

The turn for ultimate harm: a reply to Fenton

Ingmar Persson,¹ Julian Savulescu²

¹Department of Philosophy,
University of Gothenburg,
Sweden.

²University of Oxford, Oxford,
UK

Correspondence to

Dr Julian Savulescu, University
of Oxford, The Oxford Uehiro
Centre for Practical Ethics, Suite
8, Littlegate House, 16–17 St
Ebbes Street, Oxford OX1 1PT,
UK; julian.savulescu@
philosophy.ox.ac.uk

Received 27 August 2010
Accepted 21 December 2010
Published Online First
22 February 2011

ABSTRACT

Elizabeth Fenton has criticised an earlier article by the authors in which the claim was made that, by providing humankind with means of causing its destruction, the advance of science and technology has put it in a perilous condition that might take the development of genetic or biomedical techniques of moral enhancement to get out of. The development of these techniques would, however, require further scientific advances, thus forcing humanity deeper into the danger zone created by modern science. Fenton argues that the benefits of scientific advances are undervalued. The authors believe that the argument rather relies upon attaching a special weight to even very slight risks of major catastrophes, and attempt to vindicate this weighting.

THE POSSIBILITY OF A TURNING POINT OF SCIENTIFIC DEVELOPMENT

In an inventory of catastrophic threats, especially those generated by modern science, the eminent British physicist Martin Rees surmises that ‘the odds are no better than fifty-fifty that our present civilisation on Earth will survive to the end of the present century’.¹ This is of course only the roughest of estimates, but it is certain that the astounding progress of science and technology has significantly increased the risks of globally devastating catastrophes, for example by creating weapons of mass destruction and by environmental destruction.

We suggest that this justifies the claim that, all things considered, scientific development has lately been for the worse rather than for the better. This claim does not imply the absurdity that life was better in prehistoric days, before the advent of any science and technology, than it is today. It is rather that there has been a *turning point* in the development of science and technology: up to that point development has been for the better all told, but after that point it has been for the worse. Our proposal is that this point be defined as the stage at which science and technology put in the hands of humans the means of destroying or seriously damaging forever the conditions of sentient life on this planet, and that this point occurred sometime in the middle of the preceding century.

Let us assume that in general sentient life on this planet is better than non-existence, so that there is a net balance of the good over the bad. We can then define an *ultimate harm* as something that ensures that there will never anymore be such a net balance of the good. Something could be ultimately harmful by forever extinguishing sentient life, or by damaging its conditions so drastically that, in general, life will not henceforth be worth living. Our claim is that the development of science and tech-

nology turned for the worse, all things considered, at the point at which it put in the hands of humankind the powers of doing ultimate harm.

It might seem strange to claim that sometime in the middle of the last century there was such a turning point because we surely did not experience things turning for the worse. If what makes for badness is the threat of ultimate harm, things might be experienced as getting steadily better right up to the end, until the final catastrophe that eradicates all sentient life. Suppose that if science and technology had not developed to the degree that made this fateful instance of ultimate harm achievable, the average quality of life would have remained somewhat lower, but life would have gone on for much longer. Then in virtue of this longer duration, this scenario would have been better, all things considered. To be sure, it might have been even better if there had been development to the extent that brings forth the power of doing ultimate harm, but this power is never used. However, that might be a state that is not possible given human nature. So we should be alert to the possibility that a turning point of scientific development of the type we are talking about could sneak in upon us unawares.

It is therefore not true, as Elizabeth Fenton claims in her response to an earlier paper of ours,² in which we contend that modern science has placed humanity in the described precarious situation, that our argument commits us to ‘the surprising conclusion that *all* forms of scientific progress are instrumentally bad for humans overall’.³ Our argument only commits us to believing that this is true of scientific development by and large of recent date. Fenton’s main criticism of us seems to be that we have ‘dramatically undervalued the *benefits* of significant scientific advances’ (Fenton pg 150)³ and, in particular, the value of ‘non-traditional’ cognitive enhancement, ie, cognitive enhancement by biomedical or genetic means. We will in a following section turn to an assessment of this criticism as it applies to the present and future.

THE ACTUALITY OF THE TURNING POINT

So it is possible to hold that the present level of scientific development is more harmful than beneficial, without being forced all the way back to the absurd position that we would have been better off without any scientific technology at all. This is because it is possible that there has been a point in scientific development at which it took a turn for the worse, but this is not to say that we have passed such a turning point and that the present level of scientific development is more harmful than beneficial. We now want to argue that this is indeed the case, even if the risk of ultimate harm is appreciably smaller than Rees¹ estimates.



This paper is freely available
online under the BMJ Journals
unlocked scheme, see [http://
jme.bmj.com/site/about/
unlocked.xhtml](http://jme.bmj.com/site/about/unlocked.xhtml)

In our paper (Persson pgs 173–4),² we appealed to the following simple sort of case. Suppose that the level of welfare could be assigned a number, say, 100 units, and that the relevant probabilities could also be assigned numerical values. Consider now your choice of whether or not to do or undergo something in which the probability that you will gain 2 benefit units is 99% (not ‘2% chance of a small gain’, as Fenton writes), (Fenton, pg 149)³ but the risk that you will lose 100 units is 1%. An example would be an operation that will almost certainly correct a small defect of yours, eg, in your visual or motor ability, but that could conceivably go seriously wrong, although this is very unlikely, and make you blind or paralysed. Then the expected value of the possible improvement, obtained by multiplying the welfare value with its probability, is nearly double the expected disvalue of the possible loss. Standard decision theory would declare it rational to undergo the operation, but many of us would regard it as more reasonable not to do so. It would strike us as almost insane to run even such a small risk as one in a hundred of being blind or paralysed in order to gain an increase of wellbeing that is relatively insignificant.

This exemplifies the phenomenon that Daniel Kahneman and Amos Tversky call *loss aversion*, that for us ‘losses loom larger than gains’.⁴ In contrast, as they point out, most of us happily accept a very high risk, amounting to almost a certainty, of a small loss in order to obtain a vanishingly small chance of a very great gain. This is something we often do when we gamble, but then it is clear that it is only big losses to which we are averse. So ‘big loss aversion’ would be a more appropriate name. Its positive counterpart is accordingly a big gain attraction.

We believe that the explanation of the big loss aversion and the big gain attraction is that our conative–affective reactions are not finely attuned to small differences in probability, except when this amounts to a change to certainty or impossibility. Thus, an increase in the probability of a big loss from 0.85 to 0.90 might not noticeably affect our emotion of fear, while an increase in it from 0.95 to 1 could change our fear to horror.¹ So, a loss that is big enough to have the potential to stir up our fear will do so to the same degree, irrespective of whether the risk of it is 0.05 or 0.1. Similarly, as regards a gain that is big enough to stir up our hopes or wishes.

This account seems to imply that big loss aversion and big gain attraction are irrational attitudes that we should resist in favour of reactions that concord with expected utility theory. We think indeed that we should resist big gain attraction, because it could lure us to gamble until we are destitute, but we believe that big loss aversion is useful because, as we observed in our paper, ‘it is, as a rule, much easier to harm than to benefit’ (Persson, pg 173).² We could distinguish two aspects of this greater power of harming.

First, the magnitude of harm produced could normally be larger than the goodness done. For instance, we can normally kill, or cause pain to, many more individuals than we could save the life of, or relieve of pain. The second aspect of the greater power or ease of harming is that if we compare harms and benefits of a similar magnitude, there are likely to be many more ways of causing harm available to us, for instance, more ways of disturbing a well-functioning organism, or some bigger arrangement like the ecosystem, than of improving it to the same extent. This is why a random interference with a well-functioning system is enormously more likely to damage it than to improve

upon it. It follows that the risk of disturbance is generally greater than the chance of improvement of the same scale.

It is this greater power of harming, which, in our view, makes it reasonable to abide by the big loss aversion. As we have seen that there are usually many ways in which harm could be done, we should calculate with a probability that we have overlooked some of these ways, even if our investigation into the matter has been as thorough as time allows. Some of these ways may combine with the ways we recognise to create a greater probability. Furthermore, as our capacity to cause harms of great magnitude is considerable, we should take into account the risk that we have overlooked some harms that are greater than the ones we envisage. This makes reasonable an attitude of aversion, or at least precaution towards risks of big harms/losses.

Now, the progress of science and technology has magnified this greater power to cause harm to the point where we could cause what we have called ultimate harm. The negative instrumental value of ultimate harm is indefinitely high because there is no way of telling how much of a net balance of goodness it prevents, ie, how much of worthwhile life there would have been in the future had it not occurred. This fact, in conjunction with the fact that we might well have overlooked some of the factors that contribute to the risk of ultimate harm, seems to us to make it reasonable to demand that we try to minimise the risk of ultimate harm, whatever the expected gain of the alternatives might be (within realistic limits). It follows that if, as we have suggested, the development of science and technology since the middle of the last century has increased the risk of ultimate harm in order to secure relatively minor improvements of welfare overall (even if the benefits to some individuals have been profound), it has been for the worse, all things considered.ⁱⁱ

Yet, there is in fact no general aversion to this development of science—why? One reason is that there is probably not a general awareness of all the risks of ultimate harm that it generates. This might be what motivated Rees to write his book. Another reason is that there is a further feature of our mental make-up that has to be in place for the big loss aversion to come into operation, namely the *availability bias*: we are fixated on the possible occurrence of events of which we have readily available images, largely as a result of recently having experienced events of these kinds. Our emotions are geared to how vividly we imagine possible events rather than simply to how we abstractly estimate their value and probability. To take an everyday example, you might cycle to work every day in heavy traffic, without feeling any fear of being the victim of a serious accident, although you know that there is a not negligible probability that you might be. However, if you do have a serious accident, this might make you so terrified of cycling in the traffic that you are unable to do it for a long time. It is not that the detailed memory of the recent accident makes your estimate of the disvalue or probability of another, similar accident rise steeply; instead it makes you imagine more vividly what it is like to have such an accident. In the course of time the vividness of this memory usually fades, and with it the vividness of images of possible accidents in the future, so eventually your fear might subside, and you might be able to resume your cycling.

The availability bias was doubtless active in the USA in the years after the terrorist attack in New York City on 11 September 2001, to inflate a fear of future terrorist attacks. In a cross-national study made in the USA a couple of years after 9/11, US citizens on average estimated that the risk of their

ⁱThis deficient sensitivity to probabilities is not the same as the notion of a diminishing sensitivity to probabilities the greater their distance from impossibility and certainty, which alongside loss aversion, figures in Kahneman and Tversky’s alternative to expected utility theory, ‘prospect theory’ (Kahneman pg 50).⁴

ⁱⁱThis is a stronger conclusion than we arrived at in our earlier paper, in which we concluded that this development is worse in one respect. (Persson pg 174).²

being seriously harmed in a terrorist attack in the next year was 8.27%. Even though there is no way of accurately determining what the actual risk was, this figure is certainly extremely exaggerated. The risk of their dying in a motor vehicle accident was only 0.015%, and in 2001 a US citizen was statistically 15 times as likely to die in such an accident as in a terrorist attack.⁵

However, at least some of the ultimate harm with which modern science and technology threaten us is novel, and hence not available to our imagination. One example is a global nuclear war. Another example is a devastating climate change and environmental destruction to which all human beings—but especially those in affluent countries—might contribute by using products of modern technology. Here the big loss aversion seems blocked by the availability bias: we are so accustomed to the advance of scientific technology boosting our living standards that we find it hard to imagine that it will not continue to do so indefinitely into the future. So, we jeopardise the future of humanity, by not making comparatively small welfare cutbacks—contrary to the attitude of precaution that we have recommended.

THE POSSIBILITY OF A WAY OUT OF THE PREDICAMENT

How could we get out of this perilous state in which the mismatch between our technological power of action, on the one hand, and our cognitive fallibilities and moral shortcomings, on the other, lands us? To begin with, this technological advance itself produces means that could help us out of the predicament to which it has given rise. For instance, it supplies more effective surveillance techniques with the help of which political authorities could forestall the kind of terrorist attacks with nuclear or biological weapons of mass destruction that we discuss (Persson pg 166–7).² Certainly, if these means of surveillance are employed, citizens' rights to privacy will be curtailed, but the increased security might be considered to be worth the price. Similarly, if freedom of speech is restricted by a prohibition to publish information about potentially lethal pathogens and toxins.

When criticising our alleged undervaluation of the benefits of scientific progress and of non-traditional cognitive enhancement, Fenton (Fenton pg 150)³ appeals to climate change as an instance of a problem to which these benefits might provide the solution. We certainly do not want to deny that scientific breakthroughs could be a crucial asset in the attempt to curb climate change, by providing new forms of clean energy, goengineering, etc., but we do not believe that technology by itself could solve this problem. It seems likely at present that whatever the technological advances in the time at our disposal, humanity will not be able to reduce the emission of greenhouse gases to a tolerable level without making some sacrifices of welfare. Even if there were to be new, wonderful clean energy, the transition from using old dirty energy to using it would be likely to be costly: imagine, for instance, replacing all the hundreds of millions of cars powered by petrol by cars driven by hydrogen-powered fuel, and all petrol stations by hydrogen fuelling stations. So, to a considerable extent the problem of climate change is, and will remain, a moral problem: to solve it the present generation needs a will to cut down on their consumption and welfare in order to leave the planet in a more hospitable condition for future generations.

As a further respect in which scientific progress could make a necessary contribution, Fenton (Fenton pg 150)³ refers to the task 'to make more gains available to those not lucky enough to enjoy them currently'. She might have in mind both the problem of global injustice—that some countries are affluent while others are destitute—and the problem of intra-social injustice, that even in the most egalitarian societies there are still

huge differences in welfare between the best off and the worst off. Even more clearly than the problem of climate change, these problems of justice are mainly moral problems, problems that exist because of our moral failings rather than because our technology fails to produce enough material goods to endow all people on earth with a decent standard of living. In 2008 only five countries had reached the modest goal that the United Nations set decades ago of foreign aid amounting to 0.7% of a country's gross national product. The average for Organisation for Economic Co-operation and Development nations is 0.47%; the two biggest world economies, the USA and Japan, lie at the bottom, at approximately 0.2%.

Humans fail to deal with problems of climate change, global and social inequality because of their limited altruism—their capacity of having genuine concern only for a few people who are near and dear to them—their incapacity of sympathising with great numbers of people in proportion to their number, their discount of the more distant future, their feeling of greater responsibility for what they cause than for what they let happen, and so on.⁶ In our paper we suggested that it is worth exploring the possibilities of biomedical and genetic means of moral enhancement because we think that these features are so deeply entrenched in human nature that otherwise it might be hard to remove them. It is not that we believe that no moral progress could be achieved by reflection and traditional means of moral education. The fact that it is now widely recognised that all humans have equal rights might be the most important instance of moral progress accomplished by these means (mentioned by us² Persson pg168 and by Fenton pg 148).³

This example also reveals the limitations of success by these means, for although the equal rights and worth of all humans are widely endorsed, the economic inequality of the world is arguably greater now than it was before this egalitarian creed conquered the world. For instance, 'the difference between the per capita incomes of the richest and the poorest countries was 3 to 1 in 1820, 11 to 1 in 1913, 35 to 1 in 1950, 44 to 1 in 1973, and 72 to 1 in 1992'.⁷ The economic inequality of the world seems to have grown in step with the growth of wealth made possible by modern technology, in spite of the official doctrine of egalitarianism. That is to say, this doctrine has not sunk in deeply enough to shape our behaviour.

It might be suggested that the resources of traditional moral education have not been exhausted. Liberal societies, with their traditional ideal that the state interferes minimally with the lives of citizens, may have downplayed the important task of moral education. Presumably, if traditional moral education were practised so intensively from an early age that it amounts to brainwashing, it could be effective enough, but short of that, we believe that, to avoid intentional and unintentional misuse of the powers of causing ultimate harm with which science and technology increasingly supply us, we would need judicious and extensive employment of biomedical and genetic methods of moral enhancement that we are far from possessing at present. To gain possession of these methods, we stand in need of further development of science and technology, perhaps speeded up by non-traditional cognitive enhancement if the process is to be swift enough. Thus, we are prodded further into the danger zone created by sophisticated science and technology; things will have to be worse before they could get better. Among the things that could be misused are the new powers of cognitive and moral enhancement. Obviously, there is a bootstrapping problem about a judicious use of techniques of moral enhancement: it is humans with their current moral blemishes that have to apply these techniques to themselves.

Such was the dilemma we sketched in the article criticised by Fenton.³ We cannot find anything in her piece that motivates a retraction or revision of this dilemma. It seems to us that if we go wrong anywhere—obviously, these matters are so complex that it is difficult to be confident of not going astray somewhere—it is not by undervaluing the benefits of scientific progress, and of non-traditional cognitive enhancement, as Fenton stresses,ⁱⁱⁱ but by exaggerating the risks of them. People are so accustomed to the idea that scientific development reliably leads to an improved quality of life that they are hard put to believe that they might have passed the summit and are forced to descend into a dark valley before they could ascend again. Still, bleak as this picture is, we persist in thinking that it is the most realistic.

Acknowledgements This work was funded by the Wellcome Trust, Grant 086041/Z/08/Z.

Competing interests None to declare.

Provenance and peer review Not commissioned; not externally peer reviewed.

REFERENCES

1. **Rees M.** *Our final century*. London: William Heinemann, 2003:8.
2. **Persson I**, Savulescu J. The perils of cognitive enhancement and the urgent imperative to enhance the moral character of humanity. *J Appl Philos* 2008;**25**:162–76.
3. **Fenton E.** The perils of failing to enhance: a response to Persson and Savulescu. *J Med Ethics* 2010;**36**:148–51; 149.
4. **Kahneman D**, Tversky A, eds. *Choices, values, and frames*. Cambridge: Cambridge University Press, 2000:46.
5. **Sunstein C.** *Worst-case scenarios*. Boston: Harvard University Press, 2007:40–3.
6. **Persson I**, Savulescu J. Unfit for the future? human nature, scientific progress and the need for moral enhancement. In: Savulescu J, ter Meulen R, Kahane G, eds. *Enhancing human capabilities*. Oxford: Wiley-Blackwell, 2011.
7. **Seitz J.** *Global issues*. 3rd edn. Oxford: Blackwell, 2008:3.

ⁱⁱⁱWe find it particularly puzzling that Fenton³ accuses us of arguing that ‘in all cases’ the cognitive enhancement of anyone is bad news for others because we explicitly say that this is so only in cases in which ‘there is a clash of prudential ends’.² We do not believe that an enhancement race between individuals is necessary.