Biomedical Moral Enhancement – Not a Lever Without a Fulcrum

Ingmar Persson & Julian Savulescu

When we argued in *Unfit for the Future*[1] that moral enhancement – which might include biomedical moral enhancement, BME[2] – is necessary to solve the coordination problem presented by the amelioration of anthropogenic climate change, we used the familiar tragedy of the commons as a simple, small-scale model of this larger, more complex coordination problem. Stefan Schlag criticizes this argument, contending with respect to the tragedy of the commons that 'when it comes to the implementation of BME, a problem with exactly the same structure /as the tragedy of the commons/ emerges and impedes BME' (7).[3] In other words, an attempt to solve a coordination problem like the tragedy of the commons is 'self-defeating' (11). We are grateful to Schlag for his critique because it provides us with an opportunity to clarify and specify some features of our account.

The gist of our reply to him is that he overlooks a rather obvious fact: we can be sufficiently morally motivated to form an intention *in advance* to act in a certain (moral) way when a situation arises without being sufficiently motivated – lacking strength of will or character – to follow through this intention when are in the midst of the situation and actually experience the self-sacrifices this way of acting imposes on us. Thus, before we are tortured, we may form an intention not to betray our comrades, though we know the torture will be gruesome; yet, when we are being tortured, we may break down and betray them. Displays of weakness of will – well, it might be an exaggeration to brand those who break down under gruesome torture as 'weak-willed', but let that pass – are common as regards both moral behaviour and long-term self-interested behaviour. For instance, many smokers are enough motivated to kick their hazardous habit in order to form an intention beforehand to do so; yet, they aren't enough motivated to stick consistently to their intention.

Suppose that when forming their advance intentions, these people had access to some drug which would boost their motivation, so that they *would* be capable of following their intention through. Then, surely, they would voluntarily take it. So, in this sort of situation a

---

1 Oxford: Oxford U. P., 2012. We have run basically the same argument in several other places, but we will here focus on the book, as does our critic, Schlag.

2 In the book, we use the shorter term 'moral bioenhancement', but here we adopt the terminology of Schlag.

3 'The Tragedy of the Commons and the Dispensability of Biomedical Moral Enhancement', *Neuroethics*, this issue. Unprefixed page references in the text are to this paper.

biomedical enhancer could offer effective help. Why couldn't it do so in the tragedy of the commons as well? We might imagine that the herdsmen who have to cooperate in reducing the grazing of their cattle in order to prevent overgrazing of the commons are in circumstances parallel to the people in the preceding paragraph: they are sufficiently altruistically concerned about the welfare of all the herders and their families to form an advance intention to carry out the required cutbacks on the grazing of their cattle; nevertheless, when they begin to cut back and begin to *feel* the squeeze it puts on them, they backslide and defect from the agreement to cut back. Clearly, in their situation the availability of effective BME would be of help, just as it would be in the situations in the foregoing paragraph. Being sufficiently motivated to form an advance intention to cooperate, they are likely to be sufficiently motivated to undergo voluntary BME to help them implement it. For the point of forming an intention to act is to make us act in the way intended.

Schlag comes close to seeing the point we are making when he writes in the context of climate harm that BME 'might pose lower motivational obstacles to action than realising a number of more complex acts necessary for the reduction of emissions' (12). He also makes the important observation that BME 'could have the character of a general-purpose device… one effective action of enhancement might bring about extensive advantages in a wide range of future cooperation problems' (12). That is, suppose that people in general are morally motivated enough to have the intention beforehand to act morally in many types of situation. Then one kind of BME intervention could enable them to follow through their intention in *all* these types of situation. This would obviate the time-consuming undertaking of keeping an eye of their compliance in each type of situation. Schlag appears to be aware of this point when he writes 'onetime indirect stimulation of cooperation with general-purpose BME is more parsimonious' (12). Nonetheless, he drops these suggestions by reference to some unspecified 'indirect consequences' and the bald statement that cooperative behaviour still 'cannot avoid the risk of being fruitless due to free-riding of others' (12). Surely, a lot more argument on his part is called for here.

It might however be protested that we have stacked the cards in our favour by imagining the herders to possess enough moral motivation to form the advance intention to cooperate. But notice, first, that this moral motivation could consistently be thought to be present in the tragedy of the commons. It isn't required that this moral motivation is lacking, only that it isn't strong enough to withstand the self-interested motivation to defect before effective action has been taken. Secondly, if the herders' village is so small that everyone knows

everyone, it isn't unrealistic to think that many of them would have some altruistic concern for a lot of their fellows.

However, we don't need this assumption of some measure of altruism to show that BME could open a way out of the cooperation fix. Imagine that the herders are purely self-interested (except perhaps for being concerned about the welfare of their own families). What they would have most (self-interested) reason to try to accomplish would then be that all the other herdsmen (or a majority of them) undergo BME treatment to make them sufficiently motivated to carry through the cooperative endeavour, while they themselves sneak out of it, just as they would have most reason to try to accomplish that all the other herdsmen carry through the cooperative endeavour, while they themselves free-ride on their behaviour. Even so, the lever of BME can still find a fulcrum. For it might be possible for individual herdsmen to defect in secrecy from an agreement to reduce the grazing of their cattle, since this is an temporally prolonged process, although it *isn't* possible for them to arrange things so that they are let off the hook of having BME, while the rest of the other herders are subjected to BME because this treatment isn't a temporally drawn out process. It may be a matter of having a single intervention, such as a pill or an injection on one or several single occasions. Then each and every herder has to undergo BME, or be exposed to some sanction, like being excluded from the benefits of social life.

To be sure, in these circumstances undergoing BME wouldn't be *voluntary*: the herders are forced to undergo it on pain on having some sanctions imposed on them. But that fact can't be used as an objection to the employment of BME. For, as Ellinor Ostrom's work notably brings out, when people in fact succeed in avoiding the tragedy of commons, they do so by means of threats of sanctions on free-riders or defectors. Coercion in the shape of effectively implemented laws is an inescapable feature of functioning social life. Most of us – with the exception of radical anarchists – find this morally acceptable in principle.

At this juncture, we would like to draw attention to a disanalogy between the tragedy of the commons and the problem of alleviating anthropogenic climate harm: while the former might be solved by the reductions of *one* generation of herders, the latter is likely to be an *inter-generational* problem. Even if the present generation of adults effectively limit their emissions of carbon dioxide, and so on, this will be futile if the following generations, their children, grandchildren, and so forth, don't comply. However, these generations might be required to put up with *increasingly greater* sacrifices of welfare resulting from climate harm combatting measures: these generations pick up from a lower welfare level due to the cutbacks of the preceding generation, but they still have to implement further restrictions.

This might create the following situation. Suppose the present generation is sufficiently concerned about fighting harmful climate change to form an advance intention to impose substantial restrictions on carbon dioxide emissions, but that they aren't sufficiently motivated to keep up these restrictions. However, they have at their disposal BME that could supply the needed motivational boost. Now suppose they also realize that if they don't treat their children, too, they won't be morally better than they themselves were to begin with. As a result, their children mightn't be even capable of forming the advance intention to implement the necessary cutbacks, since these place greater burdens on them than the ones placed on their parents. In light of this realization, the parent generation might no longer be capable of forming the advance intention to cut down, unless they also intend to give their children BME at least to the extent that they will be capable of forming the requisite advance intention when they are grown-ups because otherwise the cutbacks of the parent generation will be wasted. Our view is that in this situation the parents could justifiably have their children BME-treated.

Needless to say, their children wouldn't undergo this treatment voluntarily. But, as Schlag notes (12-3), the fact that this treatment is non-voluntary doesn't make us shrink back from recommending the employment it. Upbringing of children must inevitably contain a lot of coercion, both directly for their own good, but also for the common good of society which by and large serves the good of the children. We force them to eat certain kinds of food rather than other kinds, to brush their teeth and wash, to go to school and to abide by certain moral rules, such as not to inflict physical harm, steal or lie. If there were some safe, effective BME which would make children more moral, we don't see why it couldn't be justifiably applied without their consent, like the other patterns of behaviour. Whether it should will depend on its risks and benefits. If we could spot would-be sociopaths or psychopaths early on, surely it would legitimate to apply BME to them if this was the only way to cure them of the personality disorder. This would be for the benefit not only for their potential victims, but also for their own benefit because otherwise they risk ending up in prison – a much worse form of coercion.

However, we think it's premature to spend much time on such speculations, since we're so far from being in possession of safe and effective forms of BME that we can have no definite idea of what they will be like. Thus, we are loath to provide any detailed prescription for how BME could facilitate tackling such an enormously complex, international cooperation problem as climate change. Suffice it to say that we think it could facilitate it along roughly the same lines as it could facilitate smaller scale cooperation problem. We aren't convinced that solving the problem posed by climate change with the expedient of BME necessitates the

establishment of a world government, as Schlag suggests (14). If, as he contends, BME creates a coordination problem 'with exactly the same structure' (7) as the original coordination problem, how could the solution of the BME-generated problem require a world government unless the solution of the original problem does?

In this response, however, our main objective has only been to illustrate how BME can help us solve some simpler cooperation problems. We believe that we have shown, contrary to Schlag, that it doesn't self-defeatingly raise structurally parallel cooperation problems. It's instead a potentially useful lever with a fulcrum in our actual nature and circumstances.

One final point of great general importance. In conclusion, Schlag writes: 'Social dilemmas are not caused by individual deficiencies but are rooted in the problematic structure of human interaction' (16). This is a pernicious confusion which he is far from alone in suffering from; we have met it again and again among our critics. What this claim implies with respect to the tragedy of the commons is that the herders' problem to cooperate isn't 'rooted in' their psychology or motivation. But it clearly is: it's rooted in the fact that their interest in pursuing the common good isn't strong enough to overcome their desire to minimize losses of their own good. It would be a mystery how problems could pop up in the dimension of 'human interaction' if they weren't 'rooted in' psychological – in particular, motivational – traits of the individual agents. It may be that these problems are solvable by the establishment of an effective system of surveillance and sanctions, or other social or political arrangements. But a more effective strategy might be to increase the agents' interest in the common good by means of BME. It may affront the self-respect and sense of human dignity of many people to have to concede that they are in need of being morally enhanced; perhaps that's something which explains why opposition to BME is so intense and extensive. But wishful thinking doesn't change our nature.

As Schlag mentions (10), we discuss a 'boot-strapping' problem in *Unfit for the Future*. Our conception of this problem is more capacious than he depicts it as being. We write that a first obstacle to be surmounted is that we have to 'be enough interested in being morally enhanced to put aside sufficient resources for research into biomedical means of moral enhancement' (124). Judging by the reception of our book, this is no minor obstacle.