

## **Moral Enhancement Can Kill**

There is recent empirical evidence that personal identity is constituted by one's moral traits. If true, this poses a problem for those who advocate for moral enhancement, or the manipulation of a person's moral traits through pharmaceutical or other biological means. Specifically, if moral enhancement manipulates a person's moral traits, and those moral traits constitute personal identity, then it is possible that moral enhancement could alter a person's identity. I go a step further and argue that under the right conditions, moral enhancement can kill the subject of the enhancement. I then argue that these conditions are not remote.

Keywords: Moral enhancement, personal identity, moral psychology

### **1. Introduction**

Although the idea that a person's moral traits contribute to that person's identity is not new, recently some philosophers and psychologists have argued on empirical grounds that a person's moral traits are essential to that person's identity (Strohming & Nichols 2014). Supported by a series of surveys about identity and what mental changes contribute to identity changes, Strohming and Nichols claim that their surveys support the "essential moral self hypothesis," which is the hypothesis that moral traits are the most essential mental features to personal identity. Details of the essential moral self hypothesis aside for now, if a person's moral traits are essential to that person's identity, then as those essential traits change, so too could the person's identity change.

Over the past few years, the idea that humans can or should improve our moral traits pharmaceutically, or by way of other biomedical means, has received a great deal of attention. This idea of moral enhancement can be summarized as the manipulation of a person's moral traits by biological means. Moral traits are construed broadly to include moral beliefs or other attitudes, moral motivation, dispositions, or emotions. The discussion has concerned primarily

whether we should engage in moral enhancement, and if so how, and secondarily whether we are physically able to engage in moral enhancement.

To my knowledge, no one has yet considered something like the essential moral self hypothesis in determining whether we should engage in moral enhancement. But given the empirical evidence, such a consideration is warranted, because if it's true that a person's identity is determined by his or her moral traits, and moral enhancement manipulates those traits, then moral enhancement may end up manipulating the identity of that person. I argue below that such manipulations can end the life of the manipulated, if the essential moral self hypothesis is true. In other words, moral enhancement can kill.<sup>1</sup>

Moral enhancement only kills under certain readings of the essential moral self hypothesis, however, while other conditions must also obtain. After setting out the different readings of the essential moral self hypothesis, I argue that moral enhancement can kill, if the essential moral self hypothesis is true, and the conditions of moral enhancement to kill are not remote. I argue further that sometimes this killing can be murder.

## **2. Moral enhancement**

Moral enhancement is the act of improving the moral traits of people through biological means. One of the purported reasons for moral enhancement is that enhancing humans' moral traits through biological means will be necessary to prevent very bad things from happening, especially since as humans' cognitive abilities increase more people have access to more serious threats (Persson & Savulescu 2008, 2013a, 2014). Others think that moral enhancement is inevitable, given the increasing use of cognitive enhancement (Harris 2011).

---

<sup>1</sup> Douglas (2008) briefly considers the potential threat of moral enhancement to personal identity. But the context of that discussion is that moral traits are not essential to a person's identity.

Among the areas of discussion over moral enhancement has been the discussion of whether moral enhancement should be compulsory or voluntary (Persson & Savulescu 2013b; Rakić 2014). It may be necessary to make moral enhancement necessary for some groups of people, such as psychopaths (Carter 2016; Wiseman 2016) while making it available to others for voluntary use. Others have argued that moral enhancement can be effective only if it is administered covertly (Crutchfield 2016).

Currently, the science and technology are not sophisticated enough to engage in widespread moral enhancement. But the required level of sophistication may someday be reached. Potential vectors of enhancement include lacing municipal waters supplies with a drug, offering drugs on a regulated pharmaceutical market, or by modern neurotherapies (Nakazawa et al. 2016).

### **3. The essential moral self hypothesis**

Strohminger and Nichols argue on empirical grounds for the essential moral self hypothesis, which is the hypothesis that “moral traits are more essential [to personal identity] than any other mental feature, including those that provide functionality, distinctiveness, or personal narrative” (2014 p. 160). Mental features that provide for functionality are things like perception and cognition. Features that provide for distinctiveness are memories, desires, or preferences. Memory is most typically associated with being a determinant of identity, even as source of a unified personal narrative. Instead, the essential moral self hypothesis is that a person’s moral traits such as psychopathy, piety, politeness, conscientiousness, honesty, loyalty, compassion, racism, empathy, and generosity are the most important determinants to personal identity.

To support the essential moral self hypothesis, Strohminger and Nichols performed a series of five studies. All five studies measured the degree to which respondents thought that moral traits, as compared to other mental features, contribute to personal identity. In all five studies, the moral traits were the traits that respondents thought contributed most to identity. In the first study subjects responded to a case in which a person loses a mental faculty after a brain transplant. Of all the lost faculties, the loss of moral faculties was most strongly associated with changes in the identity of the transplant recipient, an even stronger association than the one exhibited between identity and memories

The second study measured the degree to which a person's identity would change if he or she were to take a pill that alters mental features. Again, respondents thought that changes in moral traits resulted in greater changes in identity than did any other mental feature, including memories.

The third study asked respondents to judge which mental features would persist after a soul switched to a new body. The traits most strongly associated with persistence were the moral traits, more so than the traits related to memory. Similarly, in the fourth study, respondents judged whether traits would persist after reincarnation, and the traits judged most likely to persist after reincarnation were the moral traits.

The fifth study measured the associations between identity and a wide range of traits that are likely to be considered distinctive of the person. Then respondents imagined that they were meeting an old friend who had over time undergone changes in a list of traits. Respondents were to rate how much the change would impact the friend's identity. Like the other four studies, respondents thought that moral traits contributed most to the person's identity.

Strohminger and Nichols' studies provide good evidence that the folk—laypeople—believe that the essential moral self hypothesis is true. The folk believing that moral traits constitute identity is compatible with four more specific formulations of the hypothesis:

- (a) The folk think that identity wholly consists of moral traits and are right;
- (b) The folk think that identity wholly consists of moral traits and are wrong;
- (c) The folk think that identity partially, but essentially, consists of moral traits and are right;
- (d) The folk think that identity partially, but essentially, consists of moral traits and are wrong.

The empirical evidence from Strohminger and Nichols' studies suggests that the folk believe that, though moral traits are essential, other psychological traits contribute to personal identity. Even though moral traits exhibited the strongest associations, other traits also exhibited associations. Thus, formulations (a) and (b), though consistent with the essential moral self hypothesis, are not compatible with the results from the studies, which leaves only (c) and (d). Though the evidence supports only (c) and (d), it also supports the idea that changes in moral traits can lead to a change in identity.

I address the idea that (d) might be true—that the folk are wrong about identity—near the end. But in what follows I assume that the folk are not making categorically false judgments about personal identity, which is to say that I'll assume that a person's moral traits partially constitute that person's identity. I also assume that the respondents in the study are sufficiently representative of some larger population that makes it possible to generalize the conclusion to this larger population, whatever that larger population is. Of course, it's always possible that when generalizing from a sample to a larger population the sample is not in fact representative.

But it's impractical to ask everyone what they think about personal identity and its relation to moral traits.

Suppose that a person, S, is a subject of moral enhancement. Suppose S at time 1 has a hundred moral traits (which exhaust her moral traits).<sup>2</sup> After the pharmaceutical enhancement at time 2, the person who was S at time 1 has a hundred different moral traits with no intersection of those of S at time 1. Call this person S'. The question is: are S and S' the same person?

All of the going accounts of personal identity allow that a person's traits can change dramatically while preserving numerical identity, so long as there is some form of continuity from one change to the next. There's no reason to think that the essential moral self hypothesis would postulate anything different--that as long as there is continuity between changes in moral traits, then there is no change in personal identity. But the hypothesis also isn't yet well developed, and as such makes no claims about conditions for persistence of personal identity. To insist that S and S' are not the same person after a gradual, continuous change in moral traits would be to saddle the essential moral self hypothesis with a position it doesn't imply. But the empirical evidence and the formulation of the hypothesis are also compatible with the notion that continuity isn't sufficient for identity persistence, as presumably the respondents who thought that there were changes in identity also thought that those changes were continuous, or that the changes occurred without a break in continuity. In any case, I assume here that the change in moral traits between S and S' is total and immediate rather than gradual and continuous, that the enhancement breaks continuity.

---

<sup>2</sup> Or however traits are individuated. This is merely an example, and not an assertion that people have exactly one hundred moral traits.

People undergo changes to their moral traits all the time. But usually these trait changes don't result in different identities because only very few traits change or because the changes occur within the person's narrative in a way that allows the narrative to continue to unify the self, preserving the person's identity through the change. Usually these trait changes occur over long periods of time. But they can occur over short periods of time, as in the case Saul of Tarsus—a persecutor of early Christians—who, three days after seeing the resurrected Jesus, became Paul the Apostle. Traumatic events may also be able to break the personal narrative, change many of one's moral traits, and, if the essential moral self hypothesis is correct, result in a different person. If the break of a personal narrative—a break in continuity—can occur ordinarily, then it isn't unreasonable to suppose that moral enhancement could do this pharmaceutically.

The essential moral self hypothesis doesn't imply that moral traits are the only determinants of identity. Other psychological states or traits also help to determine identity, and Strohminger and Nichols' studies show that the folk think this is also true. It's not that these other states are irrelevant to a person's identity—a person is who he or she is because of all of these traits. It's that of all of these traits, the moral traits are the most essential to determining one's identity. If S and S' are numerically identical, then moral traits aren't essential at all, and the essential moral self hypothesis is false. Since I am supposing, in accordance with the results from Strohminger and Nichols' studies, that moral traits are essential, S and S' must not be the same person. This doesn't mean that they can't be very similar to each other. But similarity isn't identity.

#### **4. Moral enhancement can kill**

##### *4.1 A change in moral traits can cause a person to die*

If the essential moral self hypothesis is correct, then the case that the moral enhancement in the simplified example above causes S to die seems straightforward. The intervention causes S to lose all of her moral traits, and this loss occurs immediately. The result is S', who has a totally different set of moral traits. Since moral traits, we are supposing, are the most essential mental features to a person's identity, then the identity of S is no more.

However, if loss of identity results in the death of the person, then an implication may be that there is rather more death around than we ordinarily think. For example, suppose that a selfish person faces certain injury or even death, but then is rescued by a stranger. It could be that the selfish person's outlook on life changes as a result, so that he or she is charitable, selfless, kind, and so on, so much so that people say he or she "is not the same person." If it is true that the rescue caused such significant changes in the person's moral traits, and the essential moral self hypothesis is true, then the selfish person has died, which some may find counterintuitive. What may be even more counterintuitive is that the stranger caused the person's death.

Or suppose that a single event, such as an observed act of injustice, causes a significant deterioration of a person's moral traits, such as the loss of his or her faith in humanity. It could be that this person also dies, if the essential moral self hypothesis is true.

The implication of cases like these, given the essential moral self hypothesis, is that death may be more common than we ordinarily think, and this is supposed to give us a reason to doubt the claim that if the essential moral self hypothesis is true, changes in moral traits can cause a person to die.

Supposing that it's true that changes in moral traits can result in changes in personal identity, the question is: does loss of identity result in death? It seems clear to me that it does, but others may disagree. Intuitions about cases aside, consider the alternative, that loss of identity



doesn't result in death. Suppose again that S and S' are not the same person after the moral enhancement. For it to be true that the moral enhancement didn't result in the death of S, then it must be true that S is still alive (assuming that if something is alive and hasn't died, it's still alive). But it's also true that S' is alive. Thus, S and S' are both alive, and both in the same body. Instead of changes in moral traits resulting in more death, they result in more life.<sup>3</sup>

This judgment that S, the selfish person, and the witness to injustice have not died after the change in moral traits hinges on the continued existence of their identities. But if the essential moral self hypothesis is true, those people no longer exist after the change—the moral enhancement replaces the moral traits with new ones. If moral enhancement erases the existence of those identities, then it must be that they have died, because a prerequisite to being alive is existing. Thus, in cases where the change to moral traits is so significant to result in a change in identity, the former person is no longer alive, because he or she no longer exists.

An alternative is to hold that life and death are properties of bodies, not persons. If this is true, then significant changes in moral traits would not result in death, whether they result in changes in personal identity. Thus, changes in moral traits cannot result in death.

This alternative implies that S was never alive in the first place (because S is a person, and the property of being alive is the property of bodies, not persons). Notwithstanding this counterintuitive implication, it's plausible that what makes death so bad is that it is the greatest loss a person can suffer (Sinnott-Armstrong & Miller 2012). In the case of loss of identity, one

---

<sup>3</sup> The use of “alive” is admittedly awkward when referring to persons, though I use it neutrally between the notion that life and death are properties of identities and the notion that life and death are properties of bodies, which I discuss two paragraphs below.

loses everything, because one no longer exists. Thus, though on this alternative view of life and death S may not die, he or she is still subject to the resulting badness.

The judgment that S and S' are not the same person, if the essential moral self hypothesis is correct, is clear because it is built into the example that the change is total and immediate. This means that the moral traits themselves are gone, along with their function in the person's psychology. For most of us, our moral traits are connected with other psychological states and function in our personal narratives. When S's moral traits are replaced, the original moral traits no longer exhibit those connections and they no longer function in the new narrative. This is not to say, however, that S' lacks those other psychological states. Indeed, she is likely to be psychologically very similar to her predecessor, sharing memories and other psychological traits (but not personal narrative, to the extent that personal narrative is composed of or derived from moral traits). S' may think of S as her old self, or even think that she's the same old person, failing to recognize that she's not. But if the essential moral self hypothesis is correct, then S and S' are not the same person, no matter what S' thinks.

#### *4.2 When does S die?*

In the example, S and S' are not the same person, and S dies. If the moral traits are swapped without continuity, then the moment S dies is the moment that her moral traits are swapped. However, as I mention above, the essential moral self hypothesis may allow that identity can change even when the change in moral traits is gradual. In the case of gradual change, it's less obvious when S dies. Suppose S has a hundred moral traits. When the manipulation swaps out the first moral trait, does S die? Intuitively not. So, one moral trait doesn't make the difference. So, swapping another trait won't make a difference. Apply this

reasoning to the whole sequence of the moral trait changes, and at no point does S's identity change. But we know that S and S' are not the same person.

The difficulty is familiar. On the one hand, it seems that S and S' are different people. But on the other hand, when the reasoning that one trait doesn't make a difference to the identity of S is carried out, one trait at a time, the conclusion is that S is the same person throughout the intervention. Answering the question, "When does S die?" requires resolving this paradox. The likely solution will be either that the true boundaries of 'being S' are determinate but unknowable or that whether a person is S is context-sensitive, as these are two of the most common solutions to a Sorites Paradox (Stanley 2003; Williamson 2002).

It is plausible that if the essential moral self hypothesis is correct, then the point at which S becomes S' during a gradual and continuous change is sensitive to the attributor's context. Suppose that A and S are colleagues working for a charity devoted to serving people without a home. A knows S as someone who is caring and compassionate, as those are the moral traits of S's with which A is most acquainted. If S' came to the charity one day, and started exhibiting traits indicating a lack of compassion and caring, it's plausible that A may say that S isn't the same person as before. And that attribution may be true while S's family, who, though they have noticed the change in compassion and caring have also noticed stability in other traits, might say that S is still the same person. That is, A and the family might both make different identity attributions, and it's plausible that both might be right. But it's also plausible that what the boundaries of identity are, if the essential moral self hypothesis is true, are definite but unknowable—that there is a threshold past which identity changes. There are many reasons why the boundaries (i.e., threshold) might be unknowable, including that we have limited access to

others' psychology. This is compatible with there being a definite point in a gradual change of moral traits at which one's identity changes.

Whether we can know the boundaries of personal identity or whether attributions of identity can shift depending on context are interesting issues worth exploring. But a detailed evaluation of these possibilities is beyond the present purpose, which is to show that moral enhancement can kill. Thus, for the sake of argument I assume that the moral enhancement totally replaces one's moral traits, and that this total replacement is immediate. If the moral enhancement works this way, then whatever connections the moral traits have with other states is broken, and it's at the moment of replacement that the change in identity occurs—it's at that point that the person dies.

#### *4.3 Voluntary enhancement may save the person*

Moral traits not only connect with other psychological states, but also with other moral traits. We all have higher-order moral traits, or moral traits that have as their object other moral traits. An example of a higher-order moral trait is the motivation to be a better person. This moral trait would be directed at, for example, increasing the frequency of exhibiting first-order moral traits, like helping those who need help or being a better friend.

Because we all have higher-order moral traits, voluntary moral enhancement may not result in death of the person, even if most of one's moral traits are replaced. In the case of a voluntary moral enhancement, presumably the volunteers are doing so because they want to be better people. This is a higher-order moral trait that may persist throughout the change in moral traits.

In the case of a voluntary enhancement, intuitively the person persists even after the intervention. Among a person's moral traits are the person's higher-order moral attitudes, such as

the desire to be a better person. A person who volunteers for moral enhancement is a person who is acting on his or her higher-order moral traits. Ideally these would be the desire to be a better person and the belief that volunteering for moral enhancement is a way to satisfy that desire, but it doesn't matter to the person's identity that he or she volunteers to improve him or herself. What matters is that the higher-order moral trait, whatever that is, persists through the intervention.<sup>4</sup>

The important point is that if some moral traits, such as higher-order moral traits, persist through the intervention, then so can the subject's identity. Thus, ordinary cases of moral enhancement, cases in which a person improves his or her first-order moral traits by forming certain higher-order moral traits, may not result in the death of the person.<sup>5</sup> For example, the racist who reforms his behavior by having higher-order moral traits to change his first-order moral traits may keep his identity over the reformation because the second-order moral traits will persist over the manipulation.<sup>6</sup>

---

<sup>4</sup> If, however, the higher-order moral trait does not persist through the manipulation, voluntary moral enhancement is plausibly suicide, if the essential moral self hypothesis is true.

<sup>5</sup> I don't mean to give higher-order moral traits a privileged status as determinants to identity. They are merely useful examples of the moral traits that would persist in cases of voluntary moral enhancement.

<sup>6</sup> Douglas (2008) considers a strong sense in which moral enhancement threatens personal identity and a weak sense. The strong sense is that we don't consider changes in moral traits to result in a different person. The weak sense is that changes in moral traits can alter identity. The essential moral self hypothesis supports the strong sense, and the empirical evidence in support of the hypothesis rebuts Douglas's dismissal of the strong sense. Moreover, his discussion assumes that the manipulation will be voluntary. In such a case, even if the manipulation will alter a person's identity, he claims that the person can still have reasons (higher-order moral traits) to volunteer for the manipulation.

So, it matters that the intervention is involuntary. It would be especially problematic for either moral enhancement or the essential moral self hypothesis if ordinary, garden-variety moral improvement resulted in death.<sup>7</sup>

#### *4.4 Is this killing?*

Suppose that S involuntarily undergoes a moral enhancement, and that the intervention totally and immediately replaces her moral traits. If the essential moral self hypothesis is correct, then she dies as a result. Her death is the result of something that someone did, and because of this it is a killing, whether the death occurs out of negligence (for failing to understand the potential consequences of forcing upon her the intervention) or out of the administrators' intent to cause her to die. Thus, in the example, participating in the moral enhancement kills S.

#### *4.5 Can moral enhancement do this?*

In principle, if the essential moral self hypothesis is true, moral enhancement can kill. But if the moral enhancement technology never advances to the point where it becomes possible to involuntarily administer an intervention that totally and immediately changes a person's moral traits (or otherwise change one's moral traits in a way that results in death), then to worry about the possibility that moral enhancement can kill is merely a mental exercise.

Admittedly, the current state of moral enhancement technology is nascent. This makes it difficult to say that it is currently possible to involuntarily administer a moral enhancement that totally and immediately changes one's moral traits. But it also makes it difficult to say that the

---

<sup>7</sup> It's possible that some moral traits are more essential to one's identity than are other moral traits, and that the persistence of the former would be sufficient for the persistence of the person, but that the persistence of the latter would not be so sufficient.

possibility is remote. Despite the nascence of the technology, for some populations the possibility may not be remote.

For example, it is plausible that some psychopaths' moral development is entirely dependent on their psychopathy. It is also plausible that if treatments for psychopathy are developed, pharmaceutical or otherwise, the treatment would manipulate their moral traits enough that, if the essential moral self hypothesis is true, they would be killed. That is, it is plausible that the treatment could alter all of one's moral traits and do so totally and immediately—they could go from something to nothing immediately. Suppose, for example, it's true that psychopaths' moral defects arise from their failure to make a distinction between conventional norms and moral norms (e.g., Nichols 2004). If a treatment were administered that fixed that failure, then they could go from lacking the ability to having the ability, at which point the change in moral traits could be total and immediate.<sup>8</sup> Future surgeries may also be able to totally and immediately treat psychopathy. If in those cases the administration of the treatment is involuntary, and then those people would have been killed. This scenario does not seem remote. There is a long history of treating research subjects unethically. Even with IRBs and independent ethics committees, fraud and abuse still occur in human subjects research. It's reasonable to think institutionalized psychopaths or others with mental health disorders could be subject to this fraud and abuse.

Or consider someone who is selfish, abusive to one's friends and family, manipulative, and mistrustful, and has been for most of one's life. A recent obituary describes such a man:

---

<sup>8</sup> The of moral traits would, in this case, be essential to the person's identity, just as the presence of moral traits in others is essential to others' identity. So, when the total absence of moral traits is replaced with some set of moral traits by way of treatment, the identity changes.

Leslie was surprisingly intelligent. However he lacked ambition and motivation to do anything more than being reckless, wasteful, squandering the family savings and fantasizing about get rich schemes. Leslie's hobbies included being abusive to his family, expediting trips to heaven for the beloved family pets and fishing, which he was less skilled at than the previously mentioned.

Leslie's life served no other obvious purpose, he did not contribute to society or serve his community... With Leslie's passing he will be missed only for what he never did: being a loving husband, father and good friend. (Larimer, Sarah. "Brutal obit slams Texas man whose life was 'much longer than he deserved.'" *The Washington Post*, February 13, 2017)

There is recent evidence that the intranasal administration of oxytocin increases trust, generosity, and altruism, especially within one's social group (Kosfeld et al. 2005; Zak et al. 2007; Hurleman et al. 2010; De Dreu et al. 2012). Though until the research is further developed, skepticism is warranted about the use of oxytocin to influence moral behavior, it does not seem remote for this technology to be developed to the point where it could be administered to a person such as Leslie, resulting in him being kind, generous, trusting, empathetic, and caring, especially toward one's family and friends and pets. And if this were to occur, given the results from Strohming and Nichols' studies, it seems reasonable to think that Leslie's family and friends would think him a different person, even if the behavioral manifestations of the trait change unfold gradually.

Further, we already know that in some cases sudden changes in personality traits can occur, especially with the rapid onset of some mental health disorders. If this process can be replicated or reversed, then it may be possible for a person's moral traits to change significantly and rapidly in a way that implies that he or she also changes identity.



So, although the development of the technology that would facilitate the killing a person is still in the early stages, it does not seem remote that it could change the moral traits of some people—people who are morally immature or who suffer from a disorder—in a way that results in their deaths. Further, it is an open question whether the technology will advance to the point where even morally mature people could be killed as a result of moral enhancement. Thus, it does not seem remote that moral enhancement can kill.

## **5. Murder by moral enhancement**

I argue in the previous section that moral enhancement can kill. Whether the involuntary killing of a person by way of moral enhancement is murder depends on whether the killing is justified and intentional.

### *5.1 Intentional killing by moral enhancement*

Whether killing by moral enhancement is murder depends in part on whether the intent of the enhancement is to kill. And whether the killing is intentional hinges on whether those administering the moral enhancement program believe that the manipulations will result in the person's death. It is plausible that they would have that belief. The empirical support for the essential moral self hypothesis suggests that the folk believe that moral traits partially constitute a person's identity. Thus, it's plausible that the folk believe that the interventions involved in moral enhancement would result in the death of the subject of the enhancement. But the people who would be administering the moral enhancement program would presumably be experts in moral psychology. They would therefore know of Strohminger and Nichols' research, as well as any other future research that suggests that a person's moral traits are essential to his or her identity. They would likely believe that moral enhancement would result in the death of the person, if the research continues to support the essential moral self hypothesis. Having the belief

that the enhancement will kill the person and continuing to carry out the enhancement is strong evidence that the administrators intend to kill the person.

### *5.2 Unjustified killing*

Whether the intentional killing of a person by moral enhancement further depends on whether the killing is justified. And whether it is justified depends on the ethical standards under consideration. According to the traditional conception of biomedical ethics, the intentional, involuntary killing of a person by way of moral enhancement would always be murder, because according to this bioethical approach the individual is the primary decision maker and care providers have ethical duties to do what's best for the patient. Health care providers would therefore be duty-bound to not administer the intervention. There's almost no harm greater to a person than to kill her and to do so by violating her autonomy. Any deontological conception of bioethics will imply that the intervention is always murder, because it would always be unjustified.

But a more utilitarian approach may justify the intentional, involuntary killing of a person by way of moral enhancement. If the consequences are such that the harms associated with killing the person are outweighed by the benefits of doing so, then it is justified. For example, it may be justifiable to manipulate the moral traits of a dangerous psychopath and kill him in the process in order to prevent others from suffering.

At the very least, the set of circumstances in which killing by moral enhancement is justified will be just as large as the set of circumstances in which homicide is ordinarily justifiable. This includes circumstances in which the killing is in self-defense, or in cases of capital punishment.

One of the most compelling cases for the necessity of moral enhancement is that because humans face such significant threats, threats that can potentially either kill us all or cause our well being to be so low that life is not worth living, we need to urgently enhance our moral capacities to effectively respond to these threats (Persson & Savulescu 2014). For this reason, moral enhancement ought to be compulsory. But if manipulating everyone's moral traits kills everyone in the process, then it may seem that the essential moral self hypothesis undercuts one of the strongest reasons to engage in moral enhancement in the first place. There'd be no reason to prevent the ultimate harm.

Compulsory moral enhancement may still be justifiable even if the essential moral self hypothesis is true. The reason is that the interventions involved very likely would not kill everyone that undergoes them. Probably most people would retain enough of their moral traits so that their identities would persist through the intervention, whatever that threshold happens to be.

But compulsory moral enhancement may kill some people. And if the essential moral self hypothesis is true, and if one either abides by principles of biomedical ethics or, if one is a utilitarian and the benefits do not outweigh the cost, then that killing is murder.

## **6. The folk might be wrong**

Given the truth of the essential moral self hypothesis, moral enhancement can kill, and sometimes that killing is murder. Some might think that I am overstating the strength of the essential moral self hypothesis, and this overstatement is what leads to the conclusion that moral enhancement can kill. But to get this conclusion, it only needs to be the case that when all of a person's moral traits are changed in a way that he or she no longer has those traits, he or she is

no longer the same person.<sup>9</sup> The essential moral self hypothesis can't be true while this condition is false. If it's possible for a person to lose all of one's moral traits, and still be the same person, then those moral traits weren't essential at all, let alone the most essential mental features.

But some may also want to allow that the empirical research suggests that the folk do think that moral traits are the most essential mental feature, but that the folk are wrong. Of the four formulations above, one may want to deny (c) but hold (d), which is that the folk think that identity partially, but essentially, consists of moral traits and are wrong.

It is highly unlikely anyone will ever develop a way to observe whether a person's moral traits in fact are essential to personal identity, because it's highly unlikely a way to observe personal identity will ever be developed. To empirically elucidate the concept of the personal identity, asking the folk what they think is a good start.

Holding (d) is to hold an error theory of personal identity. It's claiming that the folk's widespread belief about personal identity is wrong. The folk think lots of things and they can be wrong about those things. For example, it's plausible that people are wrong about their belief that moral claims can be true (Mackie 1990), or that the richness of conscious experience is due entirely to our perceptual relation to the world (Dennett 1993). But in these cases the error theories are independently motivated--there are other good reasons that cut against the folk. It's not clear there are any independently motivated reasons to hold (d) rather than (c). It would be hasty to claim that folk can't be right.

Scholars may have intuitions about personal identity and these intuitions may motivate the rejection of folk concepts, for example, intuitions about cases involving memory and

---

<sup>9</sup> Again, continuity may not be sufficient for persistence of identity.

personal identity. And in some domains folk intuitions ought not be trusted, such as in psychology or quantum mechanics, areas where empirical research contradicts folk intuitions and beliefs. There is of course a large body of literature devoted to observing personality traits and how they change and develop over time. And some of this literature has implications for how moral traits develop and change over time, which will in turn have implications for how identities change, if the essential moral self hypothesis is true. And there are also plenty of theories that try to account for the relation between traits and identity. But to my knowledge there is no research that observes the relation between traits and identity, as it is unclear what one of the relations even is (i.e., identity). The folk theory of what this relation amounts to (the folk theory of identity) cannot be easily dismissed in the way that other folk theories can be.<sup>10</sup>

The folk might also be wrong about how moral traits are individuated. In Strohminger and Nichols' studies, the moral traits under consideration were coarsely individuated, with less than a few dozen used throughout the studies. But it is possible that moral traits ought to be individuated more finely. If they are individuated more finely, then it becomes less clear that it is possible for moral enhancement to replace all of one's moral traits. However, if moral traits ought to be individuated more finely, then it is also more plausible that identity could change *without* a total change in moral traits, because individual traits would, on their own, contribute less to the formation of a person's identity. Either way, whether moral traits are individuated coarsely or finely doesn't affect the claim that if the essential moral self hypothesis is correct, moral enhancement can kill, and in some cases murder.

## 7. Conclusion

---

<sup>10</sup> However, it is possible that the folk theory of personality, whatever that is, is contradicted by the science.

Although moral enhancement can result in killing the person, and, if other conditions hold, murder the person, it seems that killing by manipulating moral traits is not as morally problematic as killing by manipulating physical traits, such as manipulations that lead to a non-functioning brain, heart, or other critical tissues.

If Sinnott-Armstrong & Miller (2012) are right about what makes death bad (it's the greatest loss), then the intuition that killing by moral enhancement is not as bad as killing by manipulating physical traits is incorrect. People who are killed by moral enhancement lose just as much as a people who are killed by causing damage to the body. Mental features such as memories and personality traits are all lost, as are future pleasures, experiences, abilities, and accomplishments. They will be someone else's, the person that the moral enhancement creates. It may be that the person's friends and family lose less, as they may believe that their loved one lives on after the moral enhancement. But it's not others' loss that makes killing bad; it's the loss to the person him or herself. And moral enhancement can bring about this great loss.

The killing of a person by moral enhancement has other interesting implications in bioethics. One implication is that people who object to human cloning ought to also object to moral enhancement. The reasons offered against human cloning often are about threats to the clone's individuality and thus his or her autonomy (Kass 1998). Others worry about human dignity (UNESCO, 'Universal Declaration on the Human Genome and Human Rights'). When moral enhancement kills a person, it creates another. This new person had his or her traits selected, and these traits, and the person him or herself, were created in a lab somewhere using scientific instruments rather than "natural" techniques of reproduction. The new person will be genetically similar, if not genetically identical to the person who was killed. Parity of reasoning

suggests that opponents of human cloning will also be opponents of moral enhancement.<sup>11</sup>

Further, unlike human clones who would presumably undergo the development from infant to adult, the persons created by moral enhancement would lack that development, instead starting out in a developmental stage similar to that of the person who preceded him or her in that body.

A third issue relates to the safety of those patients who volunteer for moral enhancement. Such patients would typically be informed of the risks of the treatment, risks which include death. How would those administering the enhancement ensure that enough moral traits persist so that the person persists? Even more concerning, how would anyone know whether a person has been killed by moral enhancement?

If the essential moral self hypothesis is true, then moral enhancement may have a significant impact on the subject's personal identity. Much attention in the bioethics literature has been paid to the notion of moral enhancement. But none of it has been directed at the possibility that it could be potentially lethal to the subject of the enhancement. I have argued that if the essential moral self hypothesis is true, and we have good reason to think that it is, then moral enhancement can kill. Though it is a remote possibility that moral enhancement will result in the widespread killing of morally mature adults, it is not so remote that it will result in the killing of psychopaths, or those that are otherwise vulnerable.

### References

Carter, S. 2016. Could Moral enhancement interventions be medically indicated? *Health Care Analysis*. <https://doi.org/10.1007/s10728-016-0320-8>

Crutchfield, P. 2016. The epistemology of moral bioenhancement. *Bioethics*, (30) 6, 389-396

---

<sup>11</sup> And that the responses to these reasons will be equally effective.

- Dennett, D. C. 1993. *Consciousness explained*. New York: Penguin Adult.
- De Dreu, C. K. W. 2012. Oxytocin modulates cooperation within and competition between groups: An integrative review and research agenda. *Hormones and Behavior*, 61(3), 419–428.
- Douglas, T. 2008. Moral enhancement. *Journal of Applied Philosophy*, 25(3), 228–245.
- Harris, J. 2011. Moral enhancement and freedom. *Bioethics*, 25(2), 102–111.
- Hurlemann, R., A. Patin, O.A. Onur, M. Cohen, T. Baumgartner, S. Metzler, ... K.. Kendrick. 2010. Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *Journal of Neuroscience*, 30(14), 4999–5007.
- Kass, L. R. 1998. The wisdom of repugnance: why we should ban the cloning of humans. *Valparaiso University Law Review. Valparaiso University. School of Law*, 32(2), 679–705.
- Kosfeld, M., M. Heinrichs, P.J. Zak, U. Fischbacher, and F. Fehr. 2005. Oxytocin increases trust in humans. *Nature*, 435(7042), 673–676.
- Larimer, Sarah. “Brutal obit slams Texas man whose life was ‘much longer than he deserved.’” *The Washington Post*, February 13, 2017. Accessed March 13, 2017 at [https://www.washingtonpost.com/news/post-nation/wp/2017/02/13/brutal-obit-slams-texas-man-whose-life-was-much-longer-than-he-deserved/?utm\\_term=.419e93f86dd8](https://www.washingtonpost.com/news/post-nation/wp/2017/02/13/brutal-obit-slams-texas-man-whose-life-was-much-longer-than-he-deserved/?utm_term=.419e93f86dd8)
- Mackie, J. L. 1990. *Ethics: inventing right and wrong*. New York: Penguin Books Limited.
- Nakazawa, E., K. Yamamoto, K. Tachibana, S. Toda, Y. Takimoto, and A Akabayashi. 2016. Ethics of decoded neurofeedback in clinical research, treatment, and moral enhancement. *AJOB: Neuroscience*, 7(2), 110-117.
- Nichols, S. 2004. *Sentimental rules: on the natural foundations of moral judgmentt*. Oxford: Oxford University Press.
- Persson, I., and J. Savulescu. 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., and J. Savulescu. 2013a. Getting moral enhancement right: the desirability of moral enhancement. *Bioethics*, 27(3), 124–31.
- Persson, I., and J. Savulescu. 2013. Should moral enhancement be compulsory? Reply to Vojin Rakic. *Journal of Medical Ethics*, 40, 251-252
- Persson, I., and J. Savulescu. 2014. *Unfit for the future: the need for moral enhancement*.



Oxford: Oxford University Press.

- Rakić, V. 2014. Voluntary moral enhancement is a solution to Sparrow's concerns. *The American Journal of Bioethics*, 4(4), 37–8.
- Sinnott-Armstrong, W., and F.G. Miller. 2012. What makes killing wrong? *Journal of Medical Ethics*, 39, 3-7
- Stanley, J. 2003. Context, interest relativity and the sorites. *Analysis*, 63(280), 269-281
- Strohming, N., and S. Nichols. 2014. The essential moral self. *Cognition*, 131(1), 159–171.
- UNESCO, “Universal Declaration on the Human Genome and Human Rights”, adopted 11 November 1997. (n.d.). Retrieved from [http://portal.unesco.org/en/ev.php-URL\\_ID=13177&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html)
- Williamson, T. 2002. *Knowledge and its limits*. Oxford: Oxford University Press.
- Wiseman, H. 2016. *The myth of the moral brain: the limits of moral enhancement*. Cambridge, MA: MIT Press.
- Zak, P. J., A.A. Stanton, S. Ahmadi, G. McGregor, U. Bickel, H. Fehm, ... B. Sampat. 2007. Oxytocin increases generosity in humans. *PLoS ONE*, 2(11), e1128.