too. Digital creatures don’t need organic food or an atmosphere. The future is postbiological.

As I write this the WWF is reporting on the 69% decline in global wildlife populations since 1970. But really, it would appear we should relax and not indulge in sentimental regret as the whole world of sea lions and sparrows, bears and bumble bees is swept away. Viewed in a properly rational and scientific spirit, these creatures can be seen for what they are: compared to Bostrom’s imagined interstellar army of sentient bots living fake lives on distant worlds, just a pathetically wasteful and ineffective way of converting solar energy into utility, which, from the point of view of the universe, is literally the only thing worth caring about.

Of course the breezy embracing of anthropogenic ecocide horrifies some people. It certainly horrifies me. But such horror is just giving way to a sentimental and romantic nature-worship, an irrational hangup from the days when we thought of the natural world as something sacred. It’s not. God didn’t make each little flower that opens, each little bird that sings, evolution did, a blind and amoral process whose imperfect designs we can and should improve on enormously when we are good enough at technology.

I guess the proper response here is that is not so much a question of worshiping nature but of respecting it. Where that is a way of valuing that finds its main expression not in seeking to *improve* it but simply in *leaving it alone*, recognizing that it is not ours to improve. It’s the imperialist impulse again. There are so any things in the world that stand in need of improvement: Space, the lives of all those fish, the Indian subcontinent. And how can we hope to improve what we do not control?

\*\*\*

It's a widely recognized methodological tenet of contemporary moral philosophy that if starting with an attractive-seeming principle leads you to some conclusion that is highly intuitively repugnant to moral common sense you need to think again about the principle. But many now reject this methodology, taking as they do a dim view of moral common sense. They are here influenced by the writings of Peter Singer, who is something of a godfather to the EA movement, and his coathor, Katarzyna de Lazari-Radek. Singer and Lazari-Radek take a dim view of moral intuitions which they see as rather disreputable, distorting influences. Except of course the intuitions which they like, which are the ones expressed by utilitarianism. The latter are the “axioms” on which we can safely base our moral thought. This utilitarian exceptionalism is based on two premises. The first is that any authority we would accord an intuitive judgement is discredited where that intuition is susceptible to some evolutionary explanation. The second is that utilitarian intuitions are distinguished by being not so susceptible. But the second premise is very questionable biology and the first extremely questionable philosophy. That we have any altruistic impulses at all, caring for and looking out for each other, undoubtedly reflects the way we have evolved as social animals like baboons and not as solitary loners like the desert tortoise. But that, if it is true, does not remotely imply that our altruistic impulses or the odal codes which express them are in any way undermined.

The longtermists’ concerns sometimes seem a bit odd. They look to a fantastic posthuman future where we replace ourselves with digital synthetic consciousness but their greatest concern seems to be that we will be wiped out of existence by hostile out of control AI intent on replacing us, just like in the movies (think *2001: A Space Odyssey* or the *Terminator* franchise). But this looks like a strange combination of views. It seems we want to avoid being wiped out and replaced by machines long enough for us to figure out how replace ourselves with machines.

MacAskill seems to address this in chapter 4 of his book where he suggests (88) that he is happy enough if we get replaced by digital sentience providing it has the right values. But I don’t really know what right values would look like for digital sentience. Digital Dickie hanging out round Alpha Centauri in a postbiological world a few million years hence is not, we might suppose, a social animal like you and me. He has no need for love and affection and no sense of community. Did I say ‘He’? Of course I should not. He, she… it has no sex in both obvious disambiguations of that phrase. It doesn’t have children it loves and cherishes. There is no reason to suppose it is troubled at the prospect of its own death if it even has enough recognisably in the way of personal identity for concepts of survival and death even to apply to it. For something so remote from fleshy organic human experience anything recognisable as human values would likely be of little relevance and make little sense. Will such a creature ever replace us and fill the universe? I haven’t got a clue and I don’t care.

For almost all species of living creature on our planet the biggest existential threat is human beings. In spite of that dispiriting truth, I very much hope, for the sake of our children and grandchildren and their children and grandchildren, that we do not go extinct any time very so. Money spent avoiding this would be money well spent. Might some more distant catastrophe finish us off before we have time to indulge the postbiological fantasy of peopling the universe with a colossal population of blissful synthetic sentience. It likely will and the end when it comes will, like many endings, be very sad but such sadness as there is will not derive its warrant from the foreclosing of this idle fantasy.

The new longtermist strain of EA, for all its absurdities is increasingly powerful and very well funded. It feeds the hubristic fantasies of some very very rich people with lots, one might even say LOTS of money to throw about. Émile Torres has described the Longtermist narrative as “quite possibly the most dangerous secular belief system in the world today”. Perhaps that is hyperbolic. I am not sure. History might perhaps make us a little wary of philosophers whose aspiration is not to merely understand the world but to change it. It should certainly make us *very* wary when they are people who have convinced themselves that they have seen through the folly and illusion that is everyday common-sense morality, going beyond it to put their vision of our historical destiny on what they consider a proper, rational, scientific footing. And we should be *extremely* wary when they look forward to an imagined future order supposed so vast and so sublimely wonderful that advancing the cause of bringing it about is worth enormous costs to ourselves and the people with whom we share a still untransformed world, now, today. People who think in these ways are very dangerous indeed.

We should praise, ungrudging and cheerfully, generous people who go above and beyond to support medical charities in poor counties. But we should also deplore and challenge the noxious folly of Longtermism with its dismal fantasies about posthuman, postbiological futures where we – or something – will finally find secular Salvation through the transformative power of technology.

\*\*\*

Bibliographical note. The Bibles of the first phase of effective altruism were William MacAskill, *Doing Good Better* and Peter Singer, *The Most Good We Can Do*. Bostrom’s “Astronomincal Waste” is in *Utilitas* 15 2003. MacAskill’s massively hyped *What We Owe The Future*, clearly planned as the new Bible for Longtermist EA, was published earlier this year. For EA thinking about wild animals, see e.g. Brian Tomasik “Habitat Loss, Not Preservation, Generally Reduces Wild-Animal Suffering” and other essays at <https://reducing-suffering.org> and MacAskill WWOTF, 211-213. On intuitions see Peter Singer and Katarzyna de Lazari-Radek, *The Point of View of the Universe*, chapter 7. The quotation from Moore is from *Principia Ethica*, 148. The quotation from Torres is from an essay they published in Aeon (<https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo>). The WWF Living Planet report is here: https://wwflpr.awsassets.panda.org/downloads/lpr\_2022\_full\_report\_1.pdf.