Moral Bioenhancement:

Widespread Harm and Broad Cooperation

By

Morgan Lawrence Wisheart

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Master of Arts in Philosophy

Victoria University of Wellington

2015

Abstract

A controversial issue in contemporary bioethics has emerged in recent years: moral bioenhancement (MB). Human bioenhancement in general has seen its share of controversy, but it is generally agreed that there is potential to improve human physical and mental capacities through biotechnological interventions such as medicinal drugs and genetic modification. The discussion has turned to whether biotechnological interventions could similarly improve human moral capacities. Ingmar Persson and Julian Savulescu have argued that MB is imperative if humans are to survive into the future, because our current moral capacities do not equip us to address future catastrophic problems, Ultimate Harm, which will be caused by modern advanced scientific progress. I suggest related but distinct reasons why MB is appealing: scientific progress and deficient human moral capacities are jointly responsible for enormous amounts of harm all over the world, Widespread Harm, and MB has the potential to reduce that harm. Human moral capacities are deficient because of their dependence on what I call 'moral intuitions'; evolved psychological traits that shape our many societies' varied moral values, resulting in moral disagreement and the disruption of inter-group cooperation. Addressing modern problems requires a broader level of cooperation that is unlikely to be achieved by depending on our current moral intuitions. This is why our moral capacities should be improved. However, typical accounts of MB do not describe interventions that will improve our moral capacities in this way. They are focused on the vague objective of 'making people morally better', assuming that this will address human moral deficiency and that this will in turn address the resulting problems. 'Making people morally better' means making them more satisfactory to our current moral intuitions, which are the root of moral deficiency, so these MB strategies are unlikely to be effective. An alternative MB strategy, which I propose, instead focuses on the objective of modifying current moral intuitions so that they promote broad cooperation. This will result in improved moral capacities in the sense that our moral capacities will be more practically useful to us. However, because this strategy disregards the importance of satisfying our current moral intuitions, it will be morally unpalatable. This is its main disadvantage over the typical MB strategy, though it is better at handling many common objections. Ultimately, there are a number of practical concerns that cannot be completely satisfactorily responded to even by my new MB strategy, such as the issues of mandatory MB and of fine-tuning our moral capacities. These concerns may mean that MB is too risky, and therefore not the best course of action in response to modern problems rooted in scientific progress and moral deficiency, particularly since we have promising alternatives available such as traditional moral enhancement techniques and further scientific progress. The prospect of MB should continue to be investigated, but it should focus on improving upon our current problematic moral intuitions rather than better satisfying them.

Contents

Introduction

Recent scientific advances have resulted in the prospect of making people morally better through biotechnology. Can we do it? Should we do it? Some proponents answer yes to both questions. Julian Savulescu and Ingmar Persson argue that if we do not morally enhance the human population through biotechnology, there is a significant, growing risk that we will soon bring about the end of the world. There are also related problems that already cause massive amounts of harm and will cause more in the future, even if they don't threaten to cause an apocalypse. This harm is caused largely by the interactions between advanced technology and human moral deficiency. There are legitimate reasons to be concerned about the above harms, and successful moral bioenhancement (MB) could plausibly reduce or prevent them.

However MB is usually conceived of with a focus on more effectively satisfying some current set of moral intuitions. This focus is not promising. I am doubtful that MB of this kind could be effectively implemented, and it would probably not solve the problems that it aims to solve. If the objective of MB is to prevent certain kinds of catastrophic harm, we must actively focus the development of MB on that objective. I argue that motivating broad cooperation among humans through our moral capacities is potentially a better way to accomplish the objective of MB than by causing us to be more satisfactory to current moral intuitions.

It is not clear whether this new MB strategy would constitute moral enhancement. There is arguably a difference between doing what will prevent the most harm and doing what is morally right. MB may involve sacrificing the interests of many individuals for the sake of others, reducing freedom and diversity, or dangerous risks of going wrong. It is also possible that the same objectives we intend to achieve through MB could be achieved more efficiently, effectively, or safely through scientific advancement, cognitive enhancement, or traditional, non-biotechnological moral enhancement techniques such as moral education. Moreover, the plausible alternatives to MB are likely to be far more morally palatable.

Section 1 establishes some framework relevant to the moral bioenhancement debate, including typical definitions of enhancement, moral enhancement, and moral bioenhancement, and a discussion of the disastrously harmful consequences that have been foreseen should we choose not pursue MB. I argue for an improvement-

based enhancement model, that MB is not incoherent, and that there are reasons to consider developing and implementing MB because of its potential to reduce certain types of harm that presently exist and plausibly will exist in the future.

Section 2 explores some important moral concepts in order to determine more concretely what moral enhancement would involve. I emphasise the practical value of broad cooperation, and the fundamental importance of moral intuitions to moral value and moral motivation. However, current human moral intuitions do not motivate broad cooperation. This is primarily because moral intuitions are products of evolution and only promote cooperation within relatively small groups of individuals, while promoting conflict between such groups.

Section 3 outlines the typical MB strategy, which prioritises the satisfaction of current moral intuitions as a means to moral enhancement and prevention of harm. I argue that this strategy is both impractical to implement and unlikely to achieve its objectives. The problems MB aims to address could best be solved by motivating broad cooperation, which cannot easily be accomplished through current moral intuitions. I propose a new MB strategy that instead prioritises the motivation of broad cooperation, with the objective of addressing catastrophic harm. One way this might be accomplished is by manipulating our psychological moral mechanisms so that we emphasise moral intuitions that motivate broad cooperation.

Section 4 examines some prominent objections that have been raised against MB. Some dissolve under scrutiny or apply only to the typical MB strategy and can be addressed by the new one, such as the threat posed to autonomy. Some objections apply even to the new MB strategy, such as the prohibitive complexity of morality or the possible requirement for mandatory MB. We also consider how the prospect of MB measures up to traditional moral enhancement techniques, such as schooling and prison systems. Although it would be a bad idea to cease using such alternative measures, it may be equally unwise to depend solely upon them.

We will be left with a good understanding of why we might need MB, how it could work at its best, and how it compares and interacts with our best alternative options. We will also be aware of the most pressing dangers of moral bioenhancement, the unresolved controversies surrounding it, and gain an understanding of the limits of its alternatives.

1: Moral Bioenhancement and Harm

Our first task is to lay the foundations of the discussion. We need an understanding of what moral bioenhancement is, which will require that we understand the concepts of enhancement in general and of morality in general. At this stage, we only need to know enough for an understanding of why we might want to implement MB. We will more fully develop our concept of morality in section 2, and our concept of MB in section 3.

1.1: Enhancement and bioenhancement

What does it mean for a person to be 'enhanced'? In simple language, it means for a person to be made better. An enhancement improves someone's capacities. Flintlock pistols allow us to launch projectiles at high velocities, something humans normally have only very limited capacities to do. Flintlocks provided the best available improvement to those capacities during a certain time period. But modern firearms, such as the semi-automatic pistol, are indisputably better at the same job. One reason for this is that while single-barrel flintlocks can fire only once before a lengthy reload process, a semi-automatic can fire over a dozen times before a much briefer reload. If your objective is to fire lethal projectiles at high velocities, and you have a flintlock, then equipping yourself instead with a semi-automatic would constitute a clear enhancement.

Here we see the first defining point of enhancement: We need to know which capacities are to be improved. It makes no sense to ask whether giving someone a gun makes them overall better. If they are not going to be shooting anyone and instead, say, are about to run a 100-metre dash, a gun will do nothing but weigh them down. We must always think of enhancements in more or less specific terms. A physical enhancement improves an individual's physical capacities. A sprinting enhancement improves an individual's sprinting capacities. An χ enhancement improves an individual's χ capacities.

The above also reveals our second defining point: Enhancement is dependent upon comparison. We can only judge something as being enhanced relative to something else. If you give someone a flintlock and ask me whether their projectile-launching

capacities have been improved, I will want to know whether they previously had anything at all, and if so whether it was a semi-automatic or a slingshot. So enhancements are relative; no matter how capable someone is made by an intervention, we cannot say they have been enhanced unless those capacities are better after the intervention than they were before.

We can also see our third defining point: Enhancement is dependent upon context. We can only call a semi-automatic an enhancement over a flintlock for the purposes of projectile-launching under specific circumstances, namely present day, real-world circumstances. In other circumstances, the reverse might be true. If you have time-travelled back to the 18th century where 9mm rounds for your semi-automatic are impossible to obtain, a semi-automatic should not be considered a projectile-firing enhancement. So we need to consider the circumstances under which an individual lives if we are to enhance them. Enhancement interacts with societal and environmental factors. Perhaps we can improve a wheelchair-bound individual's capacities by placing them in a society where stairs are all replaced by ramps.

From the above, we can put forward the following definition:

> A χ enhancement is an intervention that improves χ capacities, under otherwise relevantly similar circumstances, relative to before the intervention.

For example: If an intervention causes Algernon to be better at swimming than he was, and all other relevant factors are kept practically the same, that intervention is a swimming enhancement. This is a broad definition because enhancement is a broad concept.

Many philosophers writing on this subject reach similar definitions. Sarah Chan and John Harris suggest that an enhancement "improves our functioning", "increases our general capabilities for human flourishing", and "is something of benefit to the individual". However, they rule out certain interventions "of dubious overall benefit", such as "breast or penis augmentation", "the taking of anabolic steroids to increase muscle mass", and "'designer' modifications which are more akin to aesthetic or fashion preferences than to improvements: hair colour, eye colour, or physique" (Chan & Harris 2007, p. 1). Presumably this means that they are unwilling to call an intervention an enhancement unless it clearly provides an overall improvement and benefit.

More recently, Harris gives a definition that is supposed to capture the ordinary usage of the term 'enhancement':

> "In the context of interventions which impact on human functioning, an enhancement is clearly anything that makes a change, a difference for the better." (Harris 2010, ch. 3)

Definitions of enhancement tend to focus on improvement and benefit with regard to humans and the things we are capable of. They mention nothing about the extent of the improvement and benefit, or about the state of the individual's capacities before the enhancement beyond the fact that they were inferior. Perhaps this omission is important. In section 1.1 we will discuss the possibility that definitions of enhancement should refer to a model of 'normal functioning'; the capacities that it is normal for an individual to have.

There might be something else missing from our enhancement definition; what counts as an enhancement is broad, because what counts as an 'improvement' is vague. This is especially so with my definition; while Chan's and Harris's definitions can be taken to include only overall improvement as enhancement, my definition acknowledges that an enhancement in one context may be the opposite of enhancement in another context. For example anyone with a job in a noisy environment, such as a construction worker, may be better at their job if they are unable to hear; the constant loud noise might be stressful and distracting. But, in most contexts of life in human societies, the ability to hear is of great benefit. Therefore a 'construction work' enhancement may arguably include an intervention that causes deafness.

David DeGrazia illustrates how broad enhancement definitions can be:

> "...any deliberate intervention that aims to improve an existing capacity, select for a desired capacity, or create a new capacity in a human being." (DeGrazia 2014, p. 361)

This covers improvement of capacities, but also suggests two other ways that an individual can be enhanced: We can select for physical capacity by causing more individuals with above-average physical capacities to come into existence than would have otherwise. If we could figure out how to cause a human to have some entirely

new capacity that no human currently has, such as a sense that allows them to detect magnetic fields, this might also count as an enhancement. This is why my enhancement definition does not refer specifically to improvement of individuals' capacities. An intervention can result in improved capacities by making it so that the individuals that come into existence have superior or additional capacities, compared with those of individuals who would have otherwise existed. These are unusual cases; enhancement is usually discussed with reference to an individual whose capacities are improved. But keep these other kinds of enhancement in mind, as they will be important in section 3.

Let us look at some actual examples of enhancement. One of the most common forms is cognitive enhancement. There are various cognitive capacities that could be improved, e.g. memory, reasoning, creativity, the content of one's knowledge. Cognitive enhancement is currently practiced almost everywhere in the world, in the form of education systems. In fact, developed countries consider education to provide such a benefit that we force children by law to receive some sort of education, in most circumstances. A normal human who has received education is probably better at cognition than they would have been had they not.

Cognitive enhancement is also achieved by gaining access to more advanced technology. Over history, we have developed increasingly effective methods of recording and sharing information: drawing, writing, radio, audio, video, and the internet. These allow individuals to benefit from the cognitive capacities of others, thereby improving their own. We have also developed powerful computers that perform calculations many times faster than any human brain, allowing individuals to delegate difficult and time-consuming cognitive work away from the brain.

Another common form of enhancement is physical enhancement. Athletes undergo rigorous exercises and commit to specific diets in order to become better than they were at performing whatever physical activity they pursue. Militaries impose intensive training, which makes their recruits better at all kinds of physical tasks. We can also find physical enhancements in the form of tools and equipment: A jackhammer allows an individual to easily break rock and concrete; a strength-enhancing mechanical exoskeleton allows individual hospital staff to safely transport

patients; a nuclear bomb mounted in an aircraft allows an individual to destroy an entire city and kill its population simply by flipping a switch.

Bioenhancements are enhancements produced through biotechnological interventions. Probably the most common form of bioenhancement is the use of medicinal drugs, which can provide a variety of enhancements, from protection against and treatment of diseases, to athletic or intellectual performance enhancement. Bioenhancement also includes genetic technologies, which are largely still in development, but which have already resulted in enhanced non-human lifeforms such as crops that are resistant to environmental conditions and diseases. In the future, genetic technologies will plausibly allow us to manipulate human capacities, improving them beyond that which is normal or adding entirely new ones. Interventions that allow us to select for certain traits in unborn human embryos, such as genetic screening followed by selective implantation or abortion, are also arguably enhancements.

The distinction between enhancement and bioenhancement is an important one. The latter is a subset of the former; all bioenhancements are enhancements, but not all enhancements are bioenhancements. Practice, education, and equipment can all improve capacities, but so can drugs and genetic technologies. We will discuss the specifics of bioenhancement soon. For now, simply note that enhancement is a far broader category of intervention than might be supposed. We do not need to imagine futuristic technology or sophisticated medication to capture the notion of enhancement; a regular morning jog is enough to improve someone's capacities and qualify as enhancement, and simple tools that have been around for hundreds of years qualify just as easily as imagined futuristic technology out of science fiction.

1.2: Well vs. better

Let us return to another distinction mentioned above: the distinction between enhancement and treatment (a.k.a. therapy). It can be argued that, rather than just making humans better, enhancements (properly called) make humans better than normal. If Algernon has liver disease and we treat it by performing a liver transplant, this clearly makes Algernon better at a number of vital biological functions; it improves his metabolic capacities, for instance. But most humans are capable of

those same biological functions without the same procedure. It is only because there was something abnormal about Algernon (i.e. he had a malfunctioning liver) that the procedure made him better, or improved his biological functioning. Maybe we ought not to classify this sort of case as enhancement, and instead call it treatment or therapy.

Some definitions of enhancement have this distinction built-in:

> "...enhancements may aim to affect various non-disease related factors that have a genetic component, including intelligence, talent, strength and height." (Holtug 2011, p. 139)

A position sometimes taken by those who emphasise the enhancement/therapy distinction is that, while therapy is permissible, enhancement is not (Buchanan et al. 2000, p. 106). If you are suffering from a debilitating condition that negatively impacts your wellbeing, there is nothing wrong with seeking treatment for it. But if you are perfectly healthy and want to improve your wellbeing, that is another matter. There might be legitimate reasons why we should not be allowed to take certain measures with the sole objective of unrestricted self-improvement. We will call interventions that should be prohibited 'impermissible'.

However, some argue that prohibiting all enhancements, when defined as improvement beyond healthy norms, is untenable. Such a position implies that humans should never attempt to become better than they normally are.

> "If this is the case, we should never seek to improve ourselves through any means, a position which runs counter to familiar notions of human aspiration and achievement" (Chan and Harris 2007, p. 3).

Humans are constantly improving themselves beyond current norms; the last few millennia have seen extensive improvements to our capacities for learning, treating disease, communication, influencing our environment, and so on. The extreme position, that any improvement beyond healthy or species-typical norms is impermissible, seems implausible.

Another position is that treatments are interventions we should justly expect to be available if we need them, whereas we should have no such expectations about enhancements (Buchanan et al. 2000, p. 108). Most citizens of developed societies

expect treatments for common injuries, illnesses, and medical conditions. Governments are commonly thought to have an obligation to ensure access to a certain standard of medical treatment for the governed population. We will call interventions that citizens should generally have access to 'obligatory'.

Most biotechnological interventions commonly classified as enhancements arguably do not fall within the kinds of interventions we can expect to be guaranteed to us. In New Zealand, it would be widely objectionable if we could generally not expect to receive treatment for common conditions like broken bones, infections, and appendicitis. But people generally do not feel that justice is violated by the lack of unconditional access to memory-enhancing drugs, steroids, or genetic therapy to be made taller, even though these interventions could be of benefit to many healthy people.

The issues raised by the enhancement/therapy distinction seem mainly to be relevant to cases of bioenhancement; few would argue that there is anything immoral or problematic about training, practice, or education, which are all forms of enhancement rather than bioenhancement[1]. The interesting problems only arise regarding biomedical interventions like drugs or genetic technologies. Sending a child to school for cognitive enhancement is usually considered permissible, obligatory, and often mandatory, but feeding them pills meant to make them smart enough to attend university by age ten is a different matter.

As is implied by the above paragraphs, positions that draw distinctions between enhancement and therapy often rest on the idea of there being some model of 'normal functioning' for humans (Ibid., p. 126). The enhancement/therapy distinction is used to formulate medical legislation such as insurance policies, and guides common intuition about which medical procedures are permissible or should be obligatory (Ibid., pp. 110-112). In spite of this, it is common to discount the importance of the therapy vs. enhancement distinction, for several reasons.

[1] This further undermines the claim that all enhancements of healthy individuals are impermissible.

First, the distinction requires sufficient understanding of what 'normal' means for a human that we can tell which interventions cross the line from therapy to enhancement, and it is not clear that we have such an understanding.

Two individuals can be in the same physical condition for different reasons. For instance, the case of Johnny and Billy (Ibid., p. 115): Johnny will grow up much shorter than average because he is human growth hormone-deficient, due to a brain tumour. Billy has completely normal health, but will grow up to be no taller than Johnny simply because he inherited genes from his exceptionally short parents. The normal functioning-based enhancement/therapy distinction might justify or demand providing Johnny with HGH supplements, but not Billy, simply because Johnny does not have 'normal' health by some definitions while Billy does.

But some have the intuition that both cases are relevantly similar. If correct, this would reveal a flaw in any reasoning that advocates handling the cases differently. Perhaps it is just societal factors that identify Johnny as having a disease but not Billy (Ibid., p. 118). There are biological factors that explain Billy's shortness just as well as Johnny's tumour explains his. The society they live in has come to think of the genetics resulting in shortness as normal, but of tumours as abnormal and therefore deserving of treatment.

A related problem is that it is normal for humans to be at risk of many harmful conditions that we typically think of as diseases, and deserving of obligatory treatment (Chan & Harris 2007, p. 1). Treating diseases arguably qualifies as enhancement under the normal functioning distinction, because it is normal for a human to be susceptible to them. But if we are to judge interventions to be impermissible on the basis that they are enhancements rather than treatments, this could imply that a large number of medical interventions that save countless lives every year are enhancements, and should perhaps be impermissible or non-obligatory.

The above kinds of reasoning have led many to believe that the normal or species-typical functioning is a poor way to distinguish between enhancement and treatment, and to determine which interventions should be permissible or obligatory (Harris 2010, ch. 3).

Buchanan claims that the above arguments are flawed ways of thinking about the enhancement/therapy distinction:

> "The point that emerges here is that the natural baseline has no metaphysical importance: It is not that we must pay some special respect to what is natural, for example, by maintaining or restoring it. Rather, the natural baseline has become a focal point for convergence in our public conception of what we owe each other by way of medical assistance or health care protection. To develop fair terms of cooperation, we should not have to resolve our disputes about these comprehensive moral views, nor should we have to settle an abstract issue in the philosophy of biology." (Buchanan et al. 2000, p. 151)

We should not worry about whether we establish what is normal with complete reliability in borderline cases, because the distinction itself is less important than the benefits it provides. We can still use the normal functioning-based distinction as an intuitive guide to the interventions we should expect to be available.

But even if we can draw the distinction reliably enough, it might be objected that the moral grounds it provides are not strong enough to justify typical judgements and legislation. In other words, the enhancement/therapy distinction is not an exact guide to distinguishing between permissible and impermissible interventions, or obligatory and non-obligatory ones, on moral grounds (Buchanan et al. 2000, p. 118).

A plausible position seems to be the following: If there is a legitimate enhancement/therapy distinction, then the most it can do is prevent enhancements from being obligatory. Arguments such as Harris's convince me that we have no reason to consider enhancements to be necessarily impermissible. Moreover, the enhancement/therapy distinction is arguably not legitimate at all, in which case we cannot even use it to determine which interventions are obligatory in the way Buchanan thinks we can.

But the most problematic part of the enhancement/therapy distinction is its basis on 'normal functioning'; there are too many cases of non-normal functioning that are now taken as standard health, and cases of normal functioning in which we tend to think something is wrong and requires treatment. The less-restrictive 'enhancement

as improvement' model that I suggested in 1.1, supported by Harris, Chan, and DeGrazia, avoids most of these problems.

We will put this issue aside until we have a better understanding of how it applies to the MB discussion. We will first need to know more about the specifics of moral enhancement.

1.3: Moral enhancement

We have established a definition of enhancement based on improvement of capacities, leading us to the following: If an χ enhancement improves χ capacities, a moral enhancement improves moral capacities.

How should we understand what it means to improve moral capacities? Morality is a controversial topic, as we will see in section 2. But for now we just need a working definition of morality to see if it's even sensible to talk about enhancing it. The first explanation of morality that comes to my mind, and the way that people commonly explain morality to children, is: 'knowing right from wrong'. It seems that things that are 'right' are morally good, and things that are 'wrong' are morally bad. It also seems that, out of two things that are both right, or both wrong, one can be morally better or morally worse than another. Perhaps someone with improved moral capacities is morally better (more morally good) than they were, or would have otherwise been, whatever that means.

For now we will suppose that morality involves the making of moral judgments; judging that some state of affairs (e.g. circumstance, action, person, attitude, etc.) is morally good, bad, or better/worse than some other. I might make the following moral judgments: that selflessly donating to Red Cross is morally good, that torturing a kitten for fun is morally bad, and that torturing a kitten for fun is not as morally bad as torturing a human infant in front of its parents for fun. What justifies these moral judgments? What constitutes a correct moral judgment? We will return to these questions in section 2. But the above statements probably make at least some sense even if you cannot articulate precisely why, or do not agree that they are correct.

Morality seems to involve more than just making moral judgments. Acting in accordance with them is also important.

Imagine that Algernon spends excess income on cartons of cigarettes, which he gives out to children, and intends to continue doing so. However, he judges that it would be morally better to donate that money to Red Cross, and that it is morally bad to provide children with cigarettes. It is not clear that we should call Algernon morally good; at least, not as morally good as someone who makes the same moral judgements and refrains from giving cigarettes to children.

Consider a different situation: Bertha donates all of her spare income to Red Cross, volunteers at homeless shelters, minimises her carbon emissions, and is a strict vegan. But when asked whether she has moral reasons behind any of this behaviour, she denies that she does. In fact, she judges that helping those in need only prevents humanity's weakest links from dying off as they should, and that she has no problem with exploitation of animals and the environment in order to promote her personal interests. However, she is convinced that there is an assassin watching her every move who will kill her if she does not act as she does. Assuming we believe her, it is not clear we should call Bertha morally good either; at least, not as morally good as someone who behaves in the exact same ways but for reasons that reflect correct moral judgments they have made.

From Algernon and Bertha it seems that (a) correct moral judgments and (b) actions taken in accordance with them are both important factors when it comes to being morally good, or one person being morally better than another. A moral enhancement will at least sometimes have to act upon both of these factors in order to consistently make individuals morally better. Some individuals may become morally better if only one of the two factors is modified. Algernon, for example, may only need to be made more likely to act upon his generally correct moral judgements. But if someone who tends to make obviously mistaken moral judgements, such as 'torturing kittens for fun is morally right', is given such an enhancement, they will become more likely to act upon those mistaken judgements and start torturing kittens for fun more frequently. So both factors, (a) and (b), are important to an individual's moral capacities.

Where does this leave moral enhancement? Considering the above, a plausible definition is as follows:

> A morally enhanced person makes more correct moral judgments than they did before the enhancement and/or acts in accordance with them more consistently.

Thomas Douglas, an early contemporary advocate of moral enhancement, provides a specific definition that we can compare ours to:

> "A person morally enhances herself if she alters herself in a way that may reasonably be expected to result in her having morally better future motives, taken in sum, than she would otherwise have had." (Douglas 2008, p. 229)

Moral judgments are implied by the term 'morally better'; at the very least we will need some way to determine which motives are morally better than others, and one way to do this is to appeal to moral judgements . 'Motives' implies that these judgments are to be acted upon, or at least contribute to action. Therefore, this definition seems to be in agreement with our working definition. Other definitions similarly emphasise the importance of motivation to morality, as opposed to either judgments or actions on their own (Sparrow 2014, p. 24). Both are important to our moral capacities.

No moral enhancement definition so far has made explicit mention of what specifically it means for something to be morally better than it was previously. Therefore, we need ways to determine which moral judgements are correct, or otherwise ways to determine which things are morally better than which other things. This concept, I suggest, is best thought of as 'moral value'. We will discuss moral value in section 2.

Is our working definition of moral enhancement coherent? Could an intervention conceivably result in more correct moral judgments and more motivation to act in accordance with them? I think so. In fact this kind of moral enhancement is something that we apparently do all the time. Societies go to great pains propagating what are conceived of as correct moral judgments. We place a lot of emphasis on parents teaching their children 'the difference between right and wrong', and media for children is rife with various moral messages designed to assist with this. Sharing is good, stealing is bad. Helping people is good, harming them is bad. Almost every

story, movie, or cartoon aimed at children ends with a "moral", a message that one is meant to take away from the story.

As adults, 'right and wrong' are major themes in almost every area of life. Leaders appeal to morality when justifying wars and policies. Religions, upon which are spent a huge amount of time, devoted effort, and resources, are arguably so popular at least in part because they purport to offer a simple, reliable guide for how to be morally good. People make decisions with significant consequences based, not upon those consequences, but upon moral judgments; some people give huge amounts of money to people they will never meet, and many others pass up countless opportunities to steal from or take advantage of strangers even when there is no chance of being caught.

We also have ways to encourage people to consistently act in accordance with these moral judgments. Schools get involved, punishing students for reasons that often seem to involve the claim that the behaviour was in some way immoral: if students are violent or disrespectful they are assigned unpleasant tasks or detention to waste their free time, to discourage them from such behaviour in the future. We also have laws against many immoral acts. Notably, harmful acts in a variety of forms are prohibited by many laws. You can be sent to prison if you perform these acts, where you will undergo 'rehabilitation' or 'correction'. Since inmates are not physically or mentally debilitated, this term implies that the inmate is deficient in some other sense that makes them unfit to live within normal society. Perhaps this deficiency is a moral one. That is, they need to be made better at morality; at correctly making and consistently acting in accordance with moral judgments.

But maybe there is a better term than 'moral enhancement' to describe these societal practices; remember the distinction between enhancement and therapy. Perhaps they do not attempt to make people straightforwardly morally better, but to treat cases of moral deficiency. Perhaps there is a normal level of moral functioning, and any who fall below this level have the equivalent of a disease or debilitating condition. The messages in media, the correctional discipline in schools and prisons, and so forth, perhaps this is all better classified as moral therapy. This disregards the 'improvement' model of enhancement, according to which these cases do count as moral enhancement after all.

But it seems as though there are people with sufficient moral capacities that we do not force any of the above measures upon them but who still take actions in order to become morally better. Support groups for addicts of various kinds (alcohol, drugs, gambling, etc.) are available and popular; people voluntarily attend them in order to change something about their own behaviour that they see as immoral, or at least harmful or objectionable in some way. Some people spend free time discussing or educating themselves on moral issues, and many attend meetings of religious institutions and ask for moral advice from the respected figures that run them. So it seems as though there are at least some, and perhaps many, examples of morally normal people further improving their moral capacities, which is arguably moral enhancement even under the normal functioning-based enhancement/therapy distinction. The above kinds of moral enhancement techniques, those which do not involve biotechnology, will be referred to as traditional moral enhancement (TME) from here onwards.

1.4: The prospect of moral bioenhancement

We have looked at general enhancement and moral enhancement, and briefly touched on bioenhancement. We will now examine the latter more closely, so that we may begin to tie these three notions together.

Drug-based physical bioenhancements are commonly used by athletes or people in other physically-demanding professions. Lean mass builders such as anabolic steroids and human growth hormone can increase growth of muscle, and blood doping agents can increase the amount of oxygen carried by the bloodstream leading to improved endurance and stamina (Wallace 2009). These interventions provide such effective improvements to physical capacities that official athletic competitions routinely prohibit their use by contestants (Ibid.). Drugs can also enhance mental capacities. Ritalin and Adderall are drugs known to improve attention and concentration (especially in those with learning problems such as ADHD), heighten neural activation, and decrease the effects of fatigue (Dubljević 2013, p. 26). Some university students have reported that they misuse prescription stimulants such as Ritalin and Adderall in order to improve their academic performance, or enhance alertness or energy (Bossaer et al. 2013, p. 3). A range of other drugs have the

potential to improve memory functions, or affect mental states in ways that improve focus, self-control, or productivity (Persson & Savulescu 2008, pp. 164-165).

Bioenhancement can also take the form of interventions that target genes. Genetic interventions range from primitive selective breeding to embryonic selection to genetic modification. This can involve, rather than improving the capacities of existing individuals, making it so that the individuals who end up existing have improved capacities relative to those of individuals who might otherwise have existed. Although genetic interventions are not currently commonplace among humans, there are many examples in non-humans that suggest they may be in the future.

Animals and crops have been selectively bred since ancient times so as to encourage certain desirable traits to be passed on to future generations, and such techniques are still used today. Selective breeding among humans has and does take place in various forms, the Nazi regime being a famous example, but historically such attempts have involved great harms and today the practice is widely considered to be morally wrong.

Tobacco plants genetically modified to be resistant to viruses were commercialized in China as early as 1992 (James 1997, p. 13), and mice have been genetically modified so that they overproduce certain receptors in the brain, resulting in improved memory (Persson & Savulescu 2008, p. 164). The scope of possible benefits to humans of similar genetic modification seems vast, though of course there is a lot of research to be done before we can be sure they will provide definite enhancements without causing problems.

From section 1.3, it seems that the concept of moral enhancement is not incoherent; it is something societies arguably already achieve through moral education. It also seems that medicinal drugs and genetic interventions can have effects that improve human capacities, thereby constituting bioenhancements, and these are not the only forms bioenhancement can take. Therefore, the prospect of moral bioenhancement (MB) so far seems plausible; it is neither logically incoherent nor practically impossible. But do we have any reasons for thinking our moral capacities can be bioenhanced in similar ways that our physical and cognitive ones can? I can see two main reasons.

First, TME is a product of interventions that act on psychological factors; we can change someone's mind about moral matters by supplying them with reasons to act differently, by modelling morally desirable behaviour, and by arousing their emotional responses (Harris 2011, p. 104). The biotechnological interventions at our disposal include drugs. Many drugs are well-known to have profound influences on the psychology of individuals, such as those used to treat depression and anxiety. If the effectiveness of TME consists in the psychological changes it causes in individuals, and drugs or other biotechnological interventions can cause those same psychological changes, it is possible that MB would operate through the same effects as TME (Kahane & Savulescu 2015, p. 137); in this case, if TME is truly effective, MB would be also. That is unless there is something missing from MB but not TME, without which true moral enhancement is impossible. We will address this question in section 4.

Second, we have reasons to think that moral capacities can be linked to biological factors. Moral psychology is not exclusively influenced by societal and cultural forces, such as conforming to norms and understanding of moral reasons. Although such influences are also important, we have reasons to believe that we are innately predisposed to think morally. We can see evidence of this both through an understanding of their at least partly evolutionary origins (which we will discuss in section 2), and in scientific studies performed on humans and other animals (Persson & Savulescu 2012, pp. 109-112). If we can link morally relevant psychology to biological factors, such as genes and innate mental processes, and we can modify these biological factors through biotechnological interventions, which we can, then it follows that we can modify morally relevant psychology through biotechnological interventions, potentially resulting in improved moral capacities.

In section 3 we will look at some prospects for specific interventions that might be used to achieve MB. For now, we will move on to a discussion of why doing so might be a good idea.

1.5: Ultimate Harm

Ingmar Persson and Julian Savulescu have authored a prominent book on the prospect of moral bioenhancement, regarding how it may be essential if sentient life

on planet Earth is to remain possible in the long-term. The advancement of science and technology is responsible for many significant improvements to human capacities. In *Unfit for the Future*, Persson and Savulescu (P&S) argue that, while greatly beneficial to us, these enhancements also provide us with the power to bring about enormous amounts of harm; enough harm to effectively end the world:

> "During the last century our power to harm reached the point at which we can cause what might be called *Ultimate Harm*, which consists in making worthwhile life *forever* impossible on this planet." (Persson & Savulescu, 2012, p. 46)

P&S argue that, as technology advances, and our power and population increases correspondingly, it becomes increasingly likely that said power will be used to bring about Ultimate Harm (UH) or that UH will be inadvertently caused. This is largely because typical human morality, "common-sense morality", makes us:

> "...ill-equipped to cope with the moral problems generated by the advanced scientific technology, over-population, and the globalization of the modern world" (Persson and Savulescu 2012, p. 12).

Why do scientific and technological advancement interact with common-sense morality in such a potentially catastrophic way?

One key reason given by P&S is that moral enhancement has not kept up with cognitive enhancement and the scientific advancement of technology (Persson and Savulescu 2012, p. 106). Although science and technology have advanced exponentially, drastically changing the conditions under which we live and the power we have over each other and our environment, our moral capacities remain mostly unchanged from the way they were during our distant evolutionary past.

Common-sense morality evolved as traits that were adaptive during that ancient time-period, before airplanes and global communications and nuclear bombs. Moreover, technological advancements have allowed the human population to grow exponentially, far beyond the size it was during the period in which our moral capacities evolved. It seems unlikely that we can presently rely upon those same moral capacities to influence our actions in ways that are likely to solve the problems that come with our newfound scientific and technological power and enormous population, or prevent us from misusing it.

For simplicity's sake, I will use the term 'scientific progress' to refer to the kind of combination of cognitive enhancement, cognitive bioenhancement, or technological advancement that P&S claim exacerbates the risk of UH through its interaction with deficient moral capacities. 'Scientific progress' also encompasses the expanded human population and quicker exploitation of natural resources that it facilitates.

An important part of P&S's reasoning is their argument that it is easier to harm than to benefit, and more specifically easier to cause death and suffering than to prevent it (Persson and Savulescu 2012, p. 14). In the course of most people's daily routines, there are frequent opportunities to harm and kill dozens of people quite easily, for example by driving a car onto a crowded footpath. Opportunities to save or benefit nearly as many lives are few and far between. There are many obvious ways to disrupt a well-functioning system (such as an organism or society) in order to harm it. But to benefit a well-functioning system we need to improve upon it, which is a much more difficult task than disrupting it (Ibid., p. 48). This means it is easier to kill an organism (arguably a great harm as doing so prevents any future benefits it might enjoy) than to save the life of an organism. This is why, although scientific progress brings with it both the potential to harm and the potential to benefit, further advancement increases humans' capacity to harm more than to benefit; it is easier to harm than to benefit.

P&S claim that common-sense morality is good enough when small groups of self-sufficient humans co-exist, but is not good enough for when a vast network of completely inter-dependent humans co-exist (Ibid, pp. 19-20). They point to several features of common-sense morality that contribute to moral deficiency

First, we have a tendency to think of moral responsibility as "causally-based"; we consider ourselves to be less responsible for harms that we allow to occur by omitting action than those we actively cause. This also means our sense of responsibility is "proportionally diluted" when the actions of many contribute to the same harm. Second, we have some sort of biases towards effects our actions have on those who are near to us either in location or in time; generally we care less about harm inflicted upon those far away from us and those in the future. Third, we are susceptible to "scope-insensitivity"; we are not more morally sensitive to harm done to large numbers of people as opposed to small numbers. (Ibid., pp. 24-30)

We will look more closely at the evolutionary explanations behind human moral capacities, and why they are deficient for promoting the right sort of cooperative behaviour for solving problems created by advanced scientific progress, in section 2.

P&S suggest two primary examples of possible causes of Ultimate Harm.

The first is the threat posed by weapons of mass destruction (WMDs), e.g. nuclear bombs and bio-weapons (Ibid., p. 46-49), used in terrorist attacks, wars of aggression, and other large-scale acts of violence. In the past, such acts have cost thousands of lives; one of the worst terrorist attacks, the 2001 September 11 attacks on the World Trade Centre in the USA, killed nearly 3000 people (IAGS 2003). Bombings of the Japanese cities Hiroshima and Nagasaki during World War II killed between approximately 150,000 and 246,000 people (RERF 2007) using relatively unsophisticated nuclear weapons.

If similar violence take place presently or in the future, it is likely to utilise technologically advanced WMDs capable of causing far more death, destruction, and suffering; potentially enough to make future worthwhile life impossible. And, as scientific progress advances, it becomes easier to obtain such harmful technology. Even today it is plausible that well-organised terrorist groups may have access to it (Persson & Savulescu 2012, p. 47). If morally depraved individuals gain access to such powerful and harmful technology, their deficient moral capacities will not prevent them from using it to cause UH.

The second suggested UH possibility is the catastrophic degradation of the environment and ecosystem of planet Earth, through overexploitation of its natural resources; a primary example of this is anthropogenic global climate change (global warming). The effects of climate change (e.g. rising sea levels, ecosystem devastation) could one day grow to such catastrophic proportions that UH is caused. P&S detail the kinds of large-scale cooperative measures that would need to be taken by citizens and governments of affluent societies if we are to address problems like climate change, and argue that features of deficient common-sense morality make it unlikely that such measures will be taken (Ibid., pp. 73-85).

Climate change is not caused by the exceptionally immoral actions of a morally depraved individual or small group, like potential WMD attacks would be. It is caused by moral deficiency of large groups; entire human societies and nations. But

it is also linked to scientific progress and moral deficiency, and so could theoretically be addressed by successful moral bioenhancement.

We should take the above two kinds of potential UH cases as representative of two distinct kinds of moral deficiency in humans. The first, suggested by the weapons of mass destruction case, is that there are a few morally depraved individuals with such deficient moral capacities that, if they obtain the power to cause UH, they will do so. Call this Individual Moral Deficiency (IMD). The second, suggested by the climate change case, is that the typical moral capacities of humans are not sufficient to address moral problems that threaten UH and require large-scale cooperative action to solve. Call this Mass Moral Deficiency (MMD). Keep in mind that while both IMD and MMD have related causal factors (i.e. scientific progress, large human population, and deficient moral capacities), they are distinct.

We will put aside the tasks of critiquing P&S's claims of moral deficiency and their arguments regarding UH, because I am interested in what the best course of action would be if our situation truly is as bad as they claim. Regardless, I think that the fundamental point is plausible: if evolution has largely shaped modern human moral capacities, this genealogy does not justify thinking that they will be adaptive in our modern environment in which scientific progress has advanced so far. Even if we disregard P&S's more technical arguments, it would not be controversial to claim that the moral capacities of most humans are not motivating the large-scale cooperative action needed to address modern problems rooted in scientific progress. It would be even more uncontroversial to claim that a tiny minority of humans cause a huge majority of harm, and that this can be blamed on deficient moral capacities.

How specifically will MB help to prevent UH? P&S suggest that MB could enhance certain traits relevant to our moral capacities, such as altruism and sense of fairness (Persson & Savulescu 2008, p. 172). This would result in a greater ratio of "moral people to immoral people" (Ibid., p. 163). I take it that this has two effects: (1) The more individuals enhanced, the less likely it becomes for technology that could cause UH to fall into the hands of someone with moral capabilities deficient enough to use it for this purpose. (2) Enhanced populations will be generally more moral, motivating solutions to problems (such as climate change) that can best be addressed by morally motivated large-scale cooperative action. In other words, moral

bioenhancement could potentially address both IMD and MMD by making people morally better.

1.6: Widespread Harm

The present focus on Ultimate Harm neglects a related category of problems faced by modern humans. I agree that UH would be bad and should be prevented, but there already exists a more tangible kind of harm that is just as relevant to scientific progress and moral deficiency as UH is.

This second category I will refer to as 'Widespread Harm'; systematic harm that affects, or is likely in the future to affect, large numbers of individuals negatively. Widespread Harm (WH) is and will likely be responsible for a huge amount of suffering, even if the effects can be avoided by many and even if it will not make worthwhile life forever impossible. Some of the causes contributing to WH have always existed in one form or another. But modern population growth and technological advancements have the potential to exacerbate or systematically perpetuate them, especially if our moral capacities remain as they are. Importantly, these causes are problems that do not appear to be solvable by individuals acting alone. They require large-scale cooperation if they are to be addressed.

> Ultimate Harm (UH): Events causing enough harm that worthwhile life becomes impossible on Earth forever, causally related to scientific progress and human moral deficiency.
>
> Widespread Harm (WH): Events contributing to large-scale harm inflicted on human and non-human life, causally related to scientific progress and human moral deficiency.

We will first look at three examples of WH and consider how moral enhancement could help to address each one. We will then say more about the distinction between UH and WH.

Our first example of WH is the 'absolute poverty' in which many humans on Earth live. At least 1.2 billion people today live in this condition:

> "...characterized by severe deprivation of basic human needs, including food, safe drinking water, sanitation facilities, health, shelter, education and information." (United Nations 1995, chapter II paragraph 19)

With modern means at our disposal, it would be relatively easy to redistribute resources so that no one has to live in danger of starving or dying from easily preventable conditions. But doing so would require the altruistic cooperation of affluent societies, in the form of measures such as donations and social reform, because such societies do not have sufficient self-interested motivation. P&S themselves note extreme poverty as a modern problem contributed to by scientific progress and moral deficiency (Persson & Savulescu 2012, pp. 62-63). Not only does scientific progress provide developed nations with the means to exploit millions of people, but it also provides them with the means to end that exploitation and alleviate their suffering. Extreme poverty is a problem caused by the same MMD that threatens to cause UH.

Consider the interests of some wealthy citizen of a developed nation, Chris. It is plausible that the suffering and deaths of people who Chris will never meet does not have any effect on his quality of life. Unless the knowledge of that harm causes Chris psychological harm, the only reasons Chris apparently has for addressing extreme poverty are non-self-interested reasons like altruism or a desire to uphold human rights of some kind; the kinds of reasons we tend to think of as concerning moral good. Extreme poverty is an enormously harmful problem that could be solved if enough of the right people took the right actions, and is therefore a problem that requires large-scale cooperative action to address it.

Our second example of WH is the compromise of the environment, ecosystem, and interests of non-human life. This WH-issue is related to the possibility of UH through environmental degradation suggested by P&S, but is broader in scope. Humans have become more conscientious in recent years about our effects on the environment and the non-human species that inhabit it. But over-consumption of natural resources, pollution, habitat destruction, and the general disregarding of animal interests are still prevalent:

> "...during the past 150 years, humans have directly impacted and altered close to 47% of the global land area." (Wilson 2002)

The influence of humans has resulted in an extreme rate of extinction; at the current rate, it will only take half of one human's lifespan for half of the world's remaining species to become extinct (Ibid.).

It is not clear that people like Chris have no reason to address this issue; we all live in the environment, and its degradation is likely to have such far-reaching effects that they will eventually impact Chris's life in some way. Even so, advanced technology makes it likely that a substantial amount of human life will be able to persist even as non-human life is steadily eradicated. Most of the effects of environmental degradation can be mitigated through the same scientific progress that contributes to it, though only for those who have access to the right technology. P&S give global climate change as an example of impending UH, meaning that they predict it could result in total extinction of all humans and other worthwhile life. This might be a difficult claim to confirm. But the claim that climate change will cause WH, through raising sea levels, causing extreme weather, and radically changing our ecosystems, is almost incontestable.

Our third example of WH is the looming prospect of an antibiotic-resistant pathogen pandemic. Infection-causing bacteria more rapidly acquire resistance to antibiotics the more frequently those antibiotics are used. This is because of their fast generational turnover and high rate of mutation. An estimated 70% of bacterial infections in the US have acquired some antibiotic resistance. Many strains of the bacterium staphylococcus aureus, for example, are now resistant to several common antibiotics and have begun developing resistance to our most powerful "last resort" antibiotics. Major contributors to the emergence of resistant bacteria are our farming industries, particularly factory farming of animals for meat; antibiotics are liberally and indiscriminately used to keep animals as healthy as possible when kept in crowded conditions of high infection-risk. (GRACE Communications Foundation 2015)

Practices like factory farming, enabled by scientific progress, exacerbate the risk that bacteria harmful to humans and immune to all our antibiotics will emerge. And even without factory farming it is possible that we will eventually need new antibiotics to treat the constantly-mutating pathogens that threaten massive harm. If more attention is not soon brought to this issue an untreatable pandemic may strike the

world and result in WH, as large numbers of individuals die from contagious diseases for which there are no cures. But it is not clear that those responsible for funding research towards a solution have sufficient self-interested justification to do so. Scientific progress could be used to address this problem if put to altruistic use, but instead exacerbates it.

First, note that WH has the same kinds of causes as the potential causes of UH put forward by P&S. Most notably: WH is caused and exacerbated largely by human moral deficiency and the advancement of scientific progress, just as is the increasing risk of UH. UH and WH are related; some WH issues contribute to the risk of, or are lesser versions of, UH issues. And all UH seems to qualify as WH; if it is to potentially make all worthwhile life on Earth impossible forever, such an event would probably have to involve large-scale harm.

UH and WH are both related to human moral deficiency. I distinguished between two types: Individual moral deficiency (IMD) and mass moral deficiency (MMD). Although P&S are concerned about the risk of IMD-caused UH, because of intentionally harmful acts such as nuclear terrorism, MMD is the main cause of all three WH examples above. Extreme poverty, the risk of untreatable pandemics, and the compromise of the environment and non-human life are all examples of problems created through scientific progress, but which could also be solved through it. The actions required to address WH, however, involve large-scale cooperation which typical human moral capacities do not motivate. If MB can address WH as well as UH, we would have another good reason to consider its development and implementation.

Note that IMD also arguably contributes to WH issues, and to our failure to address them. A few immoral politicians may knowingly influence legislation and popular opinion to benefit their own interests, rather than the interests of those currently or soon to be affected by WH. For example, by discrediting research that indicates climate change is a real problem. If this is the case, then we can expect any MB that works to address IMD-caused UH, such as terrorism, to also address the contribution IMD makes to WH.

To sum up: P&S argue that the objective of addressing UH, which is caused by moral deficiency and scientific progress, provides reason to develop and implement moral

bioenhancement; making people morally better will make it less likely that scientific progress will catastrophically interact with human moral deficiency. I add the claim that any MB that addresses UH will have the additional benefit of addressing WH, because WH is also caused by moral deficiency and scientific progress. This will increase the quality of life of many humans and members of other species, and prevent substantial harm caused through death and suffering. Therefore, assuming that MB will effectively address UH and WH, we have strong reasons for considering its development and use.

For the remainder of this thesis, I will use the term 'Harm' (with a capital 'H') to refer to all states of affairs that qualify as either UH or WH. I will take it for granted that if we have reasons to address UH, we also have reasons to address WH; the objective of MB should be to address all Harm, not just UH. However, broadening the focus of the MB discussion to include WH has an additional benefit: UH is only a possibility, while WH is presently occurring and seems likely to continue occurring without some sort of intervention. Although we are assuming that P&S's arguments are correct, they can at best show that there is a small chance of UH occurring. This leaves advocates of MB open to the objection that UH does not have a high enough probability to justify the measures proposed. WH does not draw this objection.

1.7: Well vs. better, reprise

Now that we know a bit more about the specifics of the moral bioenhancement discussion, we will briefly conclude on the discussion of the enhancement/therapy distinction in section 1.2. I claim that, even if there is a meaningful distinction between enhancement and therapy based on normal functioning, the kind of intervention called for to prevent Harm is not therapy. Successful moral therapy may be enough to prevent IMD-related UH, such as nuclear bombs deployed by morally depraved terrorists, but MMD-related UH and most WH would require moral enhancement.

Moral therapy, defined according to the above distinction, requires some account of moral 'normal-functioning' for our moral capacities. However, 'normal' moral capacities do not seem like they will accomplish the objectives of MB. It might be argued that the kinds of morally depraved individuals inclined to cause UH through

misuse of technology are only in need of moral therapy. But if we are to address the large-scale cooperative action problems currently causing and threatening to cause Harm, we will need the moral capacities of humans in general to improve beyond moral 'normal functioning'. So according to the distinction based upon normal functioning, at least, addressing MMD-related Harm will require enhancement, not therapy. Therefore, at least one half of P&S's UH warning, MMD, calls for moral enhancement and not just moral therapy. And remember that MMD is also a primary cause of WH.

If we use the enhancement as improvement model instead, we avoid the problems related to normal functioning and can straightforwardly define moral enhancement as simply the improvement of moral capacities. It could be objected that the enhancement as improvement model is too broad, as it categorises routine therapy as enhancement. But this counter-intuitiveness could be worth bearing in order to construct a sensible model of moral enhancement.

Regardless of the above, I think it will be more helpful to consider the specifics of proposed MB rather than relying on a controversial distinction that might not apply in this ill-understood area of the enhancement debate. We will have to make judgments about MB that are more sophisticated than simply whether it is better classified as an enhancement or a therapy, if we are to determine whether it can or should be done.

If you are committed to a view that places restrictions on enhancement but not on therapy, purely in principle, you ought also to place those restrictions on the kind of interventions proposed by P&S and in this thesis. If you think there is instead something about enhancement that causes problems because of the specific circumstances, you should wait and see what kinds of circumstances are likely to arise when MB is involved before ruling it out. Classifying MB as a therapy may have legal or political implications, but it is hard to see how it could call for restrictions that would not apply if it were an enhancement. Therefore, from this point I will take MB to involve moral enhancement, and assume that, whatever moral therapy might be, it is not what is under discussion here.

1.8: Thinking about Harm the right way

It is worth going over the ways in which P&S's arguments are questionable. Not only may their dire warnings of UH be mistaken in some way, but the measures they advocate in response may be objectionable.

So far we have gained an understanding of why MB could be desirable: (1) Scientific progress and (2) human moral capacities together cause WH, and increasingly threaten to cause UH. These factors are linked in ways that can be destructive, exacerbating the risk and prevalence of Harm. We can reduce and prevent Harm through interventions that affect (1) or (2).

We have discussed how MB could affect (2), providing humans with the improved moral capacities necessary to mitigate the problems caused by the first. This measure is obviously advocated by P&S. However, they also prominently discuss the benefits of directly affecting (1). The risk of UH increases in proportion to the scientific progress.

> "…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." (Persson & Savulescu 2011, p. 441)

P&S have been known to imply that we should therefore consider postponing scientific progress until we have achieved a sufficient level of moral enhancement that humans are not too morally deficient to make use of it for greater overall benefit than harm. Once moral enhancement has caught up with scientific progress, we can safely proceed with other advancement. P&S are less focused on this point in their more recent work, but it is important to understand why the need for MB might also entail such measures.

While unpalatable, this position makes some sense. The more progress we make, the more we increase the likelihood that someone smart enough, or with access to enough advanced technology, will also be immoral enough to put those smarts or

that technology to use in order to cause UH. In addition, the greater the human population and technologically-enhanced standard of living becomes, the more difficult it will be to address climate change and WH issues. We end up with the troubling and counter-intuitive implication that further scientific progress will primarily increase the risk and prevalence of Harm. Perhaps we must first achieve sufficient moral enhancement in order to make overall beneficial use of further scientific progress.

Elizabeth Fenton identifies a problem with the above implication. We could grant that we would be better off postponing scientific progress until moral bioenhancement has been developed and implemented. But the development of MB will almost certainly require significant scientific progress (Fenton 2010, p. 148). Slowing scientific progress could make the development of MB much more difficult than it would otherwise have been. If this means it takes longer for MB to become effective, the whole point of slowing scientific progress in the first place (i.e. to prevent the risk of UH from becoming too great) would be undermined.

A related problem: cognitive enhancement, which exacerbates the risk of UH, may in itself constitute moral enhancement (Fenton 2010, p. 150; Pinker 2011, pp. 650-656). The line between cognitive enhancement and moral enhancement is indistinct. Many psychological factors plausibly involved with moral enhancement are also involved with non-moral matters. For example, if, as is commonly believed, moral capacities involve the ability to reason which of two alternatives is the morally better action, moral enhancement would plausibly involve improving our reasoning abilities. But it is unlikely that we could improve moral reasoning without also improving other reasoning-related cognitive capacities such as mathematical ability. Therefore, it is arguable that postponing cognitive enhancement while advancing moral enhancement is contradictory.

Fenton also argues that trading scientific progress for moral enhancement, even if it could be done, would not be worth it (Fenton 2010, pp. 150-151). Scientific progress is likely to produce solutions to problems that threaten UH. It is partly responsible for many of these problems, and MMD is responsible for our failure to address them so far, but our best chance of finding solutions may be through further scientific

progress. We have already invented nuclear weapons; limiting scientific progress will not stop them from being available.

Promoting scientific progress, however, may provide us with ways to detect and prevent the deployment of nuclear weapons. Limiting scientific progress will not reverse the adaptations that make bacteria resistant to our antibiotics, but it may provide us with new antibiotics to which they are not resistant[2]. This is to say nothing of the problems that are not caused by scientific progress but which may be addressed by it (e.g. the risk of natural and cosmic disasters) (Ibid.). Whatever the risk posed by further scientific progress, we would need a strong argument to be justified in thinking that it is greater than the risks posed by neglecting scientific progress.

Many of the above points have been echoed by others in the literature (Harris 2011; Carter & Gordon 2015).

P&S are aware of the controversy about their suggestion that we postpone scientific progress, and are confident that the risk posed by UH can justify the measures they advocate. They explain their reasoning by appealing to the relative ease of harming as opposed to benefiting, and to two principles: 'big-loss aversion' and the 'availability bias' (Persson & Savulescu 2011, pp. 442-443).

The first principle, 'big-loss aversion' refers to a psychological tendency to be strongly adverse to any course of action that has a chance of bringing about extremely bad consequences, even if that chance is extremely small. To use an example similar to one of P&S's (Persson & Savulescu 2008, pp. 173-174), imagine a game of chance in which one must wager one million dollars for a 95% chance of winning an additional one thousand dollars, or else losing everything. Most people would be unlikely to accept such a wager, and would be no more likely to accept it if the chance of winning were even greater, say 99.99%; this is big-loss aversion. In a reverse wager, such as in which one must wager one thousand dollars for a 5% chance of winning an additional one million dollars, most people would be far more

[2] In fact, we have already discovered at least one, and research in this area is promising (Knapton 2015).

likely to accept the wager. This counterpart to big-loss aversion is 'big-gain attraction'.

P&S claim that, while big-gain attraction is an irrational attitude, big-loss aversion is useful because it compensates for the general fact that it is easier to harm than to benefit (Persson & Savulescu 2011, p. 442). As scientific progress advances, it will provide us with benefits as it always has, but also with the potential to cause harm. Because it is easier to harm than to benefit, advancement will bring increasingly greater potential to harm than to benefit. P&S think that further advancement will provide only minor benefits (winning the thousand dollar prize) but will increase the small chance of UH (losing the million dollar wager).

If big-loss aversion is so common and useful, and scientific progress really does bring with it a chance of UH, it seems as though big-loss aversion should motivate us to limit scientific progress; we should be compelled to give up the high chance of comparatively minor benefits it provides us in favour of negating the tiny chance that it will lead to UH. So why is P&S's suggestion that we postpone scientific progress such an unpopular one? Beyond the likelihood that people generally are not aware of just how dangerous modern technology can be, P&S suspect the explanation lies in the 'availability bias':

> "[W]e 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." (Ibid., p. 442)

We can imagine being struck by a car while crossing a road, because we know it happens quite frequently and some of us have direct experiences with it, so we tend to be careful about crossing roads; when we imagine being struck by a car, we experience negative emotions that make us adverse to behaviour that makes that possibility more likely. But obviously no one has yet experienced or witnessed UH, so no one is capable of vividly imagining its occurrence. It is this ability to vividly imagine a possibility that arouses the emotions involved in big-loss aversion. We are unable to grasp the risk of big-loss involved in further scientific progress

I do not find much plausible about this area of P&S's argument. First, their reasoning based on the principle of big-loss aversion is questionable, as are their analogies demonstrating it (Gunson & McLachlan 2013, pp. 879-882). Second, characterising

the possible benefits of scientific progress as 'relatively minor' seems to understate how important some of them have been in improving quality of life and preventing harms. P&S are confident they have not underestimated these benefits (Persson & Savulescu 2011, p. 444). However, this matter is debatable. Third, it is not clear that the availability bias renders the harms produced by scientific progress as difficult to imagine as P&S suggest. They themselves refer to several well-respected figures that seem fully aware of the dangers, and those objecting to their proposal to postpone scientific progress seem to be as well. It is plausible that the general human population is ignorant of the risk of UH, but they are not the ones responding to P&S's arguments.

Regardless of whether addressing Harm is worth postponing scientific progress or risking uncertain MB, as P&S advocate, we have reasons to think that neither one is necessary. This is because there is an alternative to MB, one that humans have employed for millennia: traditional moral enhancement (TME). In section 1.5 we saw P&S's claim that TME lags behind scientific progress, and because of this leaves us too morally deficient to responsibly make use of that scientific progress until we enhance our morality through more effective means such as MB. P&S doubt that TME will provide sufficient moral enhancement to address UH (Persson & Savulescu 2013, p. 130). Nevertheless, it might turn out that TME is promising enough after all that we need not implement MB.

Is MB the best course of action in order to address Harm? Answering this question will be the task of much of the rest of this thesis. MB has the potential to reduce the risk of UH, and also the prevalence of less-apocalyptic but still horrific WH. But other courses of action, such as social reform, cognitive enhancement, further technological advancements, and good old fashioned TME, all also have the potential to address Harm. These alternatives may be less risky, more effective, or easier to implement. We will come back to these alternatives in section 4, and weigh them against MB once we have investigated it thoroughly. It may turn out that we should focus our efforts on something other than MB if we are to address Harm.

We can, however, answer one part of the question right now: Along with Fenton and others, I argue that we should not intentionally limit scientific progress in an attempt to minimise the chances of UH occurring. Scientific progress offers hopeful prospects

for preventing Harm; it may provide us with the means to achieve MB, and cognitive enhancement might need to come hand in hand with MB. It would be counter-productive to postpone scientific progress for reasons based on P&S's claim that it will bring about an unacceptable risk of UH.

For now, we still have at least two good reasons to continue investigating moral bioenhancement as a possible solution to Harm, even though the alternatives exist. The first is that Harm is exacerbated by the effects of scientific progress; solutions to problems of Harm in the form of advanced technology will only be effective to the extent that we are motivated to put them to use in the right ways, and it is not clear that we will be so motivated any time in the near future. The second is that TME through means currently available to us seems unlikely to produce very quick or effective ways to address Harm, as will be argued in section 4. The need for moral enhancement, however, is immediate and dire; this may be enough to justify MB.

1.9: Summary of section 1

Moral enhancement is not an incoherent concept, even if it might be a difficult one to define. We have managed to capture one reasonable concept of moral enhancement, one that we apparently already utilise in order to make people morally better: the improvement of moral capacities through some kind of intervention. We have also found some reasons to think that we would be able to achieve moral enhancement, so defined, through biotechnological means; to achieve moral bioenhancement (MB). This is enough to think that MB is not impossible and cannot be dismissed out of hand.

Having established that there is at least one coherent sense in which we might be able to achieve MB, the question remains of whether we should. Persson and Savulescu have helped us to highlight two compelling reasons to consider developing and implementing MB: as scientific progress advances, deficient human moral capacities (IMD and MMD) will arguably cause corresponding increases in the prevalence of Widespread Harm (WH) and the risk of Ultimate Harm (UH). We will take the legitimacy of these concerns about both UH and WH (Harm) for granted.

However, there are alternative courses of action other than MB that might be able to reduce and prevent Harm: further scientific progress, and traditional moral enhancement techniques (TME) such as moral education. It remains to be seen whether MB will cause problems of its own, and whether those problems prevent it from being both worth the risk and preferable to the alternatives. In the next section we will try to construct a more rigorous concept of morality so that we are better-prepared to answer the above questions.

2: Morality as We Know It

Before we can discuss how MB might best be achieved, we must have a sufficient understanding of morality. Physical bioenhancement, for example, already exists and is commonly practiced. This is possible for two reasons: One, we have a sufficient understanding of the biological factors that affect our physical capacities; we understand how muscles work, and what the effects certain interventions will have on them. Two, definitions of our physical capacities are not controversial in most cases; if the subject is stronger, faster, or has more stamina than before the intervention, they have been physically enhanced.

Regarding MB, we are lacking in both of the above respects. Not only do we have less solid scientific bases for predicting what factors will influence moral capacities, but there is widespread and emphatic disagreement about how we should define moral capacities in the first place. There are many fundamentally distinct meta-ethical theories that attempt to explain what we mean when we talk about moral concepts, and there are dozens of finer points of distinction within groups of meta-ethical theories. Even those who can agree on a theory cannot usually completely agree on the implications of that theory. This section will establish some concepts most relevant to the objective of addressing Harm: moral value, moral motivation, moral intuitions, and broad cooperation.

2.1: Meta-ethics

We have been using the term 'moral judgements'. These are beliefs and claims about what is morally right and wrong. Moral judgements are statements about moral facts, the kind of statement that can potentially be true or false (truth-apt). The concept of truth-apt moral judgements draws one of the first major distinctions between kinds of meta-ethical theories: 'cognitivism' and 'non-cognitivism' (van Roojen 2004, section 1). This distinction determines what we think we are actually doing when we engage in moral discourse.

Consider the test-statement "torturing kittens for fun is morally wrong". The cognitivist might say that this is a claim about a moral fact. The statement could be understood as being about a moral property, call it 'moral wrongness', that is

somehow present in the act of torturing kittens for fun, and this is what makes the act morally wrong. Opinions vary widely about the exact nature of moral facts. They could be derivable from facts about the physical world. They could be meta-physical, outside of space and time. They could be subjective, or relative to culture, or constructed by humans. However we conceive of them, we need them in order to make truth-apt moral judgments. (Sayre-McCord 2015, section 2)

But cognitivist theories might be thinking about morality in a fundamentally incorrect way. The non-cognitivist instead explains apparent moral judgements in some way that does not require the existence of moral properties. When someone says or thinks "torturing kittens for fun is morally wrong", they are not attempting to make a truth-apt statement. It may sound like they are, but the actual cognition involved is not a process for making a true-or-false statement about some property involved in a state-of-affairs. Some non-cognitivists, for example, understand moral judgments as expressions of endorsement or opposition (van Roojen 2004, section 2).

Whatever meta-ethical approach we take, there is an apparent point of agreement: morality is concerned with moral value. Morally good things have positive value and morally bad things have negative value, and this value can be understood in a number of different ways depending on the meta-ethical theory you endorse. We will investigate moral value later.

The concept of normativity is usually fairly central to morality. For example, it is commonly held that moral judgements are 'normative claims'; that is, claims about how the world ought to be and what actions ought to be taken (Sayre-McCord 2015, section 3). If I steal your wallet, and you become angry at me, I might ask you why I ought not to have done it. 'Stealing is wrong!' you say. 'I understand that,' I reply, 'but why ought I not to do things that are wrong?'. If you think I'm being unreasonable, that likely means you believe moral value has some kind of normative force. Things that are morally wrong ought not to be done, and things that are morally right ought to be done. In other words, positive moral value is desirable in some way; perhaps intrinsically valuable, or else valuable because of the practical benefits that accompany it.

Moral normativity is controversial, so I will say little about it directly. What is more important is the concept of moral motivation, which is linked to normativity; if I am convinced I ought to do something, that thought may motivate me to do it.

Remember our working definition of moral enhancement (i.e. improvement of moral capacities) outlined in section 1: A morally enhanced person makes more correct moral judgments than they did before the enhancement and is motivated to act in accordance with them more consistently. This definition assumes that moral discourse involves the making of truth-apt moral judgments, because they have the potential to be correct. But we could interpret the improvement of moral capacities in some other way, for example one that is compatible with non-cognitivism. What is important is that our definition of moral enhancement should include the two key concepts identified above: moral value and moral motivation.

This is because, as we will see, these concepts are the most relevant to the objective of MB: addressing Harm.

2.2: Moral value and moral motivation

Ways of understanding moral value are as varied and contentious as the meta-ethical theories behind them. But we do not need to solve this debate. We just need to know how people tend to think about moral value, and why.

One of the most common moral values is the reduction of harm. When considering a state-of-affairs from a moral viewpoint, one important factor tends to be whether it causes any more harm than is necessary, or whether it prevents or minimises harm that would have otherwise occurred. But, while practically all respectable meta-ethical theories consider minimisation of harm to be morally relevant, it is not always thought to be the most important factor. We might consider moral principles such as honesty, respect for authority, or the pursuit of the greater good to take precedence over minimisation of harm. For example, even if you know that telling a friend that their spouse is unfaithful may cause them psychological pain, harming them, it would not be unreasonable to believe that telling them would have more moral value than lying to them and sparing their feelings.

We have space only to scratch the surface of moral value. This is because meta-ethical theories are so numerous and varied, and each one has something different to say about moral value. Three prominent meta-ethical theories are consequentialism, deontological ethics, and virtue ethics. We will briefly examine each one to gain some examples of the different ways that moral value can be understood.

Consequentialist theories find moral value solely in the consequences of actions. Certain states-of-affairs are morally better than others, and an action that brings about better states-of-affairs than an alternative action is therefore morally better than that alternative. Probably the most familiar consequentialist theories are varieties of utilitarianism, which tend to find moral value in something like 'actions that maximise happiness impartially' (Sinnot-Armstrong 2014, section 1).

Deontological theories are varied, but generally contrast well with consequentialist theories. Rather than emphasising consequences, they emphasise duties or rules. Sometimes, that which is morally right will not bring about the best consequences. Some well-known examples of morally wrong actions according to deontological theories are torture, or killing innocents. Such actions are wrong regardless of whether taking them might prevent similarly wrong actions from occurring in the future (Alexander & Moore 2012, section 2.1).

Virtue ethics theories emphasise the moral character of individuals, determining moral value based upon virtuous character traits. Virtues are a kind of disposition:

> "...a disposition which is well entrenched in its possessor, something that, as we say 'goes all the way down', unlike a habit such as being a tea-drinker". (Hursthouse 2013, section 2)

Honesty might be one such virtue, for example, though having the virtue of honesty is more complex than simply having a tendency to tell the truth.

These three cognitivist theories are examples of moral realism. Other cognitivist theories understand moral value in different ways, as mentioned earlier, and moral value for non-cognitivist theories contrasts even more. But we have seen enough for a basic understanding of moral value, so we will move on.

Important to the psychology behind moral value is that it seems as though it plays a part in motivation to act in accordance with that value. This ties into moral

normativity; perhaps morally valuable actions somehow ought to be taken, and this can provide motivation. If someone rescues a stranger's child from drowning in a public pool, and we ask them why they did so, we would not find it strange if they respond "Because it seemed like the right thing to do". Plausibly, what they mean is that their moral values indicated to them that saving the life of a child under those circumstances has positive moral value; at least, higher moral value than plausible alternatives such as not acting at all. Upon thinking that the prospective action had moral value, the rescuer was compelled to take that action. We might be able to come up with alternative motivations; perhaps the rescuer didn't want to be held accountable by the child's family for allowing them to die, or perhaps they wanted to selfishly reap the rewards that come with being considered a hero. But it seems plausible that the action's moral value played a part in motivating it.

We looked at several meta-ethical theories above, and each one offers a different explanation for why saving a child from drowning has positive moral value. The consequentialist might say that saving the child produces the best consequences. The deontologist might say that as a rational agent you should be able to understand why there is a moral imperative to save the child. The virtue ethicist might say that a virtuous person has qualities that would compel them to save the child. The non-cognitivist may express their moral values in endorsement of saving children's lives, which motivates them to do so themselves and to positively regard others who do the same. Whatever someone's reason for believing that an action has moral value, that belief can provide motivation for acting in accordance with the moral value.

But moral motivation is not synonymous with moral value. It is possible to hold some belief about moral value without being motivated to act in accordance with it. In section 1.3 we saw an example of how an individual, Algernon, can make a moral judgement, e.g. that it is wrong to give cigarettes to children, but decide to take an action that they acknowledge is morally wrong. This might be because Algernon has the motivation to do the right thing, but has stronger motivation to do something that conflicts with it; perhaps he is addicted to cigarettes and resents anyone who is not. It might instead be that Algernon has some sort of perverse desire to do the morally wrong thing, and does so purely for selfish self-satisfaction. Whatever the explanation, we cannot be sure that the belief that something has positive moral value alone provides motivation to act in pursuit of it.

What moral value definitely does is provide a target for moral motivation. To whatever extent we are morally motivated to do what we perceive to be morally valuable, it makes no sense to talk of moral motivation in complete isolation from moral value. Moral value and moral motivation are separate but related.

With so many different ways of understanding and determining moral value, how can we ever go about doing so with any confidence? Instead of focusing on finding a theory that seems to get moral value right most of the time, we could focus on moral value itself. We can ask what kinds of things tend to be thought of as morally valuable most commonly. Once we know that, we might be able to figure out why we tend to value them.

I suggest that common moral values tend to be significant factors in enabling people to form and effectively coexist within societies; motivating people to take actions that benefit collective interests over individual self-interest. Call this 'promoting cooperation'. By this I mean nothing more than that there is a correlation between many states-of-affairs that are commonly considered to be morally valuable and many states-of-affairs that tend to promote cooperation. I do not claim, for example, that the tendency to promote cooperation makes something morally valuable.

It is not hard to find support for this correlation.

Consequentialist values are usually concerned with attaining consequences that are in some sense for the greater good; utilitarianism is concerned with the impartial maximisation of good among the greatest possible number of individuals, not just one or a few (Sinnott-Armstrong 2014, section 1).

Kant's famous deontological theory is guided by requirements that take the whole of humanity into account, such as never using others solely as a means to one's own ends, and that any action we take could be advocated as a universal rule (Alexander & Moore 2012, section 2.4).

One of the most influential philosophers in recent centuries, David Hume, formulated a moral theory largely concerned with the benefit of mankind and society, not just the individual. He writes of the moral duty to avoid the deformity of vice and embrace the beauty of virtue (Hume 1751, section I). Hume seems to take it to be

obvious that at least a part of what makes a quality virtuous is it role in helping humans happily coexist within societies (Ibid., section II part II).

Hume finds moral value in qualities that are "estimable" to oneself or others (Ibid., section IX part I). People respond with approval to some qualities, virtues, and with aversion to others, vices. Hume does not think it is possible to be entirely indifferent to such qualities, to truly prefer vice over virtue, because of the corresponding aversion or approval that we experience in response to each (Ibid., section V part II). However, it is not just these experiences that constitute moral value. Moral approval responds to qualities that are useful; more specifically, those that are useful to society:

> "The social virtues are never regarded without their beneficial tendencies, nor viewed as barren and unfruitful. The happiness of mankind, the order of society, the harmony of families, the mutual support of friends, are always considered as the result of their gentle dominion over the breasts of men." (Ibid., section II part II)

Although moral value is controversial, and not all moral values held by everyone succeed in promoting successful cooperation under all circumstances, it would be difficult to find many commonly held moral values that are entirely unconcerned with collective interests over individual self-interest. Keep this apparent correlation between cooperation and moral value in mind.

When engaging in moral discourse, most normal people make no explicit mention of any of the above meta-ethical concepts. When asked to give reasons for the things they think and say in a moral context, they give simple answers that others can relate to and agree with without complicated meta-ethical reasoning or justification. 'Didn't your mother ever teach you that stealing is wrong?'; 'It is always better to be honest than dishonest'; 'Just because they have the death penalty doesn't mean it is right'; 'Treat others as you want to be treated'; 'An eye for an eye leaves the whole world blind'; 'I just know being an organ donor is the right thing to do'; and so on. Call these 'moral platitudes'.

We need a streamlined way of thinking about meta-ethical theories and moral values; one that represents how normal people think about morality. Examining

common moral platitudes is a good way to do this. We cannot simply choose the understanding of morality we like best and expect everyone else to engage in moral discourse under those terms. If you want to convince someone of some claim you make about moral value, it is probably better to appeal to moral platitudes than to a meta-ethical theory.

Moral platitudes do not emerge from an understanding of meta-ethical theories. Rather, meta-ethical theories are attempts to find a structure that best fits with moral platitudes. Moral platitudes are the way that normal people tend to express their moral values. We can use some sort of account of moral platitudes similar to Frank Jackson's. He illustrates how they work through a comparison to grammar theory:

> "We have a folk theory of grammar with some clauses we can write down more or less roughly – 'Verbs should agree with subjects', for example – but there is a lot that only expert grammarians can write down, despite the fact that what they write down is based on what we folk do when we classify a sentence as grammatical: they take our classifications and seek to articulate what is guiding us in making those classifications." (Jackson 2000, p. 130)

We can use moral platitudes to think about moral value and meta-ethical theories without getting caught up in the above debates. We can use them to understand the moral capacities of normal people, and why those moral capacities might be deficient. We can also ask how they ended up that way. The answer lies in the connections between moral value, moral motivation, and the promotion of cooperation.

2.3: The practical value of cooperation

There is a plausible biological explanation for why common moral values and moral platitudes relate to the benefit of society and the promotion of cooperation. Cooperation was of practical benefit to our evolutionary ancestors, so individuals that tended to cooperate effectively were more successful than their peers and passed on those tendencies to future generations more frequently. We have reasons to think

that this tendency to cooperate worked through psychological capacities related to what we now call morality.

> "The problem of cooperation, then, is the problem of getting collective interest to triumph over individual interest, when possible." (Greene 2013, ch. 1); "...nearly all cooperative enterprises involve at least some tension between self-interest and collective interest, between Me and Us." (Ibid.); "Morality evolved as a solution to the problem of cooperation..." (Ibid.)

Cooperation requires that individuals sometimes take actions that are not directly in their own interests in order to achieve consequences that are best for the group overall. This problem of cooperation is often illustrated in the context of 'prisoner's dilemma' thought experiments. Variations on the prisoner's dilemma have been around for decades, but the basic idea is as follows:

You and a partner have both been arrested for jointly committing a crime. You are each held in separate rooms, unable to communicate, and asked to testify against one another. You know the following: If one of you testifies and the other refuses, the one who testifies will receive only one year in prison while the other receives ten years. If you both testify, you will each receive five years in prison. If you both refuse, you will each receive only two years in prison. Refusing to testify is analogous to cooperative behaviour, while testifying is analogous to self-interested behaviour. This is because testifying is the course of action that guarantees one will not receive the worst outcome (ten years in prison), and gives one a chance to gain the best outcome (one year in prison) by sacrificing another's interests. Refusing has a chance of attaining the best outcome with regard to collective interests (two years in prison each), but at a substantial risk to one's self-interests.

This situation is meant to illustrate the choices our evolutionary ancestors had for dealing with the problem of cooperation. Those whose behavioural tendencies resulted in them choosing to cooperate most consistently attained the highest benefit to their collective interests. Groups of such cooperators had a reproductive advantage compared to less-cooperative groups whose members chose self-interest over collective interests more frequently. Through evolution by natural selection, and perhaps later through cultural and societal evolution, cooperative behavioural

tendencies became more and more prevalent as they out-competed self-interested tendencies.

Persson and Savulescu identify some important features of our evolved cooperative tendencies:

> "According to our preferred view, the core of our moral dispositions comprises, first, a disposition to altruism, to sympathize with other beings, to want their lives to go well rather than badly for their own sakes."; "Secondly, there is a set of dispositions from which the sense of justice or fairness originates. The most basic of these dispositions are, we believe, the ones that have been called 'tit-for-tat'. Evolutionary theorists have found that collectives in which this pattern of reciprocal reactions is widespread are most successful in terms of survival and reproduction." (Persson and Savulescu 2008, pp. 168-169)

Cooperative tendencies must include 'tit-for-tat' dispositions in order to allow individuals to cooperate without being exploited by the self-interested minority. If a cooperator is inclined to altruistically help others indiscriminately, they are likely to end up worse off than another individual, a 'free-rider', who is less inclined to do so. The free-rider can accept help and never return the favour. The cooperator has then sacrificed their interests for the benefit of the free-rider's interests. If this repeats, the co-operator will keep sacrificing and the free-riders will keep benefiting for free. But if our cooperative tendencies include the 'tit-for-tat' dispositions described above, the cooperator will stop cooperating with individuals who they think are free-riders. It then becomes in everyone's best interests to be a cooperator instead of a free-rider.

Our evolved cooperative tendencies are complex, and we cannot fully discuss them here[3]. But a primary function of them is to motivate humans to cooperate with one another to the extent that the benefits of cooperation are attained, but not to the extent that the risk of exploitation by self-interested free-riders becomes too great. Cooperative tendencies must motivate humans to prioritise the interests of others over self-interests, but not all the time and not just for any others.

---

[3] See *Cooperation and its Evolution* (eds. Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser) for more on the subject.

Looking to evolutionary biology for an explanation of morality might seem flawed. Morality is arguably not something that can be studied through scientific methods. Perhaps there are moral facts, but they are not the same as natural facts about, for example, the boiling and freezing points of water or the relationship between mass and velocity; they are non-natural facts, which exist outside the realm of ordinary science and explanations.

So if we are to accept any scientific claims about morality, we will need to make certain assumptions. Specifically, we will need to be open to the possibility that morality is no different from any other phenomena studied through scientific methods. If moral properties are non-natural, our usual physical experiments may be unable to help us determine anything about them. Evolutionary theory and the study of biology would be poor ways to learn anything about morality. That is unless we accept the assumption of something like 'minimal moral naturalism':

> "Minimal moral naturalism [...] simply proposes to apply relevant sciences for studying how humans manage to produce their moral valuations and perform whatever they take to be morality without any nonexistent features involved." (Shook 2012, p. 4)

If we accept such an assumption, the correlation between moral value and cooperative tendencies starts to make more sense. If moral value does not represent ineffable properties that indicate some sort of 'right'-ness beyond the understanding of our sciences, the evolutionary link is one of the most plausible explanations for it that we have. In order to promote behaviour that works as successful cooperative tendencies, humans evolved the tendency to perceive moral value; moral value serves as a guide to successful cooperation.

Unless what we think we know about morality and its origins through evolutionary biology turns out to be misguided, we have good reasons to think that cooperation is central to morality. We place moral value in the things that we do because of their connection to the promotion of cooperation. This leaves us with reason to think that moral platitudes are roughly analogous to rules for generally promoting cooperation.

2.4: Moral intuitions

Our discussion of morality so far suggests two interesting points: (1) Moral value serves as a target for moral motivation, and it is common to be motivated to take actions that cause states-of-affairs that one morally values. (2) Behaviour that allowed our evolutionary ancestors to cooperate and form societies was beneficial to them, providing a reproductive advantage.

If we combine these two points, we have a plausible explanation for why humans commonly think of states-of-affairs that promote cooperation as having positive moral value; the tendency to do so evolved through natural selection and cultural/societal influence. But there remains the question of how this tendency operates on an individual, psychological level. If we are motivated to cooperate, how do we experience that motivation? I argue we experience it as what I call moral intuitions.

Intuitions, more generally, are thought of as judgments we make or beliefs we have, related to some kind of immediate, automatic response. By moral intuitions I mean the experience of unreasoned psychological impulses that cause us to hold moral values (or, according to a cognitivist understanding, make moral judgments). I feel disgust, aversion or condemnation at the idea of enslaving or hurting someone; I am instantly compelled to go out of my way to return a dropped wallet to a stranger; I unthinkingly approve of a course of action which results in fewer deaths than an otherwise equal alternative. All of these are moral intuitions.

As I understand them, moral intuitions are non-rational mental processes that provide bases for moral reasoning and fundamentally influence the moral values we hold. Moral intuitions are most easily conceived of in terms of moral platitudes, which represent the moral intuitions we experience in terms of ordinary language. This allows us to share and discuss them with one another: “It is clearly better if only one person needs to die instead of five”, “Not if you cause that one death yourself, rather than just allowing the five happen”. Reasoning is not involved in this kind of discourse, beyond determining how to apply the moral platitudes we accept.

It might not be agreed that moral intuitions are distinct psychological factors. They could instead be classified as moral judgments or beliefs, or some non-cognitivist process. If I witness someone torturing a kitten for fun and immediately think

"That's horribly immoral!", am I experiencing a moral intuition, forming a moral belief, making a moral judgment, expressing a moral value, or something else entirely? Any of these seems appropriate. But I argue it is useful to distinguish moral intuitions. The fact that there is such a close relationship between them and moral judgements and beliefs could be explained by the fact that moral intuitions are more or less directly responsible for them.

It might also be thought that moral intuitions are simply emotions, and that we can draw no meaningful distinction between moral emotions and moral intuitions as I define them. Moral judgments are often associated with emotions. I might refrain from stealing something because the idea of doing it makes me feel guilt, which causes psychological suffering. I might donate to Red Cross because such acts tend to make me feel satisfied, and contemplating doing them is pleasant. I might feel righteous anger at the harming of a defenceless victim, making me less adverse than usual to the prospect of confrontation and motivating me to intervene. Indeed, moral intuitions are frequently involved with emotions, often the particularly strong emotions that we associate with moral judgments and beliefs.

However, I argue that moral intuitions are not exclusively emotions. Jonathan Haidt agrees, instead emphasising the automatic and unreasoned character of intuitions:

> "...intuitions (including emotional responses) are a kind of cognition. They're just not a kind of reasoning." (Haidt 2013, ch. 2)
>
> "The crucial distinction is really between two different kinds of cognition: intuition and reasoning. Moral emotions are one type of moral intuition, but most moral intuitions are more subtle; they don't rise to the level of emotions." (Ibid.)

One of the most important features of moral intuitions is that they fundamentally influence moral reasoning. Haidt argues that when we make reasoned moral judgments, the moral intuition usually comes first; the intuition causes the judgment, which causes a search for reasoning in support of the judgement. Rather than providing rational basis for moral values and judgements, moral reasoning tends to be 'post hoc'; it comes after the moral value has been accepted, after the judgement has already been made, to rationalise our moral intuitions. (Haidt 2001, pp. 822-823) This is why moral intuitions are more fundamental to moral value and moral motivation than moral reasoning is.

Studies demonstrate how post hoc moral reasoning can malfunction. This happens when reasoning given in support of moral judgements can be proven irrelevant, but the moral judgement remains unchanged; this suggests that the judgement is initially based on a moral intuition. One example is the common, strongly experienced moral intuition that incest is wrong. In one study, subjects were told the story of a brother and sister, Julie and Mark, who have sex with one another while alone together on vacation in France, on just one occasion. They use condoms and birth control pills. It is stipulated that neither of them suffer any harm as a result, and the secret makes them feel closer to one another. (Haidt 2013, ch. 2)

Subjects were asked if it was wrong for Julie and Mark to have sex. Only 20 percent of subjects said it was okay. But when asked to give reasons, subjects rapidly produced and discarded reasons, but rarely changed their minds about the initial judgement when they were unable to find a satisfactory reason to justify it. One subject first objected to the possibility of incest offspring being 'deformed, most of the time', then to Julie's and Mark's ages, then to incest being something people are not 'brought up to' do. Even after accepting that none of their reasons were relevant, most subjects maintained that what Julie and Mark did was wrong. (Ibid.)

> "...It's obvious that people were making a moral judgment immediately and emotionally. Reasoning was merely the servant of the passions, and when the servant failed to find any good arguments, the master did not change his mind." (Ibid.)

The way moral intuitions such as emotions influence moral reasoning can be more directly demonstrated. Haidt details various experiments in which induced negative emotions such as disgust can have definite effects on the moral judgments people make. For example, secretly spraying foul-smelling fart spray nearby before asking passers-by to fill out a short survey on moral judgements resulted in harsher moral judgements (Haidt 2013, ch. 3).

Moral intuitions can also help to explain moral motivation. In section 2.2, I suggested that motivation is an important part of moral value; if we believe a state-of-affairs has positive or negative moral value, we are commonly motivated to act in accordance with that value. Moral intuition is a plausible link between moral value and moral motivation. Haidt emphasises the motivation supplied by moral intuitions

over moral reasoning (Haidt 2001, pp. 823-825). Similarly, Joshua Greene points out how emotions tend to result in action (Greene 2013, ch. 5).

Reasoning is still important to our moral capacities. But Haidt argues that, instead of influencing our own moral judgments, moral reasoning primarily influences the moral judgments of others. While rational self-reflection sometimes results in influence on our own moral intuitions, our moral intuitions are far more susceptible to influence by reasoning given by other individuals (Haidt 2001, pp. 818-819). This is why he calls his model the 'social intuitionist' model. Keep this feature of moral reasoning in mind; even if moral reasoning is secondary to moral intuitions in forming one's own moral values, it still plays an important part in justifying those values to oneself and influencing the values of others.

Moral intuitions are products of evolution, but they did not evolve as a concrete set of instructions to motivate specific cooperative actions. Rather, evolution has predisposed us to form and experience some set or other of moral intuitions, specifics of which are probably determined by a combination of an individual's genetics and their experiences within their culture and society (Ibid., pp. 825-826). This results in different individuals, cultures, and societies emphasising different moral intuitions.

Moral intuitions can therefore explain moral disagreement between individuals belonging to different groups. These groups can be divided by nation, culture, socio-economic status, minority membership, etc. For example, the kinds of moral values commonly emphasised by citizens of the USA in southern states versus those in northern states. Studies have revealed what is often called an honour-culture generally prevalent in southern states, which places high moral importance in concepts that do not factor nearly as much into moral values held in northern states.

In one study, various employers were all sent relevantly similar letters apparently from a well-qualified male applicant. In the letter, the applicant admitted to violently killing another man during a bar fight; the now deceased man had claimed to have slept with their wife, laughed in their face, and challenged their manhood. Employers in southern states were more positive about the prospect of hiring the applicant, and more likely to express the opinion that the applicant had done nothing wrong (Nesbitt & Cohen 1996, pp. 74-75). The circumstances of the Northern state

employers and Southern state employers seem relevantly similar, and yet they make opposing moral judgements. It seems more plausible that those moral judgements, their moral values, are most fundamentally based on differing moral intuitions rather than differing moral reasoning abilities. Haidt draws similar conclusions, suggesting that differing moral intuitions can explain the "bitterness, futility, and self-righteousness" of moral debates such as those over homosexuality and abortion (Haidt 2001, p. 823).

Attributing fundamental importance to moral intuitions might be a mistake. Hanno Sauer represents a moral 'rationalist' position, as opposed the kind of position that emphasises moral intuitions, which can be called moral 'sentimentalist'. Sauer argues that there is "nothing intrinsically dubious" about post hoc reasoning that justifies moral intuitions. We can experience 'educated' moral intuitions, resulting from the internalisation of practical reasons we have for acting morally. (Sauer 2012a, p. 264-265)

Sauer also argues that we can rationally distinguish factors that influence our moral intuitions in morally irrelevant ways (e.g. unpleasant smells) from those that do so in morally relevant ways. In experiments such as the smelly office, we should not consider the moral judgement induced through manipulation of emotions to be a case of genuine moral judgment, because it is responding to "morally irrelevant" emotional reactions (Sauer 2012b, p. 106).

> "...genuine moral competence is characterized by the feature that if after a moral judgment has been made, a certain piece of information is added in the light of which the initial judgment is undermined, the judging subject will suspend his belief or adequately back it up with appropriate further grounding." (Ibid.)

This could explain how morality is fundamentally based on reason rather than intuition; we can use reason to divide morally-relevant intuitions from morally-irrelevant ones.

I am doubtful that it is possible to distinguish between morally relevant and morally irrelevant factors through cold reasoning alone. Any rational judgement about which intuitions are morally relevant would plausibly involve first a moral intuition followed by moral reasoning that justifies it. It is therefore hard to see how this process of moral self-reflection could reveal reasons uninfluenced by intuition, rather

than just making our moral intuitions more consistent with one another. This type of moral reasoning, 'moral consistency reasoning' (Campbell & Kumar 2012, p. 274), makes more sense with regard to moral intuitions, and can explain why we find certain factors morally relevant and others morally irrelevant. But the process is still fundamentally based upon moral intuition.

Regardless, Sauer provides plausible arguments for thinking that moral reasoning is more fundamental to our moral capacities than moral intuition is. I maintain that moral intuition is more fundamental. But we cannot engage further with this debate. We do not need to fully accept any particular sentimentalist account such as Haidt's or Greene's. But this discussion leaves us with some good reasons to think that moral intuition is more fundamental to moral values and moral motivation than moral reasoning is. But remember that I understand moral intuition as something that cannot be completely isolated from moral reasoning; they are two distinct cognitive processes, but they are inextricably interrelated.

We should take away the following from this brief discussion of moral intuitions: If moral intuitions play a fundamental role in moral value and moral motivation, as I claim, MB must closely involve moral intuitions if it is to be effective. This account of moral intuitions could easily be challenged, but we will not do so in this thesis.

2.5: Broad cooperation

I argued that cooperative tendencies evolved because of the practical benefits of cooperation. I then argued that moral intuitions are how we experience this evolved cooperative tendency. But the kind of narrow cooperation that benefitted our ancestors is no longer beneficial enough. As argued in section 1, scientific progress has resulted in vastly different circumstances for humans than those in which our moral intuitions evolved. If we are to address Harm, we need to promote a new form of cooperation that allows all humans to work together on a large scale towards objectives that are of the greatest benefits to everyone. Call this 'broad cooperation'.

We do not currently place a lot of moral value in broad cooperation, or at least not as much as narrow cooperation (i.e. common-sense morality). Narrow cooperation does have practical value as well as moral value. But if we focus on practical value instead

of moral value, it seems to me that broad cooperation would be even more practically valuable than the narrow cooperation that we currently value morally. I argue that our current moral intuitions evolved because of their tendency to motivate cooperation. However, I also argue that our current moral intuitions do not motivate broad cooperation.

Our current moral intuitions evolved because they were adaptive in an environment in which small groups of individuals cooperated among themselves, while competing with one another. This is because evolution is driven by competition for resources. Cooperation needs to give one group an advantage over competing groups, so that the cooperative group reproduces more successfully and passes on their cooperative tendencies (Greene 2013, ch. 1):

> "Imagine, for example, two groups of herders, one cooperative and one not. The cooperative herders limit the size of their individual herds, and thus preserve their commons, which allows them to maintain a sustainable food supply. The members of the uncooperative group follow the logic of self-interest, adding more and more animals to their respective herds. Consequently, they erode their commons, leaving themselves with very little food. As a result, the first group, thanks to their cooperative tendencies, can take over. They can wait for the uncooperative herders to starve, or, if they are more enterprising, they can wage a lopsided war of the well fed against the hungry. Once the cooperative group has taken over, they can raise even more animals, feed more children, and thus increase the proportion of co-operators in the next generation. Cooperation evolves, not because it's 'nice' but because it confers a survival advantage." (Ibid.)

This is the primary reason why moral intuitions did not evolve to be capable of promoting broad cooperation; it was not adaptive for them to motivate cooperation outside of the in-group. Our moral intuitions have an inherent in-group bias.

Dispositions of 'tit-for-tat' reciprocity, as described in section 2.3, also pose a barrier to broad cooperation. They evolved to allow humans to successfully cooperate without too great a risk of exploitation by free-riders. But broad cooperation requires altruism between huge numbers of people, most of whom have no reason to expect that their altruistic actions will ever be directly reciprocated. Everyone benefits from

broad cooperation overall, but our evolved tit-for-tat dispositions are unlikely to be sensitive to this.

In-group bias and tit-for-tat reciprocity mean that our moral intuitions motivate us to cooperate only with certain other individuals. Persson and Savulescu point out that the evolved tendency toward altruism, a primary moral intuitions, does not usually "extend indiscriminately to foreign and unfamiliar individuals", resulting in the common human characteristic of xenophobia (Persson & Savulescu 2012, p. 38). This is likely a defence against free-riders; we are more likely to be exploited by outsiders, so it pays to be less willing to cooperate with them. They also state that they think human moral capacities are "largely limited to in-groups as against out-groups" (Persson & Savulescu 2013, p. 129).

We cannot get too deeply into this discussion. But the above brings attention to two fundamental features of our moral intuitions that plausibly tend to undermine broad cooperation: In-group bias, and tit-for-tat reciprocity.

Broad cooperation is of particular relevance to the MB discussion because the main motivation behind MB is the need to address Harm. The causes of Harm discussed in sections 1.5 and 1.6 are all arguably related to a lack of broad cooperation among humans. MMD issues can be addressed by large-scale cooperative actions, which would be far more achievable if broad cooperation were promoted. IMD issues could be addressed through morally depraved individuals developing a broadened sense of in-group, so that they are unwilling to harm as many people as they were, though this seems less plausible. More plausible is the possibility that broad cooperation would result in more effective cooperative countermeasures to IMD; populations willing to employ increased surveillance against WMDS (Persson & Savulescu 2012, p. 59), or unwilling to tolerate self-interested leaders in power.

Instead of pointing to a lack of broad cooperation, we could leave it at the claim that moral badness causes Harm. IMD, such as terrorism, threatens UH because of immoral people attempting to do that which is morally bad. MMD, such as failure to address climate change, threatens UH because people are not morally good enough to sacrifice their own interests for the sake of future generations. But we have already seen how difficult it is to agree with each other regarding our moral values. Different

individuals and cultures will point to different causes of Harm depending on what their differing moral intuitions tell them is morally wrong.

Broad cooperation is not problematic in this way. There may be disagreement about what will best promote broad cooperation, but it will be practical, not moral, disagreement. Practical disagreement has a better chance of being resolved because it is not dependent on potentially irreconcilable differences between moral intuitions. It is dependent upon facts that can be conclusively established through scientific methods. Additionally, claiming that there is a lack of broad cooperation is not merely an identification of a problem but also a suggestion of how that problem might be addressed: we aim to broaden human cooperation.

Here are couple of more specific examples of how broad cooperation could help to address Harm.

In developed countries, policies that intend to address Harm, for example by limiting carbon emissions to counteract climate change or redistributing wealth to counteract extreme poverty, tend to be unpopular. This is exacerbated by features of the narrow cooperation currently motivated by moral intuitions. If a higher percentage of a population is willing to offer stronger support for policies that benefit not just themselves and their peers but as many people as possible, there will be a stronger incentive for politicians to base their campaigns on such policies, and a higher likelihood that such politicians will be voted into power within democratic societies.

Individuals with vastly dissimilar religions, ideologies, and cultural values conflict on many important issues. These conflicts are largely caused by the in-group vs. out-group nature of our current moral intuitions; this creates barriers to large-scale cooperative action, and makes excessively harmful treatment of out-groups seem justifiable. We saw one example of this in the American honour-culture, but there are more extreme examples in the form of cultural and religious traditions. If broad cooperation could be motivated strongly enough to overcome these intuitive conflicts, different groups would perceive greater value in cooperating with one another than in upholding the moral values that divide them.

2.6: Summary of section 2

Morality is full of disagreement. One disagreement is over what kind of meta-ethical theory most accurately describes morality, another over how moral value should be determined, and another over the roles reason and intuition play in our moral capacities. Without getting too deeply involved in this controversy, there is not a lot we can confidently claim about morality. We can, however, use common moral platitudes to illustrate how some people tend to engage in moral discourse, and use them to make some reasonable assumptions about our moral capacities.

I argued that there are links between cooperation, moral value, and moral motivation. These claims are supported by modern scientific research and by plausible philosophical theories.

I also suggested that moral intuitions provide a foundation for moral value and reasoning, and provide a link between moral value and moral motivation. Moral intuitions include emotions, but also automatic, unreasoned responses (also called 'gut reactions') that we experience in situations that we take to be morally relevant. This characterisation of moral intuitions is one of the most contentious parts of this thesis; there is a lot that could be said in response to it, and I have left it mostly undefended.

Moral intuitions are plausibly how we experience our evolved tendency to be cooperative. But this kind of narrow cooperation is not sufficient to address the human moral deficiencies that contribute to Harm, mainly because it operates through emphasising the concept of an in-group vs. an out-group. We need a radical, fundamental change in our moral intuitions in order to cooperate in a more effective way that does not disregard individuals' interests on the basis of their out-group membership. I call this theoretical kind of cooperation 'broad cooperation'.

The promotion of broad cooperation is a promising prospect for reducing and preventing Harm; the examples of Harm we looked at in section 1 could all plausibly be addressed by an increased tendency to be broadly cooperative.

If the above is correct, then broad cooperation and moral intuitions should both play a central part in any attempt we make to develop or implement MB.

3: A New Strategy

In section 1 I argued that moral enhancement means improving moral capacities. It is not obvious what it would mean to do this. Despite this, discussion of prospective MB has so far assumed just one interpretation: improvement of moral capacities means making people *morally better*. But being 'morally better' is not the only thing that having 'improved moral capacities' could mean.

This section will examine the kind of MB strategy typically discussed, one that prioritises the satisfaction of moral intuitions, and contrast it with one that instead prioritises broad cooperation. Moral intuitions still have a part to play in this new strategy, but as a means rather than an end. We will see how the typical strategy has advantages, mainly in terms of palatability, but also disadvantages in terms of effectiveness and ease of implementation. The new strategy has its own problems, most prominently that it seems to advocate morally questionable measures. Although we cannot be certain this new MB strategy is the best way to address Harm, I argue it is clearly a better option than the typical strategy.

3.1: The objectives of moral enhancement

How do we improve moral capacities?

There are various factors commonly taken to be relevant to our moral capacities, though most are controversial. Here are a few non-technical examples of factors that might indicate that an individual has high moral capacities, in the same way that an Olympic athlete has high physical capacities: Strongly feeling empathy towards other individuals; a tendency to perform actions that one takes to have high moral value; morally valuing the same things that 'morally good' individuals do.

As argued in section 1, biotechnology has the potential to manipulate psychological factors in order to improve an individual's moral capacities. So before we can develop MB we need to know what factors to manipulate, and to what extent, in order to result in improved moral capacities.

Our working definition of an intervention that improves moral capacities was one that causes an individual to make more correct moral judgments, and act in

accordance with them more consistently. We saw how this working definition is related to two broad categories of features important to our moral capacities: moral value, and moral motivation.

In order to improve our moral capacities, MB could manipulate the way we conceive of moral value, such as the types of actions we morally value the most, and the way we are motivated to act in accordance with moral value, perhaps making us more strongly motivated to do so. I suggest that MB could best achieve this manipulation by targeting moral intuition, which is of fundamental importance to both moral value and moral motivation.

However, our working definition does not describe the only end to which manipulation of our moral intuitions could be put.

John Shook explores some possibilities as to what MB might involve, identifying five moral mechanisms that it might manipulate and two distinct models for the ends to which such manipulation might be attempted.

Regarding moral mechanisms, Shook suggests the following possibilities:

> "Enhance a person's sensitivity to moral features of situations—resulting in heightened moral appreciation. Enhance a person's thoughtfulness about doing the moral thing—resulting in stronger moral decisions. Enhance a person's moral judgments that get the right moral answer—resulting in more correct moral judgments. Enhance a person's motivated choice to do what moral judgment indicates—resulting in improved moral intentions. Enhance a person's volitional power to act on a moral intention—resulting in more moral will power." (Shook 2012, pp. 5-6)

The former three mechanisms ('moral appreciation ', 'moral decisions', and 'moral judgments') relate to moral value. They are what come into play when someone considers a situation (e.g. a group of delinquents torturing a kitten), identifies the morally relevant aspects of the situation (e.g. the suffering of the kitten and the delinquents' reasons for causing that suffering), becomes interested in the moral value of potential courses of action (e.g. should something be done about this?), and determines what that morally right action would be (e.g. they should be stopped). The latter two mechanisms ('moral intentions' and 'moral will power') are to do with

moral motivation. They are what cause us to be motivated to take actions based on their moral value (e.g. I want to stop them from torturing that kitten), and to consciously choose a morally motivated action over some other action (e.g. I will stop them, even though I might get hurt in the process).

We again see agreement that moral value (e.g. moral judgments) and moral motivation (e.g. actions taken in accordance with moral values) are both considered to be relevant to our moral capacities. I also argued in section 2.4 that moral intuitions are fundamental to determining the moral values we hold, and motivating us to act in accordance with them.

It is worth noting that the way in which we modify our moral mechanisms may be just as important as which mechanisms we choose. Shook writes mainly of enhancements that 'heighten', make 'stronger', give us 'more' of something, such as 'stronger moral decisions'. But it might be argued that moral enhancement might require us to make one or more of our moral mechanisms weaker instead (Kahane & Savulescu 2015, pp. 2-5). For example: I may value honesty, but always telling the whole truth could hurt people's feelings or cause me to disclose confidential information that could be used to harm others.

To answer the question of how to improve moral capacities overall, we will need to decide what the end result of MB should look like. In section 2.5 I argued that motivating broad cooperation is a plausible way to address Harm, and that current human moral capacities do not motivate broad cooperation. Accordingly, I propose the following objective for MB: Moral bioenhancement should aim to modify moral capacities so that they motivate broad cooperation.

Before we discuss the specifics of this proposal, we will look at whether it could be achieved by the ends that Shook and others in the literature have in mind for MB.

Shook considers two possible ends, the 'generic' model and the 'morphic' model (Shook 2012, p. 9):

> Generic: "A moral enhancer might work simply by utilizing one or more mechanisms to increase the frequency and reliability of conduct that the subject believes is moral, by strengthening whatever moral convictions, motives, or habits that person already has."

> Morphic: "A moral enhancer that considerably reshaped a person's moral views and motives [...] would be useful for those who want to add a new moral conviction to their set of moral beliefs, or to entirely reverse an unwanted moral conviction."

Although the generic model involves manipulating at least some of our morally relevant mechanisms, those to do with motivation, I argue that it will not help us to address Harm by itself. Harm is caused in part by the way human moral psychology has evolved to be deficient, and although this deficiency sometimes involves a lack of motivation, at least as important is the problem of which actions we are motivated towards.

For example: An individual might happen to experience moral intuitions that promote broad cooperation, but not to the degree required in order to overcome conflicting factors. A middle-class American, Fred, might be aware of extreme poverty in developing nations, judge it to be morally bad, and be capable of donating money in a manner that would effectively alleviate some of it. This results in some degree of motivation to donate, an action conducive to broad cooperation and which reduces Harm in some small way. However, Fred is more strongly motivated to spend money on a bigger television, a new suit, or renovations on his house. These conflicting motivations have their bases in personal pleasure, not morality. Through biotechnological intervention we might be able to increase the intensity of the moral intuitions that result in Fred's moral motivation to donate the money instead. Plausibly, this could result in that motivation becoming strong enough to overcome conflicting non-moral motivation, and therefore in broadly cooperative behaviour.

Consider another individual, Gertrude. She is also a middle-class American, and is morally motivated to donate money to the same degree that Fred is. She also has conflicting motivations that are stronger. Gertrude's conflicting motivations, however, are moral ones. She is more strongly motivated to spend her excess income on helping her local community, by supplying schools with better equipment, renovating the community centre, and helping homeless individuals. But this motivation is also caused by moral intuitions, in the same way that the other motivation is. Increasing the degree to which Gertrude is morally motivated would plausibly result in her still being more motivated to help her community than to help

developing nations. These actions, while morally motivated, are not broadly cooperative like donating money to alleviate poverty is. Therefore, the generic model will often fail to promote broad cooperation.

Let us now consider the morphic model. This model is more focused on moral value, as opposed to moral motivation. We could conceivably apply a morphic moral enhancement to Gertrude, above, in order to reshape her moral values. We could attempt to remove the bias she has towards providing lesser benefits to the smaller number of people in her community, and increase the degree to which she values the interests of the numerous extremely impoverished citizens of developing nations. This kind of moral enhancement would presumably be far more complicated and difficult than generic moral enhancements, but could potentially be effective in a wider range of cases. But morphic moral enhancements must answer the question: which moral values should be promoted? It is possible that the moral values we promote end up motivating broad cooperation, making the morphic model a better prospect than the generic model for our purposes. It will depend upon the specifics of the MB strategy we adopt. We will turn to this question shortly.

There is a different distinction to be made between possible ends for MB. Kasper Raus et al. provide an overview of various definitions of moral enhancement, and claim that there is a tendency to think of it as involving either 'capacities-oriented' or 'behaviour-oriented' interventions:

> "Some commentators in the moral enhancement debate label a certain intervention a moral enhancement depending on its (real or intended) effect on a person's behaviour. Others see moral enhancements as interventions that target or are intended to target a person's capacities of moral reflection." (Raus et al. 2014, p. 6)

The example given of a behaviour-oriented intervention is one in which a paedophile, unable to understand why there is anything morally wrong about acting on his sexual urges, is sedated against his will and surgically implanted with a chip that stops him from molesting children. Presumably his urges and moral views about them are unchanged, and the chip works by incapacitating him whenever it detects certain behaviour. This intervention results in morally better actions, but does not affect the

paedophile's 'capacities of moral reflection' or, quite possibly, any of the other moral mechanisms.

I claim that behaviour-oriented interventions, understood as interventions that alter behaviour so as to make it morally better without affecting mental processes that we take to be morally relevant, should not qualify as moral enhancement. Tear-gas can prevent someone from committing assault or murder, as can a well-placed bullet. But SWAT teams do not engage in moral enhancement. They prevent people from behaving in certain ways through means that do not engage the moral capacities, usually by incapacitating them. Robert Sparrow makes this point, claiming that MB must act thought 'improved' motivation (Sparrow 2014, p. 24). This motivation must be specifically moral motivation. If an intervention instead motivates certain kinds of behaviour through non-moral means, for example by inducing a painful electrical shock whenever an individual attempts a violent action, its status as moral enhancement should be doubted.

I argue that behaviour is an important part of moral enhancement; successful moral enhancement will involve moral motivation, which is directly related to actions and behaviour. However, unless the motivation is related in some way to our moral capacities, I argue that it has nothing to do with moral enhancement. Some authors arguably include behaviour-oriented interventions in their definition of moral enhancement, and other authors object to MB mainly on the basis that it will predictably function this way. We will look more closely at these arguments in section 4. For now, I claim that behavioural influence alone, isolated from our moral capacities, is insufficient to qualify an intervention as moral enhancement.

Capacities-oriented interventions, understood as interventions that affect solely mental processes involved in moral reflection and moral judgment-making, but which leave moral motivation altogether unaffected, could qualify as moral enhancements. We could also use such interventions to promote moral values related to broad cooperation. However, in many cases, we will also need to increase the degree to which an individual is motivated to act in accordance with those moral values. And we do not seem to be able to find this moral motivation in behaviour-oriented interventions as they are described above.

Therefore I disagree with the behaviour-oriented/capacities-oriented distinction as Raus et al. and others tend to characterise it. Interventions that affect morally relevant emotions should not be considered behaviour-oriented interventions. As I argued in section 2, we cannot completely separate moral intuitions, e.g. moral emotions, from moral reasoning. There is a difference between influencing action through the modification of moral intuitions and doing so through controlling people's actions through non-moral means like a chip that incapacitates them. I claim this distinction also holds for certain kinds of emotional manipulation, but here it is less clear. Regardless, there is something mistaken about the distinction between behaviour- and capacities- oriented interventions. It seems to disregard certain kinds of genuine moral motivation.

3.2: The intuition-satisfaction strategy

There is a definite tendency among those considering the prospect of MB to think about it a certain way: It must result in morally better individuals. But what makes someone morally better than someone else, or morally better than they used to be?

One generally appealing answer is that moral value determines who is morally better than whom; someone that adopts stray kittens produces greater moral value than someone who tortures kittens for fun. However, as I claimed during section 2, moral intuitions are fundamentally important for determining moral value. How do we know that rescuing kittens is morally better than torturing them? Well, it is just obvious! We can intuitively perceive the greater moral value. Reasoning is involved too; something like: It is wrong to cause needless pain, that kitten is experiencing needless pain, and you are causing that to happen by torturing it. But I argued in section 2.4 that moral reasoning comes after moral intuition. In this example, perhaps the moral intuition can be represented by the moral platitude, 'needless pain is wrong'.

If being morally better means being more like how moral intuitions tell us we should be, then the answer to the question 'what does it mean to be morally enhanced?' is: 'to be made more satisfactory to moral intuitions'. This means that successful MB needs to modify an individual in such a way that they are intuitively morally better than they were, or perhaps better than individuals typically are.

I claim that the measure of MB's success, as it is discussed in current literature, is thought to consist in how well the enhanced individuals satisfy current moral intuitions. The specific intuitions that ought to be satisfied vary depending on the account, but we can easily think of the typical platitudes that represent them; for example 'we morally ought to minimise harm and maximise benefit', or 'altruism is morally valuable'.

Because this strategy aims to achieve MB by causing people to think and behave in ways that more effectively satisfy moral intuitions, I call it the intuition-satisfaction strategy (ISS).

Let us look at some examples of the ISS discussed in the literature. Persson and Savulescu have some kind of ISS in mind, despite the fact that their explicit motivation for implementing MB is the entirely practical end of preventing Ultimate Harm. Moral deficiency is the cause of Harm, and is itself caused by deficiency in core moral dispositions such as altruism and a sense of justice:

> "As we see it, then, the core moral dispositions, which are the foremost objects of moral enhancement, are altruism and a sense of justice as it primarily manifests itself in tit-for-tat. By classifying these as moral dispositions, we imply that, by themselves, they *always* issue in a morally correct treatment of the individuals to whom they are directed." (Persson & Savulescu 2012, p. 108)

This implies that the kinds of actions called for to address Harm intrinsically involve acting morally. We can also see indications that P&S's strategy is an ISS in the examples of kinds of behaviour that they think might be susceptible to MB: violence and aggression, racial and sexual biases (Savulescu et. al 2014, pp. 96-98).

Thomas Douglas also has an ISS in mind. He focuses on what he calls "counter-moral emotions"; those which can be uncontroversially morally bad regardless of which moral theory one subscribes to. Remember that I include morally relevant emotions in my definition of moral intuitions, which are related to motivation. Counter-moral emotions are bad because, apart from being obviously bad motives for action in and of themselves, they can also interfere with obviously good motives ("moral emotions, reasoning processes, and combinations thereof"). (Douglas 2008, pp. 230-231)

Douglas uses the examples of racial aversion and violent aggression to illustrate counter-moral emotions. He makes the reasonable claim that reducing racial aversion and violent aggression would sometimes cause some people to have morally better motives than they did. It is not that these counter-moral emotions clearly demonstrate some specified criteria as to the kinds of behaviour MB will target. Instead, they are presented as intuitively obvious examples of immoral motives that MB could potentially affect. (Ibid.)

In more recent work, Douglas writes of enhancing 'moral conformity' (Douglas 2014):

> "An agent fully conforms to morality on a given occasion when she performs an act that is at least as well supported by moral reasons as any alternative act, and she fully conforms to morality over a period of time when she performs a series of acts that is at least as well supported by moral reasons as any alternative series." (Ibid., p. 75)

If there is a link from moral intuitions to moral reasoning as I claim, the 'moral conformity enhancements' Douglas has in mind almost exactly match my description of the ISS.

It might be argued that the above MB strategies are not ISSs as I have defined the term. P&S, for example, are first and foremost concerned with the consequences that might be achieved through MB; they want to reduce the likelihood that Ultimate Harm will occur. They explicitly deride common-sense morality, claiming it is in need of a "thorough revision" (Persson & Savulescu 2012, p. 12). They are not interested in satisfying the intuitions they see as commonplace and responsible for moral deficiency, which contributes to Harm. But this is not because they are not interested in satisfying moral intuitions at all. It is because they are interested in satisfying *different* moral intuitions.

P&S's MB strategy is best characterised as an ISS that emphasises a consequentialist set of intuitions. P&S think that evolved common-sense morality in humans does not promote the satisfaction of consequentialist moral intuitions. It might seem like they do not advocate an ISS, but this confusion comes from the fact that their ISS emphasises some unpopular consequentialist moral intuitions. Arguments against consequentialism are well-known; we can use the moral values of consequentialist

theories to endorse actions that, according to many people, are intuitively morally abhorrent. For example, if five dying hospital patients could be saved by killing a healthy patient and transplanting their organs, consequentialism might seem to endorse doing so (Alexander & Moore 2012, section 1). But defences of consequentialism generally try to explain either why the theory does not endorse intuitively immoral actions, or explain why the consequences of those actions are morally valuable enough to justify them (Ibid.). Consequentialist moral theories do not disregard moral intuitions.

Even if advocates of MB did not have the ISS in mind, their critics certainly do. Many of the most prominent objections to P&S are rooted in the difficulties involved in implementing MB without conflicting with moral intuitions. We will address some of these concerns directly in section 4, but for now simply notice that they conceive of MB as involving the ISS.

Many of John Harris's concerns about MB stem from his belief that what is most important to morality is moral reasoning. We briefly discussed his views in section 2. He does not think that MB could ever be successful, because the most it could do is get people to act in ways that might seem morally better. However, the only way to achieve genuine moral enhancement is through the understanding of moral reasons, which most kinds of MB, such as emotional manipulation, will never achieve (Harris 2013a, pp. 171-172). Not only does Harris take the objective of MB to be making people morally better, but he objects to it on the basis that it will never be capable of causing people to better-satisfy his own moral intuitions, concerning the necessity of a foundation in moral reasoning.

Robert Sparrow suggests a number of possible interpretations of MB, all ISSs; like Harris he expresses concerns about acting for the "right reasons", and considers what this would mean according to the usual realist theories of moral value (Sparrow 2014, pp. 24-25). Richard Dees claims that we can only determine whether prospective MB has truly resulted in moral enhancement if we "already have a moral measuring rod"; we need to know how to make someone morally better in order to improve their moral capacities (Dees 2011, p.13).

Inmaculada de Melo-Martin and Arleen Salles even explicitly state that in the views of MB proponents such as Douglas, Persson, Savulescu, and DeGrazia, the objective

of MB is to result in morally better people (de Melo-Martin & Salles, forthcoming, p. 3). However, they are doubtful that the kinds of measures being proposed could accomplish this.

It is difficult to make clear exactly what I mean by the ISS, because the ISS is never specified in technical terms. It is instead taken as self-evident that the ISS is what we mean when we talk about moral enhancement and MB. As Shook puts it:

> "Upon first hearing, 'moral enhancement' is ordinarily taken to mean something like 'making a person more moral,' which in turn practically means 'more likely to do the morally right thing.'" (Shook 2012, p. 5)

Arguments immediately turn to whether we can make people 'more moral', or morally better, as though we all know what that means. Again, I claim the best way to characterise what those discussing MB take 'morally better' to mean is 'more satisfactory to current moral intuitions'.

ISS-based accounts of MB do not explicitly refer to the objective of satisfying moral intuitions. This is because they take certain moral intuitions to be representative of genuine moral value, in some sense. They talk of the right or moral actions and reasons, and of moral value. What I call a moral intuition, others might call, for example, a moral judgment based on moral reasons. By classifying these judgments as moral intuitions I do not mean to imply that they are nothing more than intuitions in the sense that they are vacuous, or are irrelevant to what is truly morally important. The moral intuitions referred to in the above ISSs may well represent genuine moral value, in whatever way one might take that to mean.

But claiming to understand which moral intuitions represent genuine moral value is the essence of the ISS. If the question of what moral value truly consists in were settled, moral enhancement would be less problematic. The fact that the question is not settled is a major part of many objections to MB, as we have begun to see. MB need not be understood as involving interventions that attempt to make people 'morally better' – that is, more satisfactory to our moral intuitions. If there is another way to understand improvement of moral capacities, we should investigate it.

3.3: The cooperation-broadening strategy

An alternative MB strategy, which I call the cooperation-broadening strategy (CBS), instead identifies Harm as undesirable, broad cooperation as a solution to Harm, and aims to promote broad cooperation through the biotechnological modification of human moral capacities. P&S, and other advocates of MB, tend to be vague when it comes to the question of precisely how MB will address the problems it is meant to address. This is partly because MB is currently theoretical and we do not know what its specific effects will be. But this vagueness also comes from the assumption that making ourselves morally better will reduce and prevent Harm.

However, this assumption has two main problems. First is the question of how to accomplish something as controversial as making ourselves morally better, especially through methods that are themselves controversial such as biotechnology. Second is the possibility that 'making ourselves morally better' will not in fact reduce or prevent Harm. The CBS avoids these two problems.

The CBS starts by identifying Harm as a set of practically undesirable states-of-affairs. Call this the Harm Assumption (HASS):

> HASS: States-of-affairs that might lead to enough harm that worthwhile life is made permanently impossible on Earth (Ultimate Harm) are undesirable and so ought to be prevented. States-of-affairs in which large numbers of living individuals are greatly harmed (Widespread Harm) are undesirable and so ought to be prevented, unless doing so would allow greater harm to occur. Therefore, Harm (a combination of Ultimate Harm and Widespread Harm) ought to be prevented.

This is not meant to be a precise definition. The HASS is only meant to illustrate the kind of thought behind the CBS, and the kind of thought that might justify MB. This imprecision is bound to cause problems, particularly if we try to employ the proviso about 'greater harm' to make fine distinctions.

We must also be content here with little argument for the HASS. It is important that the HASS doesn't rely on meta-ethics for justification; it does not make any moral claims, such as that Harm is morally bad. We could try to justify the Hass with an argument that Harm has negative moral value by appealing to meta-ethical

principles. But moral value is, I have argued, dependent upon moral intuition. If we can find examples that match a moral principle but that do not intuitively seem to have the same moral value, then the moral principle is thrown into doubt. It might turn out that what best reduces and prevents Harm does not intuitively or theoretically seem to be the most morally valuable course of action.

We must instead we rely on it being self-evident that Harm is undesirable and that the HASS is correct. The point is that this undesirability is not meant to have a moral component. It is rather a practical observation. If you truly think that the HASS is false, then you may have no reason to endorse the CBS.

Having accepted that the HASS is correct, the CBS ignores what our moral intuitions tell us about the moral value of actions that will address Harm. It then attempts to modify individuals' moral capacities so that they better motivate actions that will reduce and prevent Harm. As I argued in section 2.5, such actions are those which require broad cooperation among human individuals. The CBS does not advocate making people 'morally better', or claim that all broadly cooperative actions will directly contribute to the reduction and prevention of Harm. If we can cause an increased tendency among humans to be broadly cooperative, that may be enough for the CBS to have been successful.

Here are three important questions: How does the CBS promote broad cooperation? How does it qualify as moral enhancement, if it disregards moral intuitions and makes no moral claims? And how does it overcome current moral intuitions that interfere with the motivation of broad cooperation? These questions all have the same answer: We use biotechnological interventions that cause individuals to experience more moral intuitions that motivate broad cooperation. I will go into more detail regarding this shortly.

In short: CBS-based MB improves our moral capacities by causing our moral intuitions to more effectively motivate broad cooperation.

At first glance, the CBS might just seem to be an ISS with a specific set of moral intuitions pre-built into it; possibly some kind of consequentialist set, with "Harm is morally bad" at the helm. But this is not the aim of the CBS.

The CBS does not make any meta-ethical claims that certain consequences are morally better than others. A consequentialist might end up with an ISS that resembles the CBS by making certain claims such as 'we morally ought to take the actions that bring about the least harm'. This ISS would involve causing people to take the least harmful actions, and this could involve broadening human cooperation. However, this ISS would have to deal with objections that consequentialism does not provide a correct view of moral value. For example: In some borderline cases, preventing harm to one individual or group requires causing or allowing harm to another (Singer 1993, p. 289-313).

In other words, MB founded on consequentialist moral values must still make claims about what is actually morally right. In borderline cases, it is always possible to argue that the most morally valuable course of action is one that undermines broad cooperation. The discussion is still about what is morally right; what best satisfies our moral intuitions. This problem is not exclusive to consequentialist ISSs. For whatever set of moral intuitions we plug into the ISS, we will always be prioritising the meta-ethical theory that claims they represent genuine moral value, rather than prioritising the motivation of broad cooperation.

The easiest way to show the stark contrast between the ISS and the CBS is to bite the bullet with regard to the intuitively morally objectionable measures that might be advocated under the CBS.

Imagine the following scenario: Technology has advanced to the point where we have a record of every individual's genetic structure. We have also isolated a genetic sequence that is strongly associated with broadly cooperative behaviour. We can engineer a virus that is fatal in 100% of cases, but which can only be contracted by individuals who lack the broadly cooperative genes. The individuals capable of deploying the cooperation-virus are not morally bioenhanced, and their moral capacities make them unwilling to deploy the virus. They undergo MB. Their moral capacities are modified in such a way that deploying the virus, which seems as though it will best promote broad cooperation, seems intuitively morally valuable enough that they are now willing to do so. According to the ISS, this MB has been unsuccessful; it caused people to perform an intuitively immoral action they would not have performed before the intervention. But according to the CBS, assuming that

deploying the virus successfully promotes broad cooperation better than any alternatives, the MB has been successful.

I am unconvinced that such extremely intuitively immoral actions would be inevitable if we were to implement CBS MB. It would depend on the specifics of what we mean by 'broad cooperation', and how moral intuitions are modified in order to promote broadly cooperative actions. CBS-based MB may be able to motivate cooperation broad enough that the interests of every individual will be taken equally into account whenever an enhanced individual contemplates an action. Actions that severely harm or compromise the interests of large numbers of individuals, even if doing so may benefit an even larger number of individuals, are arguably not broadly cooperative.

However, we should be aware that the CBS may end up motivating actions that conflict with our current moral intuitions if they are necessary for broad cooperation. Plausible actions would be those caused by reducing bias towards our kin and in-group. For example, imagine that a parent must choose between buying their own children luxuries that will slightly improve their quality of life, on one hand, and spending that money on a community programme to provide free school lunches to impoverished children, on the other. The effects of CBS-based MB may cause the parent to choose the greater benefit for strangers' children over the lesser benefit for their own, when they otherwise would not have. It could be argued that we have a greater moral obligation to our own children than to non-kin, and that CBS-based MB caused the parent to make an immoral choice. This is a minor example; it is plausible that the actual effects could be more morally objectionable. This is probably an unavoidable consequence of the CBS, and should be kept in mind.

3.4: Old vs. new

Put as plainly as possible, here is the difference between the ISS and the CBS: The ISS attempts to make us more satisfactory to our current moral capacities, while the CBS attempts to make our moral capacities more effective at addressing Harm. This is what I mean by improvement of moral capacities. Not 'better' moral capacities from a current moral viewpoint, but moral capacities that offer greater practical benefit than they currently do.

Let us look more closely at the advantages and disadvantages each strategy brings with it.

There is a good reason that the ISS is the first kind of MB that comes to mind: it appeals to our moral intuitions. Putting aside the question of whether it could ever be successful, the objectives of ISS-based MB seem admirable; who would not want everyone to be generally better people? The potential implications have wide-ranging appeal. If criminals were made morally better, they would be less likely to commit crimes like theft, murder, and rape. If politicians were made morally better, they would be less likely to lie to us, enact policies that benefit the privileged few at the expense of the downtrodden many, or allow wars that harm millions. If everyday citizens were made morally better, they would be more likely to selflessly help those who need it, and would contribute more to society.

The reason for this wide-ranging appeal is that any ISS-based MB that we imagine being successful will have, by definition, satisfied our intuitions about being morally better. If it does not, it will not seem to have made anyone morally better, and so will be considered either a failure.

But the ISS has disadvantages. The ISS is simple at first glance: We can reduce and prevent Harm by making people morally better, and we can do that by making them better at satisfying moral intuitions. But this strategy is too simple. It leaves us with a number of difficult but vital questions, two of the most pressing being:

(1) Which moral intuitions do we satisfy to make people morally better?

(2) Will 'making people morally better' reduce and prevent Harm?

Question (1) draws out the most problematic feature of the ISS: It requires us, on some level, to specify the moral intuitions that are the most important. Different societies, and even different socio-economic groups within the same society, can promote substantially different moral values, driven by emphasis on different moral intuitions, as argued in section 2.4. And, as we saw in sections 2.1 and 2.2, there are numerous meta-ethical theories that disagree on what is morally valuable. How can we know whether an intervention is a moral enhancement, according to the ISS, if there can be no agreement on whether it genuinely results morally better people? Shook draws out the problem nicely:

> "Cultures notoriously disagree on many moral matters. For starters, try to conceive of an "etiquette pill" and wonder how it could work the same in Bombay, Baghdad, and Boston. A single morality pill could hardly be less improbable. Even if we set aside the problematic diversity inherent to cultural morality to favor some reasoned ethical system instead, we must choose a system." (Shook 2012, p. 4)

> "There will be endless disagreement over whether modifications are genuinely moral enhancements, and no universal accord on what constitutes moral enhancers should be predicted." (Shook, 2012, p. 12)

If the ISS is to remain true to its objectives, the interventions it involves will implicitly be ruling on which moral intuitions are the ones that represent how morality actually is; the ones that are in some sense correct. As long as MB prioritises interventions that cause people to seem morally better by our current moral standards, it is hard to imagine how it could be achieved without first clearing up all of the disagreement about those current moral standards. If the objective of MB is simply to make people 'morally better', then it will always be debatable whether it has truly succeeded unless we can come to an agreement about what it means to be morally better. I am sceptical as to whether this will ever be possible, and confident that it will not be possible for a long time. In the meantime, moral deficiency will continue to contribute to Harm.

Perhaps moral disagreement is all ultimately dissolvable. It might be possible to identify objective moral values shared universally. The reason for apparent moral disagreement might be that different cultures and individuals are in different circumstances, and respond to situations differently based on the same moral values. Or perhaps some cultures and individuals are simply not educated or well-situated enough to determine genuine moral value. After all, the concept of common moral platitudes is not incoherent.

Shook is more pessimistic, noting both that the prospects we have for shared objective moral value seem insufficient to solve moral disagreement, and that we have so far been completely unsuccessful in justifying any less-universal set of moral values (Shook 2014, p. 10).

I will not attempt to further respond to these optimistic arguments, beyond pointing out that any ISS will still need to do the difficult work of specifying these universal

moral values. Effectively it will have to accomplish the very task that I and others are sceptical about: resolving moral disagreement.

Now we turn to the second most problematic feature of the ISS. The assumption that making people morally better will address Harm is just that: an assumption. In section 1 we took for granted the plausible claim that moral deficiency is substantially responsible for the risk and prevalence of Harm. But it is a different matter to claim that addressing this moral deficiency will also address the Harm that it has caused, and avert the Harm that it has made more likely.

Here is one example. Imagine that ISS-based MB results in a heightened sense of fairness in all individuals. This would intuitively seem to make everyone morally better. Citizens of certain western nations, that have long benefited from high-omission use of fossil fuels, may find it fair that everyone in the world reduce their carbon emissions so that those most likely to be affected by climate change will not be harmed as badly. However, citizens in nations that are still developing, thanks to their relatively recent ability to use fossil-fuels, may find it fair that they continue to do so while other nations who have already benefitted must stop entirely. A heightened sense of fairness may only increase the intensity of this sort of conflict, and does not seem as though it will resolve the MMD issue of global climate change.

We have two primary points against the ISS. The first is that ISS MB will be prohibitively difficult to develop and implement. The second is that, if ISS MB is somehow implemented, we do not have good reasons to think that it will successfully prevent or reduce Harm. A better way of thinking about MB involves a rational understanding of what kinds of behaviour will help us to accomplish our objectives, and the willingness to change our current moral intuitions to help bring it about; the CBS. In Shook's critique of MB he closes with a thought that strikingly reflects the idea behind the CBS, though he clearly does not have the CBS in mind:

> "Modifying conduct in line with what people already regard as ordinary moral behavior can never replace thoughtful adjudication among conflicting moral duties or adjusting our social norms for improving the human condition." (Shook 2012, p. 12)

This brings us to the advantages that the CBS has over the ISS. The first two are simply that the CBS is not as prone to the two primary problems that the ISS has.

The CBS does not require us to resolve meta-ethical disagreement in order to achieve its objective, because that objective is the motivation of broad cooperation rather than making people morally better. And, also because of that objective, the CBS is more likely to be successful in reducing and preventing Harm; broad cooperation is valuable specifically because of its potential to address Harm.

A third advantage is that the harm addressed by the CBS may be broader in scope. Accomplishing the objective of the ISS (i.e. making people morally better) will at best address harm caused by deliberate or negligent moral badness, as Harris points out (Harris 2011, p. 111). But the CBS may provide solutions to harmful problems unrelated to moral deficiency. This is because broad cooperation is first and foremost concerned with practical benefit rather than moral value. Broad cooperation will better enable humans to address all kinds of harmful problems by enabling us to focus our unified efforts towards them. Some examples include cosmic disasters such as asteroids and solar flares, natural disasters such as earthquakes and volcanoes, and outbreaks of lethal diseases.

A fourth advantage involves the problematic relationship between Harm and scientific progress discussed in section 1.8; we need scientific progress to address Harm but, because it is easier to harm than to benefit, further scientific progress is more likely to be overall harmful rather than overall beneficial. The CBS will help address this conflict, as it will promote widely-beneficial scientific advancement of all kinds through broadly cooperative scientific progress rather than the selfish variety that can make money and solidify a nation's power. The CBS may also remove barriers to beneficial scientific progress, making it more likely that scientific progress is overall beneficial rather than harmful, contrary to what P&S predict. Currently, promising avenues of scientific progress are often not pursued because there is no potential for making money from them, or they do not otherwise benefit those with the resources to pursue them. If we could promote broadly cooperative motives, then perhaps we would be able to develop useful technology that otherwise never would have been pursued.

The CBS also has its downsides. The most prominent is one that we have already discussed: the CBS requires us to disregard our current moral intuitions and allow them to be modified according to what will best motivate broad cooperation. The

prospect of this will, by the very definition of moral intuitions, seem morally wrong; at least, morally wrong according to current moral intuitions. Interestingly, however, once CBS-based MB has been successfully achieved, those who have undergone it will no longer find the actions it motivates to be morally wrong. Their moral capacities will have been modified so that the broadly cooperative actions it motivates seem like the morally right ones.

Another objection I can see to the theory behind the CBS is an accusation of hypocrisy. The CBS rests on the HASS; an assumption that Harm should be prevented. If a main problem with the ISS is that we cannot make any assumptions about which moral intuitions should be satisfied, how can we make a comparable assumption about preventing Harm? This would indeed be a problem if the HASS was an assumption about morally valuable consequences. It is instead an assumption about what is practically beneficial; it is in the interests of a huge number of people for Harm to be reduced and prevented. This is a less problematic assumption than that any set of intuitions represent genuine moral value.

The above discussion leaves us with a pressing question with regard to the CBS. With its focus on practical matters and demotion of moral intuitions, does it still qualify as a *moral* enhancement strategy? It could be argued that the CBS is instead a proposal for how to achieve certain practically beneficial consequences regardless of moral considerations. In section 1 we identified the potential to address Harm as a good reason to consider developing and implementing MB. But not just any intervention that addresses Harm automatically qualifies as MB, or more generally moral enhancement. What could be lacking in the CBS that excludes it from the category of moral enhancement?

One possibility is that the CBS ignores moral reasoning, and therefore can never result in genuine moral enhancement. We will look at objections along these lines in section 4.

For now, however, note that the CBS does involve reasoning. Reasoning is our only guide to the actions we practically ought to take in order to achieve specified ends; we will need to use reason to determine which moral intuitions will motivate broad cooperation. But this reasoning is not moral reasoning. It is practical reasoning, to determine which actions will bring about specified ends.

One other way the CBS may fail to qualify as moral enhancement is through its moral counter-intuitiveness. The broadly cooperative actions it might motivate can conflict with our current moral intuitions, making them seem immoral. Moreover, even implementing the CBS, intentionally modifying our moral intuitions, seems immoral. But, as I have argued, there is no universal agreement about which moral intuitions represent genuine moral value, and we should not expect further examination of our moral intuitions to reveal genuine moral value. The same actions do not seem like the right ones to everyone. This is because we have no particularly good reasons to think that our moral intuitions (i.e. what feels morally right) are a good guide to what actually is morally right:

> "The problem, I think, is that we've been looking for universal moral principles that feel right, and there may be no such thing." (Greene 2013, ch. 1)

Instead, we should make the actions that are of the most practical value the ones that seem morally appealing. As long as we are unwilling to go against our current moral intuitions, which help to cause the moral deficiency which contributes to Harm, we are unlikely to achieve worthwhile MB.

The main point that qualifies the CBS as a moral enhancement is that it achieves its objectives through the effects it has on our moral capacities. As I argued in section 3.1, we cannot call just any intervention that produces certain desired behaviour a moral enhancement. It needs to produce that behaviour through moral motivation; our actions need to be motivated by our moral capacities. The CBS determines the type of behaviour that would be most beneficial, through cold, non-moral reasoning that ignores moral intuitions (broad cooperation, because it will address Harm), but motivates that behaviour by modifying our moral intuitions. The CBS results in improved moral capacities, but does not make us 'morally better' by current moral standards.

3.5: Implementing the CBS

I have remained vague so far about what the CBS would entail. I am not familiar enough with the relevant science to suggest the best ways to bring about the necessary changes in our moral intuitions. Regardless, I will try to be more specific

about some prospects for implementing the CBS that seem not to be too farfetched. It is beyond the scope of this thesis to predict what the effects of specific measures will be. We will start with the effects we should try to achieve, before moving on to the interventions that might be capable of achieving them.

The CBS aims to motivate broad cooperation by manipulating the moral intuitions humans tend to experience. Why is this?

Even if we agree that the satisfaction of moral intuitions should not be an objective of MB, we cannot simply discard our moral intuitions. The influence that our moral intuitions have over our actions cannot be easily ignored. Moreover, why must we be in constant conflict with our intuitions? Evolutionary theory suggests plausible explanation for why we have them, one which does not endow them with objective moral authority. We have good reasons to think that moral intuitions are more relevant to our moral values, and are better at motivating our actions, than cold reasoning. And we can plausibly modify which moral intuitions we experience through the use of biotechnology; they are psychological factors, and psychology is unavoidably influenced by biology.

We should try to make our moral intuitions work for us, as they did for our ancestors, but against modern problems rooted in scientific progress rather than the problems that originally drove the evolution of moral intuitions. This, plausibly, means motivating broad cooperation rather than the formerly evolutionarily advantageous behaviour motivated by current moral intuitions. This will allow us to rely on them to motivate actions that help prevent Harm. Not only that, but it will also mean that the broadly cooperative actions, which address harm, and our moral intuitions, which currently tell us that such actions are morally wrong, will no longer be in conflict.

In section 2.5, I argued that current moral intuitions fail to motivate broad cooperation primarily because of the way that they evolved through competition between groups. Many innate cooperation-promoting tendencies exist in humans, but only function at the level of in-groups. Out-groups (individuals perceived significantly different from those in one's in-group) seem deserving of different treatment according to our moral intuitions. If we could find a way to broaden moral intuitions to include the whole human population as an in-group, broad cooperation could be achieved.

One way to do this might be to find a big and different enough out-group for the whole human race to compete with, such as an intelligent alien species. This might allow most people to broaden their perception of in-group to include all humans, as opposed to the out-group of the aliens. But this is farfetched, and would only trade the problems caused by failure to cooperate with one another for those caused by failure to cooperate with the aliens.

We could instead aim somehow to directly modify which moral intuitions humans tend to experience so that the aspects of them shaped by inter-group competition are diminished. This could plausibly involve radical changes to some moral intuitions that are widely considered of central importance to morality, two of the most notable being bias towards kin and reciprocity.

Moral intuitions are complicated, so it is difficult to say whether modifying them directly is even a realistic prospect. We will have to turn to science to explain the physical brain processes involved in the psychological factors that constitute our moral intuitions. Probably the closest specific measures we have seen suggested so far are in the context of Thomas Douglas's ISS; we can modulate emotions using biotechnological interventions, and many of our moral intuitions are based on morally-relevant emotions.

However, we may not need to directly apply MB to each individual in order for the CBS to be successful. This is because moral intuitions are susceptible to the influence of others in one's peer group. Modern research suggests that new societal norms can be propagated by some individuals prominently modelling behaviour (Harré 2011, pp. 35-52). This kind of social norm-propagation means that moral intuition-modifying MB could be used as a catalyst for broadly cooperative norms; if MB results in some people who experience cooperation-broadening moral intuitions, others they interact with may be influenced to also experience them.

The above could be supported by Haidt's social intuitionist model discussed in section 2.4; it suggests strong links between reasoning supplied by others and influence on moral intuitions, so if people observe and rationally discuss norms there is potential for broad cooperation-promoting moral intuitions to be transmitted between them.

Let us imagine that we have decided modifying individuals' moral intuitions in order to promote broadly cooperative behaviour is the best way to implement CBS-based MB. We still need to figure out the specific biotechnological interventions that might achieve this.

Drugs are the most familiar option. One commonly mentioned in the context of MB is oxytocin. This hormone is released naturally in mammals in response to a variety of stimuli, and is an integral component of intimacy and sexual reproduction. But it is also associated with social recognition, pair-bonding, and maternal behaviours, among other social behaviours. Because of its positive associations with behaviours thought to be morally significant, "such as trust, sympathy, and generosity", it has been proposed as a moral bioenhancer. Experiments have shown that it increases cooperation in certain contexts. (Savulescu et. al. 2014, pp. 93-94)

However, experiments have also suggested that oxytocin is effective in the above ways only between individuals who consider one another to belong to the same in-group, and may have opposite effects with regard to outsiders (Ibid.). This would clearly be ineffective for promoting broad cooperation, as it exacerbates the problems caused by the in-group bias of our current moral intuitions.

Another possible drug-based MB is psilocybin; the psychoactive compound found in several species of mushroom. It has been shown to result in persistent "altruistic/positive social effects", morally relevant changes that appear to be deeply incorporated into the personality (Tennison 2012, pp. 411-413). We can find promising trends among discussions of psychedelics such as psilocybin, indicating that they result in expanded states of awareness from the personal to the universal. They can promote a sense of 'ecological awareness', of everything being 'interconnected', and a perspective of a 'collective' (Dickins 2013). These are plausibly the kinds of state of mind that could result in altered moral intuitions that are able to overcome self-interest and in-group bias, motivating broad cooperation.

In the future, genetic modification may provide means to produce broadly cooperative tendencies in individuals. Our genes are responsible for biological factors that influence our psychology, and this includes moral mechanisms. It might be theoretically possible to apply such interventions to adults. But the better option would be interventions that act upon unborn human embryos. Rather than altering

existing human individuals, we can attempt to ensure that future generations are born predisposed to be more moral than previous ones.

Halley Faust poses a thought experiment in which a theoretical genetic trait has been identified, the MoralKinder haplotype (MK+). This haplotype:

> "...gives a high probability in any particular circumstance that a child will have enhanced moral decision-making, act more virtuously, and produce a kinder and more empathic person." (Faust 2008, p. 399).

Preimplantation genetic diagnosis (PGD), prior to conception via IVF, allows us to test prospective embryos for genetically-linked conditions such as Huntington's disease and implant only those that do not carry the genes linked to such conditions. Faust theorises that similar methods could be used to ensure that embryos that carry MK+ are conceived instead of ones that do not.

If we could identify something similar to MK+ that instead indicates individuals in which it is present will be inclined to experience broadly cooperative moral intuitions (call it BC+), this could be of use to the CBS. Depending on how advanced genetic analysis technology becomes, detection of BC+ could be a simple matter. We could test all prospective IVF embryos through PGD and implant only those that carry BC+. We could test all sexually conceived embryos and abort those that do not carry BC+. Perhaps one day we could even genetically modify every embryo to ensure that they carry BC+. This would plausibly result in new generations being genetically predisposed to experience broadly cooperative moral intuitions.

There are at least two main problems I can see with MB through biotechnological embryonic interventions. The first is that bioethics is full of objections to many kinds of interventions that target human embryos. Issues of PGD and abortion are some of the most controversial topics in ethics, as are any interventions intended to produce infants possessing desired traits and lacking unwanted ones. There are some good reasons for this controversy; biotechnological embryonic interventions have the potential for abuse and discrimination. For example, should we allow prospective parents to abort or genetically modify their unborn baby, based on personal preferences about the traits they will possess? We could prevent the baby from contracting a harmful genetic condition, but also prevent them from having a certain sexual orientation, or being a certain gender, or perhaps even being predisposed to

have certain talents. We cannot go further into these issues here, but note that they may incite additional opposition to MB.

The second problem is that even if we can genetically modify or select for offspring so that they carry BC+, their moral capacities will be shaped by their environment as well as their genetics. For example, if a child is genetically predisposed to experience the moral intuition that their family is no more morally important than the rest of their society, but their society instils in them the importance of family through TME, they will likely experience conflicting moral intuitions. The influence of society on moral capacities is likely to be a factor in any MB attempt, but in this case it is even more prominent; individuals' moral intuitions are likely to be more easily influenced by society during their developmental period than they are during adulthood. If enhanced new generations are raised by unenhanced old generations, the influence of old moral intuitions will likely interfere with MB. CBS-based MB will therefore work best if combined with traditional moral enhancement designed to supplement it. We will discuss this possibility in section 4.

3.6: Summary of section 3

In this section we tried to clarify what people usually mean when they talk about moral bioenhancement, and whether that concept is anything like what MB should be. We stuck to our original definition, 'improved moral capacities', and considered what this might mean. We considered possible ends for MB, and the moral mechanisms that could be modified in order to achieve those ends. I suggested the motivation of broad cooperation as an end.

We saw how the typical MB strategy involves interventions that aim to make people morally better; to be more satisfactory to some set or other of moral intuitions. Call this the intuition-satisfaction strategy (ISS). Advocates of the ISS argue that this will reduce or prevent Harm, because making people morally better will address the moral deficiency that contributes to it.

I argued that the ISS would be difficult to implement, because in doing so we would inevitably favour some moral intuitions over others; in order to choose the right ones we would first have to resolve the majority of moral disagreement. I am doubtful that

this task can be accomplished, especially in the near future. There is also a substantial chance that the ISS would not reduce or prevent Harm, because it does not make this task its direct objective.

I suggest an alternative MB strategy: the cooperation-broadening strategy (CBS). It attempts to improve our moral capacities by making them more practically useful to us; by modifying them so that they motivate broad cooperation. The CBS does not make any moral claims about the value of actions, so does not need to solve meta-ethical disagreement. It also sets the motivation of broad cooperation as its objective, which I have argued will address broad cooperation better than the vague objective of making people morally better. But the CBS is morally unpalatable, because it disregards current moral intuitions.

Therefore, calling the CBS moral enhancement might be questionable. I argue that this is only from the perspective of ISS-based moral enhancement; if an intervention can only qualify as moral enhancement by making us seem morally better according to our current moral intuitions, we will never be able to use it to produce the radical changes called for to address Harm. Because the CBS alters our moral capacities so that they are more useful to us, i.e. so that they motivate broad cooperation, I claim that it improves our moral capacities in a different sense and so can still be called moral enhancement.

We then considered, if we were to develop and implement CBS-based MB, how specifically we should go about doing so. We need to identify biotechnological means through which we could modify individuals' moral intuitions. Drugs are a possibility, particularly psilocybin, but it may be difficult to find a drug-based intervention that can be effectively applied broadly enough to achieve the CBS's objectives. I am more optimistic about embryonic interventions, which could ensure that members of future generations are more inclined towards broad cooperation.

Although this section discussed the benefits that might be attained through the CBS, there remain some serious concerns about prospective MB in general. Although the CBS is less prone to some of these concerns than the ISS, we must still discuss them in order to determine whether the CBS's benefits are worth the potential risks that come with it. This will be the task of the final section of this thesis.

4: Problems

Objections to MB tend to fall into two categories, as noted by Nicholas Agar (Agar 2015, pp. 38-39). The first kind of objection, 'in-principle', criticises the theory of MB. No matter how we go about it, MB would be bad. This is because the concept of improving human moral capacities through biotechnological means involves some kind of incoherency; whatever we accomplish through our MB attempts could never be genuine moral enhancement. The second kind of objection, 'in-practice', allows that successful MB is at least theoretically possible. If the real world were different than it is, it might be possible to improve moral capacities through biotechnology. Unfortunately, as matters actually stand, attempts to achieve MB will "predictably lead to bad outcomes" (Ibid., p. 28). Keep in mind that these distinct kinds of MB objection require different kinds of responses. It is no use responding to an in-practice objection by pointing out that, under hypothetical circumstances, MB would be successful. The objector may already agree.

This final section will discuss several established objections to MB. We will briefly acknowledge several less-important objections before fully discussing three primary ones; one is an in-principle objection, and the other two are in-practice objections. Although none of these objections specifically target CBS-based MB, they can nevertheless apply to it. However, the CBS has advantages over the ISS when it comes to responding to these objections. Therefore, in this section we will compare the ISS to the CBS in the context of each objection. We will see that the CBS can respond to these particular objections better than the ISS.

4.1: Disregarding less-important objections

We will start by briefly acknowledging a few well-known objections and discarding them without much argument.

One of the most prominent objections to MB is one we have already discussed: we cannot agree on what it means to be made morally better. Not only do moral values differ between cultures and societies, they also differ within the same society. If an intervention causes an individual to disregard the interests of their immediate family in favour of the interests of their country at large, for example, I predict you could

find just as many people who would call that individual morally better than they used to be as those who would call them morally worse. If there are objectively true moral values, and we need to know what they are before MB can be achieved, MB might be prohibitively difficult. If there are no objectively true moral values, making someone objectively morally better and therefore achieving ISS-based MB might be logically impossible.

This objection nicely explains the biggest problem with the ISS. In summary, it asks: whose moral intuitions do we satisfy? Since the ISS's primary objective is the satisfaction of the right moral intuitions, it will first have to answer this difficult question. It seems unlikely that this can be accomplished, as I have argued. But, as discussed in section 3, the CBS is well-equipped to deal with this problem, so we can discard it here without further examination. The CBS does not make any claims as to which moral values should be promoted; only which actions are broadly cooperative and which moral intuitions will motivate those actions. This defence does result in the problem of moral counter-intuitiveness, but as we have seen the CBS must simply bite that particular bullet.

A commonly-discussed objection to MB is that it threatens to undermine freedom or autonomy. John Harris is the most prominent representative of this objection. The idea is that we cannot be truly free unless it is literally possible for us to take actions that are morally wrong.

> "Autonomy surely requires not only the possibility of falling but the freedom to choose to fall, and that same autonomy gives us self-sufficiency" (Harris 2011, p. 103)

Under some understandings, MB would attempt to remove that possibility; successful MB could involve an intervention that ensures that the enhanced individual always takes morally right actions. This eliminates the 'freedom to fall' that Harris argues has moral value.

We will discard this objection without much discussion for two reasons. One, others have already discussed it at length (Savulescu et. al 2014, pp. 99-108). Two, I do not think it is plausible; it stems from a problematic concept of freedom as requiring that, whatever action an individual chooses to take, it is possible that they could have done otherwise.

We cannot discuss the metaphysics of freedom and autonomy here, but it is a simple matter to compare a naturally morally virtuous person with someone who has been bioenhanced to be just as virtuous. Neither of them could ever bring themselves to choose to take an immoral action. Why should we think that the unenhanced individual is any less free than the enhanced one?

> "The sense in which it is 'impossible' for morally bioenhanced people to do what they regard as immoral will be the same as it is already for the virtuous person: it is psychologically or motivationally out of the question. People who are morally good and always try to do what they regard as right are not necessarily less free than those who sometimes fail to do so." (Persson & Savulescu 2013, p. 128)

Moreover, the CBS may entirely avoid this objection. It does not aim to make people unable to take immoral actions. It might not even make people unable to take non-broadly cooperative actions. All it needs to do is motivate people to take broadly cooperative actions enough of the time for Harm to be addressed. Even if 'enough of the time' ends up meaning '100% of the time', it is not clear that always choosing to take certain kinds of actions necessarily makes someone unfree.

However, there is a sense in which I think freedom and autonomy could be legitimately threatened by MB, which Harris is also concerned about: MB might have to be mandatory. This is a more interesting objection, so we will give it the full discussion it deserves later in this section.

Our final less-interesting objection applies to the ISS, but possibly even more prominently to the CBS. The broad concern is that MB will be bad for those enhanced.

MB might be abused to control the behaviour of populations, leading to oppression. Immoral authoritarian regimes could force interventions on their populations, preventing anyone from dissenting to moral values that are beneficial for the regime. To make the people work hard and never complain, simply make that seem like the most moral way to be and remove their willingness to be any way but the most moral. A docile population can serve the interests of the ruling few at the sacrifice of their own interests (Sparrow 2014, p. 28).

Relatedly, certain kinds of MB, including the CBS, may make enhanced individuals more vulnerable to exploitation by unenhanced individuals. Masahiro Morioka suggests that when morally enhanced individuals interact with morally non-enhanced individuals, the latter will benefit at the expense of the former. Morioka's example is an overcrowded lifeboat, which will sink unless one person is cast overboard to die. If only one lifeboat occupant has undergone MB, they would volunteer to sacrifice themselves. This may be because MB causes them to experience moral intuitions motivating actions in collective-interests over self-interest. This specific case may not be overly concerning. But Morioka predicts the use of MB to dominate, use, and exploit as slaves those who undergo it. (Morioka 2014, pp. 122-123)

Oppression and exploitation concerns are of particular relevance to the CBS. We have reasons to think that our current moral capacities evolved specifically to allow cooperation to occur while preventing uncooperative free-riders from exploiting co-operators. We discussed in-group bias and tit-for-tat reciprocity in section 2.5. Motivating broad cooperation could involve promoting moral intuitions that undermine these psychological mechanisms that prevent cooperation from collapsing due to exploitation by free-riders.

Note that these potential problems are caused by a lack of broad cooperation. Self-interested authorities and free-riders are not concerned with collective interest, particularly that of those outside their in-group. These are precisely the problems that CBS-based MB aims to address, and only arise when enhanced individuals interact with non-enhanced individuals. Therefore there is an obvious way that they could be avoided; make sure that every individual is enhanced. This brings its own problems, which will be discussed later in this section. For now we will move on to full discussions of some other objections.

4.2: Right and wrong, reasons and understanding

The MB strategy proposed by Thomas Douglas involves direct modulation of an individual's moral emotions, leaving them with morally better motives than they had before (Douglas 2008). This is something that seems to be most plausibly attainable through the use of bio-technology, i.e. drugs that act on our psychology to affect how

we feel, or genetic modification to manipulate our emotional behavioural impulses. Moral emotions produce moral motivation. Drugs that alter the emotions we experience could therefore allow us to manipulate the motives we have, to choose only the morally best motives. The CBS shares similarities with this strategy, advocating the direct manipulation of psychological traits in order to motivate certain behaviour. However, where Douglas seems to advocate manipulating emotions so that they motivate behaviour more in-line with other moral intuitions, the CBS places the emphasis on manipulating moral intuitions more generally in order to motivate broadly cooperative behaviour.

John Harris objects to the notion of moral enhancement via direct emotional manipulation. His primary objection is that no matter how morally good the consequences produced through MB, anything that does not achieve these consequences through the right effects on our moral capacities cannot be called a successful moral enhancement:

> "It seems to me that moral enhancement, properly so called, must not only make the doing of good or right actions more probable and the doing of bad ones less likely, but must also include the understanding of what constitutes right and wrong action." (Harris 2013a, p. 172)

Harris is of course thinking of ISS-based MB. He claims that cognition and rational thought are essential for true morality. No action can have genuine moral value unless freely chosen, for reasons that the agent understands and accepts. Consider a form of MB that involved the modulation of emotions that produce morally bad motives such as racial aversion. If we apply MB to a racist individual, who consequently no longer feels fear and disgust that they used to when in the presence of African-Americans, that individual may act in ways that are morally better (according to common moral intuitions). If the enhanced individual is a judge, to take Douglas's example (Douglas 2013, p. 161), they may stop giving unfairly harsh sentences to African-American criminals.

But these morally better actions were not motivated by the right sorts of moral capacities, Harris might argue. The judge was not given reasons why harsher sentencing is wrong, which were subsequently understood, accepted, and acted upon. The motivation to give harsher sentences was instead directly removed through

emotional manipulation. The only way to make the judge truly morally better would be by making him understand, through rational discourse, the reasons why African-Americans should not be given harsher sentences. This might involve getting him to agree that people whose crimes are relevantly similar should receive equal sentences, and then by showing him comparisons between his past cases in which unequal sentences were given. Perhaps the judge could also be encouraged to interact and empathise with African-American people, in an effort to counteract the emotional aversion he experiences toward them that interferes with the above reasoning. The result is the same as the one achieved through MB: the judge stops giving unfairly harsh sentences and experiences a moral intuition that motivates this action. But perhaps it is only through the influence of reasoning that the judge can be said to have improved morally.

Fabrice Jotterand describes similar concerns in a more technical fashion:

> "...the way proponents of moral bioenhancement conceptualize morality is problematic because it does not give careful attention to the nature of human moral psychology and to how we make moral judgments (the combination of moral emotions and moral reasoning). They reduce morality to the manipulation of human behavior (the capacity to respond morally) and fail to address the question of the content of morality including beliefs, ideas, values, conceptions of the good and the just, and human flourishing" (Jotterand 2014, p. 1)

One thing that MB cannot do is improve our moral capacities by explaining moral reasons to us so that we understand and accept them. If true moral enhancement requires this, then MB of any kind will never be successful.

Thomas Douglas characterises these objections as 'superficiality concerns', and sums them up with reference to Harris:

> "We could instead take Harris' claim that the direct modulation of emotions could not produce 'moral enhancement properly so-called' to be the claim that, although such modulation could increase moral conformity, there is some deeper variety of moral improvement that it would not produce, or, at least, that it would produce to a lesser degree than the more traditional ways of improving moral conformity that Harris favours. Thus, it would result in a kind of moral improvement that is,

> in one respect at least, more superficial than that produced by these more traditional means." (Douglas 2014, pp. 77-78)

Many superficiality concerns may simply be objections to behaviour-oriented interventions, as described in section 3.1. This would be the argument that such interventions should not qualify as moral enhancement, which I agree with. But I do not agree that the manipulation of moral intuitions, including morally-relevant emotions, is the same as influencing an individual's behaviour through non-moral means such as administering pain or incapacitating them with tear-gas. If my arguments are correct, then this understanding of superficiality concerns does not pose a problem for either the ISS or the CBS.

Beyond this, there are two concerns that might be expressed by the above kinds of objections:

(1) Understanding moral reasons behind moral motivation is just as important as, or more important than, the motivation itself, if we want our actions to have genuine moral value. TME entails understanding moral reasons, while MB does not, so TME results in less superficial moral enhancement than MB. (2) We cannot afford to depend upon the kinds of factors that could be modified by MB to inform us as to what is objectively right and wrong. We need purely rational support for our moral values and judgments. Otherwise, we cannot know that we are being motivated towards the action that is truly the right one.

These concerns may be, but are not necessarily, related. Perhaps we must understand moral reasons so that we can rationally assess whether we are justified in thinking that we hold objectively true moral values. However, even if there is no objectively true moral value, it might still be important that we understand the moral reasons behind our moral motivation; arguably I am not acting morally if I just happen to do what I am inclined to think is right without caring at all about whether or why it might actually be the right thing to do.

Both objections could be problems for the ISS, but we have space only to briefly mention why. On at least some views, such as Harris's, if we want to make people genuinely morally better then they need to act in the right ways for the right reasons. Interventions such as those proposed by P&S and Douglas will cause people to act morally better through causing them to understand and accept moral reasons.

Therefore according to Harris's moral intuitions, at least, such interventions do not result in morally better people. It also does not seem as though we can depend upon the direct manipulation of moral mechanisms to result in objective moral value unless we can resolve moral disagreement, which the ISS struggles to do.

Objection 2 is not a problem for the CBS, because the CBS does not intend to motivate actions that are objectively morally valuable. It aims to motivate broad cooperation, which will address Harm.

Objection 1 may be a problem for the CBS, depending upon controversial facts about meta-ethics and moral psychology. If my account of moral capacities emphasising moral intuitions is correct, then objection 1 seems to be asking too much of MB. Morally relevant emotions and other automatic responses, the kinds of factors Harris and others object to, are never entirely separable from our moral reasoning and values. The judge in the above example will end up endorsing the same moral reasoning whether he is convinced of that reasoning first, which ends up altering his moral intuitions, or has his intuitions altered first, which provides basis for future moral reasoning. TME still aims to influence moral intuitions, but takes a different route than CBS-based MB does.

However, my account of moral capacities might be mistaken. If it is the case that moral reasoning is entirely separable from moral intuitions, CBS-based MB might cause individuals to behave morally and experience broadly cooperative moral intuitions, but be unable to rationally endorse those intuitions. In this case it is plausible that an individual who acts and intuits the same as another individual, who also understands supporting moral reasons, will be motivated by factors that are not related to moral capacities. The distinction is complicated. In response I can only appeal to the arguments I made in section 2, which are admittedly controversial.

But even if one or both of the above objections is correct, the CBS would still have achieved its stated objective: the motivation of broad cooperation. And therefore it could still effectively address Harm. But because it would not motivate broad cooperation through moral capacities as I claim it does, it could lose its status as a moral enhancement. It would instead be a behaviour-oriented intervention, which I am unwilling to classify as moral enhancement. Moreover, if the actions motivated

by the CBS conflict with objective moral values, it might mean that the CBS is objectively immoral.

The superficiality concern, therefore, could be a problem for the CBS in that it could undermine its status as MB. I argue that this is not the case. But even if it is, the CBS could still have practical value; it could still address Harm. Whether that would make its implementation worth the potential moral cost is another question. But the superficiality concern is a different sort of problem for ISS-based MB because it could undermine the ISS's ability to make people 'morally better', which is its objective. This part of the problem does not apply to CBS-based MB.

4.3: Moral complexity

I have argued that moral intuition is the most fundamental part of our moral capacities, and the part that MB should attempt to modify. But that does not mean moral intuitions are a single, simple type of influence that can be modified with precision by MB. Interventions that modify any part of our moral capacities, especially our moral intuitions, could produce unexpected and negative consequences if we do not know exactly what the effects will be, and it might be unreasonable to expect that we ever will. Agar sums up the concern:

> "One feature of such interventions makes them dangerous. They are piecemeal interventions. They target specific psychological influences on moral judgment. This piecemeal approach differs from the means by which moral improvement typically occurs. When successful, moral education typically strengthens many contributors to moral thinking. A change in your affective responses is typically accompanied by changes in the moral reasoning you perform. It is the piecemeal nature of changes to moral inputs produced by moral bioenhancement that makes it dangerous." (Agar 2015, p. 40)

Because MB will presumably be of this piecemeal variety, there is a risk that any such intervention will disrupt the complicated relationships between our moral mechanisms that allow our moral capacities to function normally. This could result in unintended and intuitively immoral judgments, motivation, and actions.

For example, boosting an individual's empathy may strengthen morally relevant emotional bonds with their own children to such an extent that they are willing to impose excessive sacrifices on strangers to obtain benefits for their own children. Conversely, a different intervention might boost an individual's motivation to act in accordance with a reasoned utilitarian moral principle, one which places no greater value in the interests of their own children than in the interests of strangers. Agar's example: A medical scientist who receives such an intervention may morally endorse testing experimental HIV treatments on their own child, speeding up progress on the cure but requiring that they infect their child with HIV (Ibid., p. 43).

Normal moral capacities involve a delicate balance of moral mechanisms. A single piecemeal MB intervention could result in moral values and motivations that are intuitively immoral, because certain mechanisms have been elevated to the point where other equally important ones are neglected. MB is dangerous because it risks causing moral motivation that leads to intuitively immoral acts such as the examples above.

Harris holds related concerns about the complexity of our moral capacities. In his extensive debate with Douglas about the proposition of direct modulation of moral emotions, Harris points out that some emotions (e.g. strong aversion) might cause a person to act immorally in some contexts (e.g. to commit assault) while being appropriate or perhaps morally required in other contexts:

> "Could we in short have the sorts of feelings that are appropriate and indeed, it might be argued, necessary to morality, if we did not feel a strong aversion to, for example, someone who deliberately and unjustifiably killed or tortured those we love?" (Harris 2011, p. 105).

Sparrow, too, points out that moral behaviour is context dependent:

> "Even drugs [...] which encourage feelings of compassion and empathy, will lead people to behave *less* morally in many circumstances where acting morally requires agents to follow abstract moral principles or respond to the demands of distant others rather than those close to them." (Sparrow 2014, p. 25)

Prohibitive moral complexity concerns are expressed in different ways by various critics of MB. Chris Zarpentine suggests that the effects of MB will likely interact with

environmental factors, making it unpredictable and difficult to achieve desired results with (Zarpentine 2013, p. 145-146). de Melo-Martin and Salles give the example of racial aversion; the emotions associated with it, such as fear, are not uniformly negative or morally wrong. Fear is complex, and morally appropriate in many contexts:

"...it is implausible that a biomedical intervention would be so finely tuned as to be able to reduce fear to a particular object only." (de Melo-Martin & Salles, forthcoming, p. 8)

Some of these thoughts tie into part of the superficiality concern; an intervention that affects behaviour alone can never, in-principle, result in genuine moral enhancement. But its more interesting aspects are related to an in-practice objection.

The moral complexity objection is undoubtedly a valid concern, especially for the ISS. If our goal is to make individuals overall morally better, we cannot use interventions that only produce moral intuition-satisfying behaviour in certain contexts, whether or not that behaviour is motivated by the right kinds of moral capacities. With enough fine-tuning on an individual level through the simultaneous use of different interventions, it might be possible to modify moral capacities so that they satisfy moral intuitions under almost every circumstance. But I ultimately agree it is more likely that ISS-based MB would have to work in the problematic piecemeal fashion predicted by Agar, if it is to result in intuitively morally better people in all situations.

Under the CBS, there is at least one response available. However, I am doubtful whether it will seem satisfactory to anyone with currently normal moral capacities.

The CBS specifically intends to modify current normal moral capacities; it should not be a surprise if, after undergoing CBS-based MB, an individual begins acting in ways that would have seemed morally prohibited or unnecessary before. If the individual still acted in accordance with the satisfaction of normal moral intuitions, the CBS would have failed. This is because current moral intuitions are the root of moral deficiency.

Imperfect CBS technology might fail to motivate broad cooperation through this intuition manipulation, because the technology required to do so may have to be as

complex as our moral capacities. Agar and others are doubtful as to whether such complex technology will ever be successfully developed. This is a valid concern that I cannot address without a full understanding of the relevant science.

4.4: Universal and mandatory moral bioenhancement

Plausibly, MB will only be effective if practically every individual receives it. Call this universal MB. To address IMD-caused Harm, we may need to ensure that not even a single morally depraved individual remains who may be capable of causing it; as technology advances, it will become easier for a single individual to cause enormous amounts of harm. To address MMD-caused Harm, the vast majority of individuals will need to be enhanced; whole societies are responsible for this kind of Harm, and if only a few individuals within a society have improved moral capacities they may not have enough of an influence.

Some objections to MB become less troublesome if we assume that MB will be universal. One example is that if everyone becomes a more effective broad co-operator, without exception, there will be no moral disagreement about the moral value of broadly cooperative actions. There would also be no moral intuition conflicts, either with those of other people or with broadly cooperative actions. Universal MB would also ensure that exploitation and oppression of the morally bioenhanced by uncooperative free-riders would be non-existent. This is because no one would be willing to free-ride after undergoing MB; everyone would be inclined to think first of collective interest rather than self-interest, to refuse testifying in the prisoner's dilemma.

The main problem with universal MB is that it would be unrealistic to suppose that a majority, or even a substantial minority of the human population, would volunteer to receive it. The people most likely to cause of IMD-related Harm are political leaders, extremists, and desperate or idiotic or exceptionally harm-inclined individuals; the powerful and self-interested, or the morally depraved. Some of these people are the least likely to volunteer for MB, and also the most difficult to force it upon because of their elusiveness or their positions of authority. On the other hand, the people responsible for MMD-related Harm, members of general populations, would also be unlikely to volunteer for MB. This is because the normal moral intuitions that fail to

prevent them from contributing to Harm are integral to their societies and cultures, and therefore do not seem to those people to be moral deficiencies. Moreover, an intervention that modifies those moral intuitions will obviously conflict with them, and will therefore seem immoral.

Therefore, if MB needs to be universal, it arguably also needs to be mandatory; it would have to be forced upon people. Call the concept of universal, mandatory MB 'moral bioenforcement'[4].

Concerns that MB would undermine freedom and autonomy seem more valid in response to the need for moral bioenforcement, rather than to the effects it may have on psychology as briefly discussed in section 4.1. Moral bioenforcement is arguably an impermissible violation of democratic principles that protect the freedom we consider ourselves to possess in most developed societies. Freedom to decide what medical interventions are enacted upon us, to refuse to put substances into our bodies that will have significant biological effects, is plausibly more valuable than any benefits to be gained through MB.

Even if we permit moral bioenforcement, regardless of the costs to freedom, there still remains the significant challenge of universally implementing it. It can be argued that it would be impossible to implement MB universal enough to be effective and avoid causing problems. Harris argues that MB would have to be "universal and exceptionless", that it would not be effective otherwise, and, as we know from experience with simple interventions such as the polio vaccine, "that it would be impossible to ensure anything like universal coverage" (Harris 2013b, p. 286). Shook states a similar opinion:

> "Depictions of entire societies or a whole planet undergoing empathetic moral enhancement will remain utopian fantasies. One country after another will decline moral enhancement until the "worse" countries have done it, and each country would want their neighbors to go first. Even if everyone cared a whole lot more about each other, people will still care about other priorities like personal survival,

[4] Masahiro Morioka also uses this term, and seems to have a similar concept in mind (Morioka 2014, p. 121).

> family responsibilities, career advancement, equitable distributions, class distinctions, social equality, retributive justice, restoration of homelands, political freedom, and so on." (Shook 2012, p. 11)

The root of these difficulties is that implementing universal MB is a challenge that requires broad cooperation, and the very reason we need to implement universal MB is that humans are incapable of broad cooperation. The upshot is that most people, and especially those people most in need of it, will refuse to voluntarily undergo MB, or allow it to be forced upon everyone.

Despite the above, there are reasons to think that moral bioenforcement might not be necessary in order to achieve the goals of CBS-based MB. We may be able to work our way in small steps towards universal MB. Once the beneficial effects become apparent, more and more individuals and societies may voluntarily undergo it. We could start with MB as an option for criminal offenders in exchange for lighter sentences, or perhaps even as a mandatory part of rehabilitation. We could require that leaders in positions of power undergo MB at the start of their term. If the worst and most influential individuals can be enhanced, norm-propagation as described in section 3.5 could result in gradual influence of entire populations.

This kind of incremental MB implementation is not ideal, as it may take a long time. In that time, the effects and risks of Harm become greater as scientific progress advances while moral deficiency remains unaddressed. Moreover it will result in a long period in which morally enhanced individuals interact with morally non-enhanced individuals, exacerbating problems of moral counter-intuitiveness and exploitation.

If MB must take the form of moral bioenforcement, the prospects for it actually happening are unlikely but not impossible. It may be practically achievable, and the risks may be outweighed by the benefits. Moral bioenforcement might be similar to how many countries presently treat education. Children are forced to receive education, whether or not they or their parents want them to, because the society in which they live widely regards education to be beneficial. In fact, as discussed in section 1, education systems incorporate TME techniques.

CBS-based MB could arguably be beneficial in the same way as education; beneficial both for the individual who receives it and for the rest of society. Moreover, if MB

primarily occurs through embryonic interventions, the issue of consent may be less problematic; parents may still need to consent, but the prospect of altering an unborn child's personality for the better may be more palatable than altering that of a fully-developed individual.

Regardless of the benefits, it is hard to avoid the problems that arise if moral bioenforcement is the only option for successful MB. This is one of the main points against MB; moral bioenforcement may be required, and this will almost certainly negatively impact upon democratic principles of freedom. However, if the alternative is rampant Harm, this sacrifice may be worth it.

4.5: The better option

None of the three prominent objections discussed in this section can be completely put to rest. If there are alternative options to MB for addressing Harm, and they do not carry the same risks and problems, it may be better to focus our efforts on them instead of MB.

Our alternatives to MB include further scientific progress. This encompasses technological breakthroughs that might directly address causes of harm, and the possibility of further cognitive enhancement that could perhaps intrinsically involve some kind of moral enhancement. We briefly discussed these possibilities in section 1.8, and I see no reason that they are not hopeful (unlike Persson and Savulescu). Even though scientific progress is partly responsible for the risk and prevalence of Harm, it could easily provide means to address it in the future.

But there is another alternative I want to focus on. Most opponents of moral bioenhancement, having made their arguments against it, are quick to add that we already have perfectly good means of morally enhancing people:

> "These tried and tested methods include the traditional ones of bringing children up to know the difference between right and wrong, to avoid inflicting pain or suffering on or doing harm to others; and instilling in them habits of respect for others." (Harris 2011, p. 104)

In section 1, we briefly examined Persson's and Savulescu's arguments that TME cannot sufficiently address human moral deficiency. However, many argue that P&S are too pessimistic about the prospects for TME (Zarpentine 2013, pp. 143-145; Harris 2011, p. 105; Morioka pp. 121-122).

Indeed, it seems as though TME works quite well. Consider all of the so-called moral progress humans have made over recent decades. Extreme poverty has declined rapidly, from 43% of the global population in 1981 to 18% in 2010 (World Bank 2014, p. 1). Victims of historically extensive discrimination and oppression, such as women, racial and religious minorities, and LGBT individuals, have received numerous rights in many societies, which they were previously denied. Stephen Pinker's *Better Angels of Our Nature* details the occurrence of violence in human societies; more specifically, that it has been steadily declining over the entire course of human history (Pinker 2011).

If we can measure improvement of moral capacities by any of the above factors, which seems plausible, it would follow that human moral enhancement has been constantly occurring as our societies have become more advanced. Many of them could also arguably be linked with a reduction in in-group bias, and an indication that we may one day achieve broad cooperation through TME. These considerations may count against the claim that TME is not effective enough to address human moral deficiency.

One possible response is that, although moral enhancement does occur through TME, the resultant improved moral capacities are still based on evolved, deficient common-sense morality. TME has altered the moral capacities of many individuals in ways that are beneficial, such as by making violent behaviour less morally acceptable. However, these benefits are just variations on the same moral capacities that were effective in benefiting small scale human populations in ancient times before scientific progress. To claim that TME will ever be able to provide the radically new kinds of moral capacities required to address harm is a different matter. TME's problems are similar to those of ISS-based MB: the objective of each is to make us 'morally better', which means 'more satisfactory to current moral intuitions'. As I argued in section 2.5, current moral intuitions do not promote broad cooperation,

and so will arguably not address Harm; we will still experience moral intuitions founded on inter-group competition.

Despite the progress made by TME, it is not hard to find evidence of the lingering inability of current moral intuitions to promote broad cooperation. First, we have all the examples of Harm presented in sections 1.5 and 1.6. There is still war, terrorism, oppression, unsustainable societal practices, and excessive inter-group animosity almost everywhere in the world. Even if all of these problems are decreasing, they may not be doing so fast enough. They also may be superficial. Persson and Savulescu point out that people tend to regress to "barbarous behaviour" with "frightening speed", "when political conditions allow it" (Persson & Savulescu 2012, p. 106). We cannot be sure that TME is resulting in genuine, lasting broadly cooperative tendencies.

However, even if TME is presently insufficient, it arguably has the potential to become sufficient in the future. Perhaps improved TME will follow advancements of other kinds, similar to how technological advancements have resulted in widespread improvements to standards of living. Reasoning can catalyse changes in moral intuitions, and reasoning is what led us to the conclusion that broad cooperation will address Harm. Perhaps cognitive enhancement and technological advancement through scientific progress will result in novel solutions to the problems caused by current moral intuitions (Fenton 2011). One possibility is the development of new standards of child-rearing, so that parents are no longer entitled to more or less indiscriminate influence over their children's moral intuitions.

But even advocates of MB should acknowledge that TME is worthwhile. Moral reflection and the propagation of moral reasons will play a role even in a future where MB is commonplace. There is no reason why MB and TME need be mutually exclusive; MB does not replace our current moral capacities, but alters them so that they are improved. We will still think about moral value, perform moral reasoning, and experience moral intuitions; moral discourse and education will continue even among morally enhanced individuals. Effective MB may be altogether impossible without help from TME:

> "Moral bioenhancement will of course rest on a conventional moral education: children would still need to be taught correct values, and the importance of acting

> on values, etc., just as cognitive enhancers do not work without education and study. But the moral bioenhancement may allow the education we routinely give our children to be more effective." (Savulescu et. al., 2014, p. 102)

This line of thought, especially combined with a CBS that includes embryo interventions, is one of the most plausible and hopeful options we have with regard to addressing Harm through MB.

4.6: Summary of section 4

We discussed three primary objections to MB: the superficiality concern, the moral complexity objection, and the objection to moral bioenforcement. Although these objections usually target ISS-based MB, they can apply to the CBS as well.

The superficiality concern encompasses arguments that, in-principle, MB could never result in moral enhancement that is as genuine as that which can be achieved through traditional moral enhancement (TME). I argued that the superficiality concern does not cause the same problems for the CBS as it does for the ISS. One problem it might cause for the CBS, depending on whether my account of moral intuitions is correct, is that of undermining its status as moral enhancement. Another is that, if there is genuine, objective moral value that can only be perceived through cold reasoning, CBS-based MB may interfere with this and cause us to morally endorse genuinely immoral actions. In either of these cases, however, the CBS could still motivate broad cooperation and address Harm. It just would not qualify as a moral enhancement.

The moral complexity objection encompasses arguments that our moral capacities are so complex that, in-practice, it would be prohibitively difficult to successfully improve them through biotechnological interventions. Attempts to do so are too likely to fail and produce negative consequences. I argued that, while the complexity of our moral capacities does pose a significant challenge to prospective MB, these concerns are related mainly to the desire for MB to satisfy our current moral intuitions; to the ISS. MB that successfully addresses Harm will have to modify our current intuitions, which is what the CBS aims to do. We should expect the results of CBS-based MB to conflict with our current moral intuitions. It is a separate question

whether doing so is morally permissible or practically advisable. The moral complexity objection, however, does point out that any MB attempts, including CBS-based ones, will be risky due to the highly complex nature of the moral capacities they must manipulate. This is an important problem, but one that can only be solved by technological advancement.

The objection to moral bioenforcement argues that MB would have to be universal to be effective, and the only way to make MB universal is to make it mandatory. There is some hope that MB could remain at least partly voluntary, though this would result in sub-optimal implementation. But if moral bioenforcement truly is necessary it will probably involve severe violation of principles of democratic freedom. This may ultimately be unavoidable, but could be a sacrifice worth making in order to address Harm.

Because we cannot completely assuage the above concerns, CBS-based MB is left looking risky at best. We need a good reason to justify that risk if we choose to take it by developing and implementing MB. The main reason we have is MB's potential to address Harm. But there are reasons to think that TME and further scientific progress also have the potential to address Harm, and do not carry the same risks that MB does.

Conclusion

In section 1 we considered how moral bioenhancement (MB), defined as the improvement of human moral capacities through biotechnological interventions, might be beneficial. I argued that MB has the potential to help prevent Ultimate Harm (UH) that threatens to end all worthwhile life on Earth, and reduce Widespread Harm (WH) that contributes to enormous amounts of suffering. This is because both UH and WH (Harm) are partly caused by two kinds of human moral deficiency. I also argued that, despite the possibility that MB is not the best way to address Harm, we have good reasons to seriously consider developing and implementing MB. This is, of course, unless we can find reasons that MB is more problematic than its alternatives.

In section 2 we considered how morality really works, exploring different meta-ethical theories and the different ways they explain moral value and moral motivation. I emphasised the importance to morality of two core concepts. The first is moral intuition; it links moral value and moral motivation, and explains how humans experience the evolved tendency to cooperate. The second is broad cooperation; a more inclusive form of cooperation than was beneficial to human evolutionary ancestors. Our moral intuitions result in the human moral deficiency that contributes to harm. This is because the promotion of broad cooperation will address Harm, but current moral intuitions do not motivate broad cooperation. Therefore, MB should promote broad cooperation, but must involve moral intuition.

In section 3 we considered the possible ends to which MB could be put. I argued that MB discussions invariably have only one interpretation of improved moral capacities in mind. MB's objective is assumed to be making people 'morally better'; the satisfaction of current moral intuitions. I call this the intuition-satisfaction strategy (ISS). I presented an alternative MB strategy that prioritises the promotion of broad cooperation. Its objective is to improve our moral capacities by modifying the moral intuitions we experience so that they motivate broad cooperation. I call this the cooperation broadening strategy (CBS). The CBS is easier to implement than the ISS, because the former does not require us to resolve all meta-ethical disagreement in order to be considered successful, while the latter does. The CBS is also more likely to address Harm than the ISS, because broad cooperation will directly counteract the

problems that contribute to Harm, while satisfying current moral intuitions will only make people intuitively morally better.

In section 4 we examined several of the most prominent criticisms that have been raised against the prospect of MB, and saw how the CBS can help mitigate some problems raised against the ISS. However, we were still left with several pressing concerns still partially unsolved, especially the likely requirement that MB will have to be universal and mandatory. We considered how the potential benefits of MB stack up against those of further scientific progress traditional moral enhancement (TME), and whether this leaves MB looking as though it is worth the risks and problems that may accompany it. Although I am unconvinced that TME will ever be as effective by itself as MB, I think that a combination of the two would be preferable to and more effective than either one alone.

I am ultimately still uncertain about the best course of action regarding MB. The alternative measures for addressing Harm available to us could potentially work well enough, without carrying the same risks as MB. Moreover, scientific progress and TME will be necessary in order for MB to be successfully developed and implemented. The key question still remains: could they address harm alone? We have seen some reasons to think they could not, but the matter is far from settled.

The arguments in this thesis leave many prominent criticisms of MB looking less problematic than they did. This is largely because objections to MB have so far focused on the problematic ISS, while the CBS is a better prospect. The problems that we could not address have at least been thrown into doubt. The objectives of MB are worth accomplishing, and there are reasons to think MB will be successful. We do not presently have any indisputably better way to reduce and prevent Harm. CBS-based MB has enough going for it, and little enough going against it, that its development and implementation should be seriously considered.

References:

Agar, N. (2015). Moral Bioenhancement and the Utilitarian Catastrophe. *Cambridge Quarterly of Healthcare Ethics, 24*, 37-47.

Alexander, L., & Moore, M. (2012). Deontological Ethics. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition). from http://plato.stanford.edu/archives/win2012/entries/ethics-deontological/

Bossaer, J., Gray, J., Miller, S., Enck, G., Gaddipati, V., & Enck, R. (2013). The Use and Misuse of Prescription Stimulants as "Cognitive Enhancers" by Students at One Academic Health Sciences Center. *Academic Medicine, 88*(7), 1-5.

Buchanan, A., Brock, D., Daniels, N., & Wilker, D. (2000). *From Chance to Choice*: Cambridge University Press.

Campbell, R., & Kumar, V. (2012). Moral Reasoning on the Ground. *Ethics, 122*, 273-312.

Carter, J. A., & Gordon, E. C. (2015). On Cognitive and Moral Enhancement: A Reply to Savulescu and Persson. *Bioethics, 29*(3), 153-161.

Chan, S., & Harris, J. (2007). In Support of Human Enhancement. *Studies in Ethics, Law, and Technology, 1*(1), 1-3.

de Melo-Martin, I., & Salles, A. (Forthcoming). Moral Bioenhancement: Much Ado About Nothing? *Bioethics*.

Dees, R. H. (2011). Moral Philosophy and Moral Enhancements. *AJOB Neuroscience, 2*(4), 12-13.

DeGrazia, D. (2014). Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of Medical Ethics, 40*, 361-368.

Dickins, R. (2013). Preparing the Gaia connection: An ecological exposition of psychedelic literature 1954-1963. *European Journal of Ecopsychology, 4*, 9-18.

Douglas, T. (2008). Moral Enhancement. *Journal of Applied Philosophy, 25*(3), 228-245.

Douglas, T. (2013). Moral Enhancement via Direct Emotion Modulation: A Reply to John Harris. *Bioethics, 27*(3), 160-168.

Douglas, T. (2014). Enhancing Moral Conformity and Enhancing Moral Worth. *Neuroethics, 7*, 75-91.

Dubljević, V. (2013). Prohibition or Coffee Shops: Regulation of Amphetamine and Methylphenidate for Enhancement Use by Healthy Adults. *The American Journal of Bioethics, 13*(7), 23-33.

Faust, H. S. (2008). Should We Select for Genetic Moral Enhancement? A Thought Experiment Using the MoralKinder (MK+) Haplotype. *Theoretical Medicine and Bioethics, 29*, 397-416.

Fenton, E. (2010). The perils of failing to enhance: a response to Persson and Savulescu. *Journal of Medical Ethics, 36*, 148-151.

GRACE Communications Foundation. (2015). Antibiotics. from http://www.sustainabletable.org/257/antibiotics

Greene, J. (2013). *Moral Tribes Emotion, Reason, and the Gap Between Us and Them* (E-book ed.): Atlantic Books.

Gunson, D., & McLachlan, H. (2013). Risk, Russian roulette and lotteries: Persson and Savulescu on moral enhancement. *Medicine, Health Care and Philosophy, 16*, 877-884.

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review, 108*(4), 814-834.

Haidt, J. (2013). *The Righteous Mind Why Good People Are Divided by Politics and Religion* (E-book ed.): Penguin Group.

Harris, J. (2010). *Enhancing Evolution* (E-book ed.): Princeton University Press.

Harris, J. (2011). Moral Enhancement and Freedom. *Bioethics, 25*(2), 102-111.

Harris, J. (2013a). 'Ethics is for Bad Guys!' Putting the 'Moral' Into Moral Enhancement. *Bioethics, 27*(3), 169-173.

Harris, J. (2013b). Moral Progress and Moral Enhancement. *Bioethics, 27*(5), 285-290.

Harré, N. (2011). *Psychology for a Better World.* from http://www.psych.auckland.ac.nz/psychologyforabetterworld

Holtug, N. (2011). Equality and the treatment-enhancement distinction. *Bioethics, 25*(3), 137-144.

Hume, D. (1751). An Enquiry Concerning the Principles of Morals. *Works of David Hume* (E-book ed.): MobileReference.

Hursthouse, R. (2013). Virtue Ethics. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition). from http://plato.stanford.edu/archives/fall2013/entries/ethics-virtue/

IAGS. (2003). *How much did the September 11 terrorist attack cost America?* , 2014, from http://www.iags.org/costof911.html

Jackson, F. (2000). The Location Problem for Ethics: Moral Properties and Moral Content. *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (pp. 113-138): Oxford Scholarship Online

James, C. (1997). *Global Status of Transgenic Crops in 1997.* Ithaca, NY: ISAAA.

Jotterand, F. (2014). Questioning the Moral Enhancement Project. *The American Journal of Bioethics, 14*(4), 1-3.

Kahane, G., & Savulescu, J. (2015). Normal human variation: Refocussing the enhancement debate. *Bioethics, 29*(2), 133-143.

Knapton, S. (2015). First new antibiotic in 30 years discovered in major breakthrough. from http://www.telegraph.co.uk/news/science/science-news/11331174/First-new-antibiotic-in-30-years-discovered-in-major-breakthrough.html

Morioka, M. (2014). Some Remarks on Moral Bioenhancement. In A. Akabayashi (Ed.), *The Future of Bioethics: International Dialogues:* Oxford University Press.

Nisbett, R. E., & Cohen, D. (1996). *Culture of Honor: The Psychology of Violence in the South* (First ed.): Westview Press.

Persson, I., & Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy, 25*(3), 162-177.

Persson, I., & Savulescu, J. (2011). The turn for ultimate harm: a reply to Fenton. *Journal of Medical Ethics, 37*, 441-444.

Persson, I., & Savulescu, J. (2012). *Unfit for the Future* (First ed.). Great Clarendon Street, Oxford OX2 6DP United Kingdom: Oxford University Press.

Persson, I., & Savulescu, J. (2013). Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement. *Bioethics, 27*(3), 124-131.

Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined* (E-book ed.): Viking Penguin.

Raus, K., Focquaert, F., Schermer, M., Specker, J., & Sterckx, S. (2014). On Defining Moral Enhancement: A Clarificatory Taxonomy. *Neuroethics, 7*(3), 263-273.

RERF. (2007). *Frequently Asked Questions: How many people died as a result of the atomic bombings?* , 2015, from http://www.rerf.or.jp/general/qa_e/qa1.html

Sauer, H. (2012a). Educated intuitions, Automaticity and rationality in moral judgement. *Philosophical Explorations, 15*(3), 255-275.

Sauer, H. (2012b). Psychopaths and Filthy Desks: Are Emotions Necessary and Sufficient for Moral Judgment? *Ethical Theory and Moral Practice, 15*, 95-115.

Savulescu, J., Douglas, T., & Persson, I. (2014). Autonomy and the Ethics of Biological Behaviour Modification. In A. Akabayashi (Ed.), *The Future of Bioethics: International Dialogues*: Oxford University Press.

Sayre-McCord, G. (2015). Moral Realism. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition). from (forthcoming URL) http://plato.stanford.edu/archives/spr2015/entries/moral-realism/

Shook, J. R. (2012). Neuroethics and the Possible Types of Moral Enhancement. *AJOB Neuroscience, 3*(4), 3-14.

Singer, P. (1993). *Practical Ethics* (Second ed.). The Pitt Building, Trumpington Street, Cambridge, United Kingdom: Cambridge University Press.

Sinnott-Armstrong, W. (2014). Consequentialism. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). from http://plato.stanford.edu/archives/spr2014/entries/consequentialism/

Sparrow, R. (2014). Better Living Through Chemistry? A Reply to Savulescu and Persson on 'Moral Enhancement'. *Journal of Applied Philosophy, 31*(1), 25-32.

Tennison, M. N. (2012). Moral Transhumanism: The Next Step. *Journal of Medicine and Philosophy, 37*, 405-416.

United Nations. (1995). *Report of the World Summit for Social Development*. 2015, from http://www.un.org/documents/ga/conf166/aconf166-9.htm

van Roojen, M. (2014). Moral Cognitivism vs. Non-Cognitivism. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition). from http://plato.stanford.edu/archives/sum2014/entries/moral-cognitivism/

Wallace, E. (2009). Drugs in Sport. 2015, from http://www.ulster.ac.uk/scienceinsociety/drugsinsport.html

Wilson, E. O. (2002). Speciation and Biodiversity. from http://www.actionbioscience.org/biodiversity/wilson.html

World Bank. (2014). *Prosperity for All / Ending Extreme Poverty: A Note for the World Bank Group Spring Meetings* 2014. Washington DC, USA: World Bank: The World Bank.

Zarpentine, C. (2013). 'The Thorny and Arduous Path of Moral Progress': Moral Psychology and Moral Enhancement. *Neuroethics, 6*, 141-153.