

Direct vs. Indirect Moral Enhancement

ABSTRACT. Moral enhancement is an ostensibly laudable project. Who wouldn't want people to become more moral? Still, the project's approach is crucial. We can distinguish between two approaches for moral enhancement: direct and indirect. Direct moral enhancements aim at bringing about particular ideas, motives or behaviors. Indirect moral enhancements, by contrast, aim at making people more reliably produce the morally correct ideas, motives or behaviors without committing to the content of those ideas, motives and/or actions. I will argue, on Millian grounds, that the value of disagreement puts serious pressure on proposals for relatively widespread direct moral enhancement. A more acceptable path would be to focus instead on indirect moral enhancements while staying neutral, for the most part, on a wide range of substantive moral claims. I will outline what such indirect moral enhancement might look like, and why we should expect it to lead to general moral improvement.

INTRODUCTION

Determining the nature of morality and promoting philosophical ideals of moral behavior have historically been of great concern amongst philosophers and society as a whole. These concerns have involved not only trying to determine what is good and right but also trying to determine how to ensure that people will in fact be good and do the right thing. While the former has received a great amount of philosophical attention, the latter has—until recently—been somewhat overlooked. The attention that has been given to the question of how to inculcate values has focused on traditional methods such as education, (dis)incentives, and social pressure. In addition, recent scientific developments open up the prospect of influencing individuals' moral dispositions and behavior through biological interventions, particularly in the form of chemical, neurological, or—more speculatively—genetic manipulation.

There is a strong *prima facie* case for permitting, developing, and promoting moral enhancements. Moral failings are, almost by definition, problematic and indisputably worth overcoming. At the individual level, we try to convince ourselves or others into being moral—keeping promises, respecting others’ rights, acting kindly, and so on. And at a societal level, we endorse various policies aimed at promoting moral behavior (e.g., reducing crimes like theft or assault, preventing environmental damage, and subsidizing altruistic behavior through tax credits). Moral enhancement could be characterized as simply another intervention of this sort—one that could ultimately be more effective than previous efforts to induce morality and so especially worth promoting.¹ But even generally endorsed policies can have crucial flaws, as I will argue in the case of certain forms of moral enhancement.

In what follows, I will investigate a particular set of difficulties that emerge from the existence as well as importance of moral disagreement, and how to overcome them. The upshot will be that large-scale programs of direct moral enhancement contain a serious flaw, insofar as they will suppress dissent. However, I will argue that there is a path forward for moral enhancement. Moral enhancement can be best achieved by focusing on more indirect means, such as improving certain reasoning processes that will reliably lead to moral improvement.

There is, then, an important distinction between direct and indirect moral enhancements. A given intervention is a direct moral enhancement when it is designed to bring someone’s beliefs, motives, and/or actions in line with what the enhancer² believes are the correct moral beliefs, motives, and/or actions.³ So, for instance, if an enhancer believes that it is wrong to kill an innocent, then he would be performing a direct moral enhancement by inculcating the belief that murder is wrong, or by inculcating the motive or inclination to avoid murdering. An indirect moral enhancement, on the other hand, is designed to make people more reliably produce the morally correct ideas, motives, and/or actions without specifying the content of those ideas, motives, and/or actions.⁴ And though indirect moral enhancements do not rely on the particular substantive commitments of the enhancer, they will rely on the connections between certain processes and the correctness of moral beliefs, motives, and actions.⁵

THE VALUE OF DISAGREEMENT

We have some reason to be wary of direct moral enhancement because of the value of moral disagreement. Moral disagreement—while potentially

inhibiting consensus-building—is actually an important feature of society, one which would be under threat by some programs of direct moral enhancement.

The arguments in this section are primarily designed as an objection to widespread programs of direct moral enhancement—programs aimed at morally improving a large portion of a society (be it local, national, or global). The weak claim that individual, isolated instances of direct moral enhancement are permissible or justified (e.g., Douglas 2008; Douglas 2011) will not be directly addressed. Though the subsection on individuality might be applicable to some individual instances of moral enhancement, the discussion of indirect enhancement will open the door to a liberal policy of allowing individuals to morally enhance themselves as they please as a form of akrasia reduction. Instead, the primary target will be the stronger claim that we should embark on large-scale efforts to morally enhance entire groups of people (Persson and Savulescu 2008; Persson and Savulescu 2010; Walker 2009). There is, again, a somewhat compelling *prima facie* case for that stronger claim. Social policies generally have as their ends the betterment of the group; moral enhancement of the group should lead to a more moral society, which is surely desirable. But as we will see, that betterment may come at a serious, indeed prohibitive cost.

There are a variety of ways by which large-scale direct moral enhancement might be brought about. Most obvious would be federal coercion, with the state mandating that its citizens (or future, unborn citizens) undergo interventions such as moral education or pharmaceutical injections aimed at instilling the proper ideas, motives, or actions. Directly coercive policies are rather problematic for reasons other than disagreement—it is generally considered illegitimate for the state to interfere with people’s bodies without their consent (though notably not in all quarters; see, e.g., Fabre 2006). There are, however, other “softer” forms of large-scale enhancement that might not seem so objectionable. Incorporating moral instruction into public school curricula would not be substantially more coercive than current educational policies. The state might also selectively ban certain interventions it considers detrimental to moral belief, motive, or action, while allowing those that it judges to be moral improvements. Alternatively, the state could subsidize certain enhancements or put forward propaganda campaigns aimed at promoting widespread adoption of interventions aimed at moral improvement.⁶ The arguments in this paper will apply to all such methods of widespread direct moral enhancement, as they all employ a similar, objectionable method to

achieve moral improvement—bringing as large a portion of the population as possible in line with what the promoter of moral enhancement considers to be moral.

The following will, to a large extent, mirror arguments put forward by John Stuart Mill in *On Liberty*, especially chapters 2 and 3. The main aim of *On Liberty* was to defend the harm principle (a necessary, but not sufficient, condition on the acceptability of interference with an individual's liberty is that the intervention prevent harm to other individuals). In order to argue for that principle, Mill offers strident defenses of freedom of opinion as well as the value of individuality. This means not just the freedom to hold a dissenting opinion but also the freedom to criticize others and have the idea debated in public without censorship—thus applying not only to attempts to morally enhance beliefs and motives but also actions. While of course Mill was not thinking of the sort of biological interventions available today, his arguments are very much applicable to these relatively recent developments and speak powerfully against widespread direct moral enhancement. The utilitarian underpinnings of Mill's theory are not meant to be endorsed here, but I take his arguments for freedom of opinion to be generally sound and convincing (for reasons outlined below) whether or not one accepts utilitarianism.

THE SOURCES OF THE VALUE OF MORAL DISAGREEMENT

The value of moral disagreement can be derived from three somewhat interrelated sources: moral fallibility, reasoning, and individuality. Moral fallibility will entail a strong instrumental reason to preserve moral disagreement in a society, while moral reasoning and individuality are values threatened by the absence of moral disagreement. I will discuss each in turn; together, they constitute a cost to direct moral enhancement that will not be outweighed by the alleged benefits.

Moral fallibility

We are, without a doubt, fallible creatures. This is especially true when it comes to morality; the level of disagreement over moral issues should make us reluctant to claim true certainty about many moral claims. Indeed, proponents of direct moral enhancement will admit as much about the general population—if people were not morally fallible, there would be no need for direct moral enhancement. This, however, holds true for the moral claims made by proponents of direct moral enhancement themselves.

Any given program of direct moral enhancement, then, will run the serious risk of being wrong-headed.⁷

Why should this be a particular problem? After all, every government action—indeed, every action—is subject to moral fallibility. Such does not seem to be a general reason against action, so perhaps it is not a problem for a program of widespread direct moral enhancement. But Mill points out that certain sorts of actions are problematic in the face of fallibility—specifically, actions intended to stamp out dissent:

There is the greatest difference between presuming an opinion to be true, because, with every opportunity for contesting it, it has not been refuted, and assuming its truth for the purpose of not permitting its refutation. Complete liberty of contradicting and disproving our opinion, is the very condition which justifies us in assuming its truth for purposes of action and on no other terms can a being with human faculties have any rational assurance of being right. (Mill 1999, 62)

The idea is that, by suppressing dissent, we cut off a crucial avenue of coming to adopt the correct moral beliefs and policies. Without dissent, conventional wisdom will go unchallenged and moral progress becomes essentially impossible. This might not be a problem if we were infallible (i.e., already knew all the relevant moral truths), but because we are not, such actions will prevent the revision of morally odious policies that, at the time of suppression, seemed perfectly sound. Dissent is instrumentally valuable, then, as a constant check on the validity of the conventional moral wisdom of our time. Morality, in other words, should be allowed to evolve.

This position is particularly compelling if one is generally optimistic about moral progress. If we expect moral ideas held by the public to, by and large, become more and more in line with the truth, then there is very strong reason to want to preserve the ability of moral ideas to evolve. And there is some reason to think that human history has generally trended towards more morally upright positions. The gradual trend of societies towards toleration and civic inclusion of marginalized groups (e.g., women, minorities, foreigners, and more recently animals) as well as the spread of democratic political values are examples of this trend. But even if one thinks, on the contrary, that public morality has either not progressed, or indeed has regressed, one should still support preserving the ability of society to evolve so long as there is some potential for improvement. Allowing evolution does indeed run the risk of further regress. However, preventing moral evolution means shutting off any opportunity to improve

or to return to the (allegedly) morally superior days of yore. In other words, it means preventing one form of direct moral enhancement, the sort that occurs over generations and can lead to massive social improvements. This should be a serious concern for anyone who thinks that moral enhancement is a valuable enterprise.

Mill's intention was, admittedly, to argue against traditional forms of suppression of ideas (such as censorship and religious persecution). Still, the point is equally valid against programs of large-scale direct moral enhancement, where the ultimate effect is to reduce dissent. Consider direct moral enhancement involving inculcating certain ideas or motives. In order to increase people's attachment to those ideas or motives, there must be at least a relative reduction in contrary ideas or motives. This relative reduction is extensionally equivalent to the suppression of dissent: reduction in the extent to which people endorse or follow those contrary ideas or motives. This is an inevitable result of a large-scale program of direct moral enhancement; just as health enhancement reduces the burden of disease and cognitive enhancement reduces intellectual deficiency, direct moral enhancement reduces the prevalence of certain moral beliefs, motives, and actions that differ from the enhancers'.⁸

Yet, unlike disease and intellectual deficiency, those supposedly deficient moral states serve a valuable purpose, to challenge the conventional wisdom and ensure that "the means of setting it right are kept constantly at hand" (Mill 1999, 63). This applies even to enhancements aimed purely at improving behavior (e.g., Persson and Savulescu 2008); while such do not strictly require making people assent to what the enhancers take to be right, it is difficult to imagine ensuring widespread compliance with a particular moral standard without ensuring widespread agreement with that standard. The mechanisms for such compliance are almost always internal, after all—influencing how people think and process, in turn affecting how they act.⁹

Put another way, direct moral enhancement is problematic to the extent that it impedes moral progress. A private organization promoting direct moral enhancement to bring a few minds to its way of thinking may seem innocuous. But like any social problem, this becomes a serious issue if such "conversions" are widespread. The large-scale shift towards the organization's way of thinking via direct interventions reduces the diversity of thought in society in a way that does little to preserve true moral progress. An alternative scenario where a number of different organizations promote different moral enhancement programs is still

problematic, to the extent that divergence from moral positions endorsed by such organizations is suppressed. Furthermore, we should be suspicious of the moral reliability of the motives of well-funded organizations (as well as governments) who have strong incentive to push people not towards what is actually more moral, but towards what better serves those organizations' interests.

A natural response is to target specific, morally uncontroversial traits—for example, altruism and justice (Persson and Savulescu 2008) or wanton aggressiveness and racism (Douglas 2008). Given our general agreement and certainty surrounding the morality of such traits, it may seem unproblematic to bring the minority of egregious moral backsliders in line. However, while we may be supremely confident in the broad morality of such traits, legitimate disagreement emerges as soon as we delve a little deeper. Firstly, we need to determine the content of the trait—what, for instance, does justice consist in, and what notion of well-being will be deployed when altruistically promoting others' interests? There is certainly no great consensus in the philosophical literature on such issues, and arguably across society as well. An enhancer with a reasonably fine-grained intervention (necessary to isolate *only* those particular traits and not spill over into other areas of thought) would have to pick a controversial conception, impeding progress that might move beyond or even reject contemporaneous ideas. And secondly, the enhancer will as a matter of course be taking a stance on the relative strength and balance of such traits. How much altruism does morality demand? When is violence justified? Is race ever a relevant factor in decision-making? The demands of one set of values often conflict with others, and a direct enhancement program presupposes an answer concerning those issues. But we should not presume that, at present, we have a proper grasp of the appropriate balance of values; direct moral enhancement, in practice, threatens moral progress by presupposing just such a balance.

Alternatively, a proponent of direct moral enhancement might simply insist that the instrumental cost of such dangers to moral progress is outweighed by the instrumental benefit of moral enhancement. Avoiding catastrophes like nuclear war or environmental devastation through moral enhancement (Persson and Savulescu 2008; Persson and Savulescu 2010) is too important to let quibbles about cessation of moral progress get in the way; better to have a morally stagnant society than no society at all. Interestingly, Mill anticipated this objection, and has a useful reply. Such claims about the importance of uniformity have been made countless

times in the past to justify programs of suppression; for instance, even the wise philosopher–emperor Marcus Aurelius held that doctrinal unity was so supremely important to social cohesiveness that Christianity must be brutally suppressed. But we, like Marcus Aurelius, are fallible; while we can tolerate such fallible action insofar as errors can be corrected over time, the unique aspect of suppression is that it prevents such correction. One might think that direct moral enhancement of a certain sort is necessary to avoid one form of catastrophe; but we cut off the possibility of performing moral corrections that might avoid further catastrophes down the line. Such corrections are indeed necessary to ensure not only our well-being, but perhaps even our very survival.

Reasoning and deliberation

Dissent is important for ensuring moral progress, but it also has value independent of such down-the-line effects. We may want a society that not only has the correct sort of moral beliefs, but has them for the right sort of reasons. This value is somewhat hard to pin down, but Mill argues it is important for people to come to their ideas (especially moral ideas) through reasoning rather than external authority (Mill 1999, 80–81).¹⁰ This is directly an argument against social conformity and in favor of rational deliberation, but it also indirectly suggests that direct moral enhancement problematically cuts off our reasoning processes. Instead of coming to believe or act on a given moral proposition because it is the most reasonable, we would come to believe or act on it because a particular external agent (the enhancer) said it is best.

It is quite plausible to think that there is value in the process itself of deliberating over a moral proposition, both within one’s own mind and in discussion with others.¹¹ Part of this value might be instrumental (promoting better, more accurate ideas, as per the preceding subsection), but there is also a compelling sense in which reasoning is valuable in itself for the reasoner. This idea goes back at least to the *Apology*, where Socrates famously argues that “the greatest good of a man is daily to converse about virtue . . . and the life which is unexamined is not worth living” (38a). We need not adopt this apparently extreme a view of the value of moral discourse and contemplation to agree with its core insight—there is something intrinsically good about such reasoning, something worth promoting and protecting.¹²

Yet, widespread direct moral enhancement would in all likelihood reduce that discourse. People tend not to debate or reflect much on issues

about which there is no doubt or dissent. Discourse and contemplation are motivated in part by the existence of disagreement. This makes sense, after all; the purpose of such discussion tends to be the discovery of truth through reason. Disagreement prompts one to either doubt whether one's own ideas are actually true, or (perhaps more commonly) attempt to correct the opinions of others through discourse. But even the latter motivation can lead to real revision—in the process of defending a sincerely held position, one might come to notice its flaws and correct accordingly. As we have seen, widespread direct moral enhancement would vastly reduce that disagreement; it would therefore also lead to significantly less moral discourse (at least over issues pertinent to what is being enhanced) that seems so valuable.¹³

Perhaps one could try to preserve the value of reasoning in a morally unified society by promoting debates where one side pretends to hold views that everyone takes to be false, and the other holds views that all assent to. But that misses out on the apparent value of the discourse (indeed, it comes close to the sort of sophistry decried by Socrates). Reasoning consists not just speaking certain words, but actually and sincerely entertaining the possibility that one is wrong—and being open to revision of one's beliefs or actions in the face of error. Structured discussion without true disagreement would be only a vague facsimile of the lively moral debates that pervade modern society—more akin to the “Two Minutes of Hate” from *Nineteen Eighty-Four* than true deliberation.

Even without a unified society, moral enhancement will be problematic to the extent that one's ideas can no longer be said to be the product of one's own reasoning. Direct moral enhancement makes one's own reasoning process more or less obsolete; even if one goes through the motions of reasoning, the enhancers have “rigged the game,” so to speak, to ensure the desired outcome. This will be the case even with some non-coercive direct moral enhancement. To the extent that people voluntarily give up on reasoning, they will be abandoning something that is of immense personal value.

Alternatively, one might promote a form of direct moral enhancement that operates on reasoning processes. Suppose, for instance, that an enhancer believes altruism is good, and wants to make people more altruistic. But this enhancer also believes that the best way to bring about more altruism is to improve on certain reasoning processes (in line with the strategy discussed in the section on indirect moral enhancement below). Such a strategy, on its face, could respect the importance of people's

ability to reason and think for themselves. However, this strategy still falls into the pitfalls listed above. The forms of reasoning inculcated will be narrowly construed as those leading to the enhancer's way of thinking. This runs the risk of the enhancer making everyone think the same, in order that everyone comes to be in agreement with the enhancer's way of thinking. What's more, the narrow conception of reasoning promoted by the enhancer would be justified not based on the merits of the reasoning process itself, but the conclusions that process draws. This puts the cart before the horse in ensuring that, ultimately, we do not come to believe and act in certain ways because those beliefs or actions are supported by the best reasoning processes, but instead because they are supported by whatever reasoning processes the enhancer thought efficient at bringing about certain effects.¹⁴

Finally, one could attempt to simply alter people's basic intuitions or emotive responses that inform reasoning processes. The idea would be to leave the reasoning process itself intact and thus ostensibly preserving the role of reasoning in the agent's actions. This program would perhaps be more desirable than a program that attempted to work around or subvert one's reasoning processes. Still, the value of reasoning processes is nevertheless impugned under such a program. When the enhancer decides to alter intuitive or emotive responses, there is a clear interference with the agent's reasoning and deliberation. Intuitions or emotions are clear components of any reasoning process, after all—they will be weighed up against other considerations, are themselves subject to various introspective forces, and may determine the content of various premises used in explicitly deductive reasoning. This manipulates an agent's reasoning process for the sake of bringing about a particular outcome, without any regard for the integrity or value of that process itself.

What's more, the difference between the enhancer and the enhancee would be over what are the proper basic moral intuitions to hold. But why should we believe that the enhancer's basic intuitions are any more reliable than the enhancee's? The correctness of moral views is not independently verifiable—not without further normative assumptions, anyway. And faults of reasoning cannot be appealed to, as these basic intuitions operate independently. So we can't use some external grounds for privileging one party's moral intuitions. One could posit that the enhancer has a special moral faculty that others lack, or some special metaphysical access to the moral truths. But such special privileges seem absurd and unjustified—where are these special abilities supposed to come from, and moreover

what evidence is there that others lack them? There does not seem to be a general reason to privilege the enhancer's over the enhancee's intuitions, and so there is not much reason to think that any alteration of those intuitions in favor of the enhancer's would lead to moral improvement.¹⁵

Individuality

A related source of the value of disagreement is the importance of individuality. It is not only important that one's beliefs, motives, and actions are the result of reasoning, or lead to good outcomes; it is important that those beliefs, motives, and actions are one's own. Mill takes this individuality to be a singular human value:

It is not by wearing down into uniformity all that is individual in themselves, but by cultivating it and calling it forth, within the limits imposed by the rights and interests of others, that human beings become a noble and beautiful object of contemplation. (Mill 1999, 109)

Lack of individuality is a threat to such cultivation insofar as it makes people more like steam-engines—passive, mechanistic beings under the instrumental control of another—than men and women of true character (Mill 1999, 107). Though Mill did not use the term autonomy, it could be said that lack of individuality is a threat to autonomy, insofar as it would make acts and thoughts less autonomous and more driven by the enhancer's motives.¹⁶ Diversity of opinion—especially moral opinion—would importantly preserve that autonomous character, ensuring that people retain their individual nature that we value so much.

Moreover, individuality is not just about the structure of conscious moral thought; it is also about the origin of those thoughts. It is important that we ourselves, and not others, are the originators of those thoughts. The steam-engine analogy is disturbing not because steam-engines conform, but because they have no independent will of their own. A society that imposes its will on the actions and thoughts of its members thus robs those people of their individuality; the group's will (or the will of the group's thought-leaders) is substituted for the individual's will. Similarly, direct moral enhancement replaces the individual's will with the will of the enhancer.¹⁷

This danger is apparent in traditional forms of moral enhancement, but the nature of biological enhancements makes the will-substitution especially problematic. Society has traditionally affected thoughts only indirectly and externally—putting pressure on those who dissent (either

with coercion or softer methods), while rewarding those who assent. People will tend to adjust their ways of thinking in response to those social forces, but at least that adjustment is within people's control. One can choose whether to listen to an argument, accept punishment as a reason to revise one's moral views, or revise one's views in the face of social pressure. External pressure can sometimes be extremely difficult to resist, but biological moral enhancement typically admits of much less individual control. Such interventions are especially direct in how they affect one's mind; one's brain-chemistry is itself altered, and the intervention itself consists in manipulation of one's very personhood, rather than that alteration being a contingent effect. While one could, if aware of the intervention, attempt to compensate for its effect (say, taking less seriously one's intuitions when one knows they're strongly affected by some manipulation), the effect will still at its outset bypass and subvert conscious processing in order to bring about improvement. Some analogy can be drawn to psychological manipulations that also bypass reasoning processes, so the difference between biological and non-biological means is not completely categorical—but biological approaches can potentially be more efficient than such external manipulations, to the extent that biological enhancements can isolate particular psychological functions and bypass the reasoning process more easily than in traditional approaches.

Against this, it might be pointed out that already people's moral beliefs, motives, and actions are entirely caused by external forces. No one is an uncaused causer. Individuality, then, is an illusion with no real value; there is nothing bad about coming to have ideas or inclinations due to direct moral enhancement because the alternative is to have those ideas or inclinations due to the vagaries of genetics, environment, and society. We all are essentially like steam engines already (albeit with consciousness). Because there is no way to be otherwise, it is pointless to oppose any action on the grounds that it might make us lose control.

This objection, of course, paints a portrait of human nature that will strike many as being quite bleak. In addition to making us not much more than self-aware steam engines, it implies that the notion of moral responsibility is mistaken. In order to ascribe moral responsibility to an agent, one must be able to distinguish between cases where an agent is and is not responsible for his or her actions. We can contrast the following cases: (1) Alex punches Ben because she intensely dislikes Ben and (2) Alex punches Ben because Carlos has utilized a mind-control device that compels Alex to do so. There is a clear intuitive difference in

Alex's moral responsibility for hitting Ben in the two cases, but the view that individuality is an illusion would deny this. The fact that Carlos has substituted his will for Alex's in (2) makes no material difference, as there were some analogous external influences (genetics, upbringing, etc.) to determine Alex's actions in (1) as well. But this rejection of a difference is a deeply implausible implication, one that suggests we should reject the individuality-as-illusion account. Instead, we should prefer an account where the origin of some action (whether internal or external) is morally relevant.¹⁸

Compatibilism is just such an account.¹⁹ There is not space to fully explicate or defend this view here, but it is quite appealing to think that notions like responsibility can be preserved in the face of all these external determinants of action. There is something valuable and morally transformative about volition and personal choice, such that we can indeed properly attribute responsibility to people whose actions result from such personal choice and see responsibility as reduced when personal choice is reduced. Yet distinguishing (1) from (2) involves something more—not just valuing personal choice, but privileging certain origins of choice over others. Cases like (2) are particularly problematic because another agent is involved—one agent's will is substituted for another's. This particularly denigrates the extent to which the agent can identify with her own actions, affecting not only moral responsibility but the broader moral connection between who she is and what she does.

If that is right, then it is no stretch to attribute value to individuality and object to direct moral enhancement on the grounds that it involves will-substitution. Having a certain moral nature due to another's will diminishes this value of individuality, in a way that having that moral nature due to natural chance or reasoned deliberation does not. (Indeed, we might form a hierarchy of value: belief or action due to reason is better than it being due to nature; but both are to be preferred to being due to another's will.) Certain direct interventions will be more deleterious to individuality than others, but they all rely on a privileging of the enhancer's moral view over that of the enhancee. In this way, direct moral enhancement involves substituting the enhancer's will (qua deciding what is good or desirable) for that of the enhancee.

Summation

Taken together, our fallibility and the value of reasoning and individuality indicate serious instrumental and non-instrumental costs engendered by a

program of widespread direct moral enhancement.²⁰ Such could impede moral progress and leave its members without the moral deliberation and individuality that we rightfully find of value. How weighty these considerations are is a matter of debate. For my part, I believe the above arguments show that the value of moral disagreement is sufficient to make widespread promotion of direct moral enhancement an untenable and wrongheaded policy. But perhaps, despite what has been said, one still thinks that the catastrophic consequences of not engaging in direct moral enhancement are too great to avoid embarking on widespread direct moral enhancement, or that some ideas, motives, or actions are so clearly wrong that ridding people of them cannot be objectionable. To such people, I would nevertheless urge that they take these costs seriously and do all they can to mitigate them, to whatever extent possible.

INDIRECT MORAL ENHANCEMENT

As defined above, indirect moral enhancements do not involve the aim of instilling particular moral ideas, motives, or behaviors in people. Instead, indirect moral enhancement focuses on improving the processes by which moral ideas, motives, and behaviors are generated. Importantly, this allows the enhancer to remain neutral on a wide range of substantive moral positions. The enhancer cannot be completely substantively neutral, as the identification of factors that will lead to more reliably moral outcomes is itself a substantive issue. However, the range and type of substantive issues within the scope of the enhancer are severely limited. It also means that the success of an enhancement will not be measured by whether an enhancee has the correct thoughts or actions, but rather that certain capacities are operating in a certain way.

MODES OF INDIRECT ENHANCEMENT

What, precisely, would indirect moral enhancement look like? John Harris, though critical of more direct emotional moral enhancement, has suggested that cognitive improvements could be moral enhancements (2011). However, Harris's comments on this front are vague. He appears to suggest that cognitive improvements can be moral enhancements insofar as they make people's non-moral beliefs more accurate, which can be morally relevant when those non-moral beliefs figure in a moral argument. For example, take the moral statement, "all acts of killing are morally wrong," and the non-moral statement "Joan killed Anne." The moral question of whether Joan acted wrongly crucially depends on the

non-moral question of whether Joan killed Anne. Perhaps certain cognitive improvements could make judgments about evidence that Joan killed Anne more accurate, improve the memory of a witness, or even improve one's ability to grapple with the concept of killing as it is applied to this particular case. In that way, cognitive improvements could indeed make judgments of morality more accurate. And these improvements would be classified as indirect moral enhancements, insofar as the enhancements make moral ideas more reliably accurate without the enhancer being committed to particular moral outcomes.

Still, this form of indirect moral enhancement is rather limited, improving only derivative moral claims. Proponents of moral enhancement typically aim at wider-scoped improvements that might affect moral ideas, motives, and behaviors that are not derived from particular non-moral beliefs. I will now propose two means of indirect moral enhancement that target more than just non-moral beliefs.²¹ The first—improving reasoning—is a close cousin of Harris's proposal, but operates on a broader set of domains. The second, *akrasia* reduction, is more distinct, though its classification as indirect enhancement may be controversial. It should be noted that all these enhancements will be discussed at the theoretical level; particular interventions that might be used will not be identified, as such empirical analysis is outside the scope of this paper. Still, the various means of indirect moral enhancement identified here should point the way towards what sort of interventions could be deployed in practice.

Enhancing reasoning

“Reasoning” is a vague and ill-defined term. Rather than offer a speculative definition here, I will instead point to several features of human thought that could sensibly be understood as aspects of reasoning. The important point is not that each fits under a broader concept, but just that improving on each would predictably and reliably lead to improved moral thought, motives, and action. As indirect moral enhancers are by and large neutral on particular substantive issues, this reliability will not be determined by the likelihood of bringing about particular outcomes but by the expected connection between the processes and the proper moral ideas, motives, and behaviors—whatever those may be.

Perhaps the easiest target for indirect moral enhancement would be improving logical understanding. This concerns people's ability to make proper logical inferences and deductions, spot contradictions in their own beliefs and those of others, as well as formulate arguments in a way that

can highlight the true point of contention between interlocutors. Logical reasoning will be especially important for the formation of proper moral beliefs. Whatever the correct moral views are, we should expect those views to be consistent with one another. Also, proper justification of moral ideas will require coherent argumentation. It is of course possible that some moral views do not easily admit of logical analysis (for example, basic, non-inferred intuitions). Still, more complex views will require at least some logical reasoning to ensure that people do not descend into moral incoherence. Insofar as we would expect people's motives and behavior to be informed by their moral judgments, logical coherence then should improve the reliability of those outcomes as well.

Another morally important reasoning process is conceptual understanding. Having what Descartes referred to as clear and distinct ideas is morally relevant in a number of ways. In order to understand the implications of a particular moral idea (say, killing is wrong), one must have some grasp of the notions involved (in this case, not just wrongness but what exactly constitutes killing). Vague and distorted ideas will lead to unreliable inferences, inducing behaviors that are not in line with someone's considered judgments.²² Moreover, in evaluating competing moral ideas it is crucial to fully understand the various claims that are being made. Misapprehension of the concepts can easily lead to misidentification of which argument is stronger or which position more compelling. And again, the moral reliability of motives and behaviors are at least to some extent informed by the reliability of a person's moral ideas. Improving someone's ability to grasp a wide variety of moral concepts will then help reduce moral errors.

Proper reasoning also involves moral ideas, motives, and behaviors being influenced only by morally relevant factors. It is far beyond the scope of this paper to identify what exactly those relevant factors are, and indeed any answer would be very contentious. Less controversial, however, are the factors that are generally not morally relevant—what are generally referred to as cognitive biases. For example, confirmation bias involves giving inordinate weight to factors that confirm one's pre-conceived ideas. This is a threat to morality insofar as it inhibits moral improvement (recall the section on fallibility above). Other relevant biases include the fundamental attribution error (dissonance between self-judgments and the judgments of others), conformation bias (conforming to the sometimes-absurd beliefs of others), and framing bias (how a question is posed or phrased can affect the outcome). By mitigating those biases, we could ensure that moral ideas

are derived only from morally appropriate factors, which would inevitably be more reliable than the sorts of irrelevant concerns that affect a wide range of biases. There may be other ways to improve reasoning processes as a means of indirect moral enhancement, but those three are at least the most salient. This is not to say that improving reasoning in the above ways will necessarily make someone more moral. However, we can reasonably predict that they would lead to more moral outcomes by and large—even if not everyone affected by the enhancements is so improved.²³

Akrasia

Akrasia refers to acting against one's better judgment. "Better judgment" is typically understood not in normative but cognitive terms, referring to consciously-entertained opinions about what is all-things-considered best to do. Often, reducing akrasia appears to be in someone's interest; say, a person recognizes that they need to lose weight but frequently succumbs to the temptation to eat sweets. Here, it will be argued that akrasia reduction can also reliably make people more moral, at least in their behavior and possibly motives.²⁴ Two avenues of such akrasia reduction will be discussed. The first involves internally reducing someone's weakness of will. The second is somewhat more complicated, and involves allowing direct moral self-enhancement.

Weakness of will affects morality in a very straightforward way. Someone recognizes that some course of action is morally ideal or morally required, but nevertheless fails to carry out that action. For instance, someone might recognize the moral imperative to donate significant sums of money to charity because that money could save a number of lives, yet remain selfishly tight-fisted. This is a failure of someone's consciously-held moral judgments to be effective. They are, more importantly, apparently outweighed by considerations that are not explicitly moral and/or not fully cognizant of all the relevant moral considerations (self-interest may be morally relevant, but the ultimate disinclination to donate fails to take into account the overriding altruistic reasoning that is factored into the moral judgment that one should donate). By reducing weakness of will in people, we could help ensure that those considered judgments more often properly inform action. This sort of approach has the added benefit of potentially improving someone's autonomy, assisting them in controlling the course of their own lives (contrasted with the more heavy-handed, individuality-suppressing direct moral enhancements).²⁵

One problem with this approach is that considered moral judgments are not always right. Jonathan Bennett (1974) has astutely observed a case of weakness of will leading to right action from the Mark Twain novel *Huckleberry Finn*. Huckleberry Finn is travelling with runaway slave Jim, contemplating whether or not to turn Jim in. On consideration, Huck believes that helping Jim escape is morally wrong (specifically, a wrong to Jim's owner). However, when the moment comes to turn Jim in, Huck has a moment of what he thinks of as weakness and relents. It is arguable that Huck did indeed fail to turn in Jim for the right sort of reasons (his sympathies towards Jim), but it is fairly clear that at least Huck considers this to be a moral failing. This case illustrates a broader point that reducing weakness of will would not reliably make people more moral, as weakness of will is just as likely to lead to the right action as the wrong one.

However, the Huckleberry Finn case is so striking because it is so unusual—that is, it is unusual that weakness of will would lead to the morally right action. And for some intervention to count as a reliable moral improvement, it need not always improve someone's morality—just do with relatively high frequency.²⁶ It is noteworthy that the Huckleberry Finn case is indeed a work of fiction and so somewhat contrived. When contemplating instances where someone doesn't do what they think they should do, the most frequent cases that come to mind are much closer to the donation case than the Huckleberry Finn case. More significantly, we have good reason to expect that weakness of will is more likely to lead to a moral failing. The cognitive functions that generate the moral judgment are sensitive to a wider range of considerations than the inclinations that drive weakness of will, which are typically much narrower. In addition, the fact that considered judgments are considered—the product of at least some deliberation, consideration of competing arguments and evaluation of persuasiveness—should give us considerably more confidence in the veracity of those judgments than those that drive weakness of will.

Reducing someone's internal weakness of will is not the only way of reducing akrasia; alternatively, one could attempt to manipulate the akratic desires and inclinations directly. This may not lead to improvements across a wide range of domains in the manner of general reduction in weakness of will, but it could at least locally make people's actions follow their considered judgments. At first glance, this approach may appear to run afoul of the objections noted above. After all, this appears to be a reversion to just the sorts of direct moral enhancements that are inimical to dissent, and in particular go against the value of individuality (the enhancer would

be imposing his or her ideas about the importance of akrasia reduction on the enhancees).

To avoid this problem, akrasia reduction would not involve forcing individuals to alter their basic inclinations to match their considered judgments. Instead, the state would simply be allowing direct moral self-enhancement, and perhaps researching means by which such self-enhancement could be brought about. At no point would people be pressured or cajoled into undergoing the intervention, obviating the problem of enhancers imposing their wills on others. Moreover, because people's antecedent considered judgments would be preserved, the ability and propensity of people to dissent would be maintained.

Allowing direct moral self-enhancement will generally lead to moral improvement for the same reason that reducing weakness of will would: we reasonably have more trust in people's considered judgments than their more basic inclinations. This is even clearer in the case of moral self-enhancement; only those cases where one's considered judgments are of sufficient force to motivate seeking out a (potentially invasive) enhancement intervention would be improved. This might indicate a limitation in how large-scale a program of allowing moral self-enhancements could actually be. Perhaps only a handful of people would be sufficiently motivated to undergo the changes. Be that as it may, such a program would at least have the advantage of avoiding the objections adduced above. Even if only a small number of people actually underwent moral self-enhancement, there is good reason to think that those who do will generally be improved.²⁷

PRESERVING DISAGREEMENT

A key advantage of indirect moral enhancements is that they can avoid the above objections based on the value of disagreement.²⁸ Consider first fallibility. The problem with direct moral enhancement is that it could induce inordinate affinity for the present, likely flawed moral views of the enhancer. But indirect moral enhancement need involve no such presentist bias. Instead of being directed towards particular moral ideas, motives, or behaviors approved of by the enhancer, people's ideas, motives, and behaviors would be the result of newly-improved processes. As long as these altered processes are not overly restrictive, this should lead to more opportunity for moral growth. Social moral improvement would occur as these processes generally lead to more accurate recognition of the proper moral ideas, motives, and behavior.

The importance of reasoning and deliberation will also remain intact under indirect moral enhancement programs. As indirect enhancers are not picking processes based on their likelihood of generating particular outcomes, we should not expect immediate convergence and a cessation of debate. On the contrary, disagreement and discourse will continue as people use their newly-improved capacities to more acutely grapple with a wide variety of moral dilemmas. In fact, one of the most promising ways to effect indirect moral enhancement (discussed above) will be to improve people's reasoning abilities. If one takes the value of both internal and external deliberation seriously, one should be encouraged by enhancements aimed at improving those deliberative capacities.

Individuality is arguably more difficult to preserve. Indirect moral enhancement still involves the enhancer imposing his or her ideas on the enhancee (in this case, ideas about the most morally reliable processes). Still, the risk of will-substitution is significantly diminished. With direct moral enhancement, individuality is threatened as one's moral ideas, motives, and behaviors are generated not by one's own autonomous thought and deliberation but ultimately by the judgment of the enhancer. Indirect moral enhancement, by contrast, involves no such tight connection. The enhancer is not imposing his or her particular values on the enhancee. The enhancee's ideas, motives, and actions are still properly the result of his or her own thought and deliberation. Indeed, it may well turn out that the enhancee endorses values the enhancer strongly disagrees with; perhaps that is just the inevitable result of improved deliberation, as the enhancer's own moral errors are corrected. This possibility breaks down the asymmetry between enhancer and enhancee, preserving the sense in which someone's ideas are properly their own.

Still, we must be careful. The notions of reasoning and deliberation described above are by no means universally endorsed. Individuals of certain faith traditions, or dialethists who believe in the existence of true contradictions, will reject some of the aspects of reasoning endorsed here. It might appear, then, that indirect enhancers would be imposing their reasoning values on others. However, none of the features of reasoning were claimed to be intrinsically valuable. It is not a success condition of indirect moral enhancement that enhancees come to believe that it is in itself morally better to be more coherent. Rather, the claim is that these processes will make people more reliable at determining for themselves what is of value. This could be understood as the inculcation of an idea that is related to value, even if it is not a value itself. But this inculcation is

not a threat to individuality any more than instruction concerning certain empirical facts about history, science, and so on—they are more general facts about the world, rather than particular claims about value.²⁹

CONCLUSIONS

In the preceding, I have attempted to identify a key problem with proposals for widespread direct moral enhancement. They threaten the value of disagreement, and more particularly could lock us in to fallible moral ideas, subvert reasoning (both internal and external), and threaten individuality. These concerns may not decisively militate against a program of widespread direct moral enhancement, especially if such a program is necessary to ensure the survival of humanity. Still, they point towards a serious flaw with such proposals, and we should seek to avoid them if at all possible.

Indirect moral enhancements are attractive in part because they avoid running afoul of the value of disagreement. When enhancers do not commit themselves to the morality of particular outcomes, they will allow individuals significant room to think and act for themselves—deliberations will continue, ideas evolve, and people still govern their own thoughts. In addition, indirect moral enhancement has some independent appeal. There is indeed a plausible connection between reasoning properly and coming to the correct moral conclusions. Indeed, arguably the entire enterprise of moral philosophy is predicated on such an idea that thinking these matters through can lead to the right moral answers—or at least point us in the right direction.

These arguments also suggest a few avenues that moral enhancement should take. Traditional schemes of moral enhancement such as moral education should aim to be as substantively neutral as possible. Government agencies and charitable organizations may still take strong moral stands, of course, but they should take care in the use of direct moral enhancement to instill those views on others. This is not so implausible, as it coheres well with liberal neutrality in education prompted by Rawlsian thought (Waldren 2011). Public moral education, on this view, should be more philosophical than dogmatic (if present at all)—teaching students how to think through moral issues, without insisting on what the proper moral answers are. There is perhaps more leeway for allowing more dogmatic private moral education of one's children, which is in some ways similar to allowing direct moral self-enhancement. Generally, parents are given significant license to raise children as they see fit, and children's limited

autonomy will help mitigate some of the concerns expressed above. Moreover, a state policy of simply allowing such behavior will avoid the issue of inhibiting general disagreement in society.

As for more novel biomedical means, we should be focusing less on interventions that will bring about particular outcomes—altruism, non-violence, and so on. Instead, we should research interventions that can improve the processes that reliably lead to good ideas, motives, and actions. Research into cognitive enhancement, then, would be geared not just towards making people “smarter,” but better able to process, understand, and analyze moral issues. The downside to this approach is lack of full verifiability. Enhancers will not be in a position to evaluate whether a program of indirect moral enhancement has succeeded in producing moral improvements. This is because any such attempt to verify outcomes would only lead those interventions to become direct moral enhancements—they would be deemed a success or failure based on the enhancer’s own moral views, which is precisely the situation that indirect moral enhancement seeks to avoid. Still, researchers will not be completely in the dark—one can evaluate the effects of interventions on the processes themselves, and make judgments of success on those grounds.

A program of indirect moral enhancement will, then, perhaps not be as obviously revolutionary and salutary as some (such as Persson and Savulescu) would hope. Still, we should have confidence in people’s abilities to reason for themselves and more often than not—when those reasoning processes are functioning properly—come to the right conclusions and actions on their own. And if there is not ultimately a convergence of moral thought after such a program has been enacted, this is not a reason to despair. On the contrary, we should be happy to have a wide diversity of thought in society, and encourage the inevitably vigorous discourse that will follow. It may not always lead to the right answers, but it should in any case tend in the right direction.

ACKNOWLEDGMENTS

For invaluable comments on various versions of this paper, I thank Roger Crisp, Tom Douglas, Guy Kahane, Jordan Ridgewell, Wenwen Fan, the two anonymous reviewers for the *Kennedy Institute of Ethics Journal*, members of the Oxford Applied Ethics Discussion Group, attendees at the 2012 Central States Philosophical Association Conference, attendees at the Seventh International Conference on Applied Ethics, attendees at the 2012 Dutch Conference on Practical Philosophy, and attendees at the 2013 International Conference on Enhancement: Cognitive, Mood and Moral.

NOTES

1. On this understanding and the definitions of direct and indirect enhancement given below, moral enhancement would strictly speaking include external interventions like argumentation and punishment. However, the literature on moral enhancement has primarily focused on what might be called ‘internal’ enhancements (especially bioenhancements), which operate via altering internal ideas, emotions, and motives (thereby generating good behavior) rather than merely incentivizing behavior through external means. Most of this article will focus on such internal interventions, though moral education will be discussed at various points. See footnote 9 below for a brief discussion of punishment. As for argumentation, while it may technically be a form of direct moral enhancement, it contains a structure (appealing to reasons, leaving it up to the interlocutor what to decide) that, in my view, will avoid the bulk of the objections given here.
2. “Enhancer” will be understood as the individual or individuals who intend to use moral enhancement to generate improvement in society; depending on the context, this might be the state, a charity, or some other body—though one with the power to implement the program in at least a somewhat large portion of the population.
3. I will not assume that the enhancer’s beliefs are correct, as this would be question-begging against the fallibility objection below. Indeed, some have argued that moral enhancement just consists in enhancing what one believes to be moral (see, e.g., Shook 2012). Still, even if one adopts an objectivist account according to which moral enhancement occurs only when the enhancer is correct, the further objections concerning reasoning and individuality will apply.
4. As indicated by the notion of “design,” the direct/indirect distinction will rely in part on the enhancer’s intentions. The direct enhancer aims at inculcating particular ideas, motives, or behaviors; the indirect enhancer aims at facilitating moral improvement but does not (when it comes to determining what counts as moral enhancement) prejudge whether a resultant idea, motive, or behavior is moral.
5. Shaw (2014) has a somewhat similar account, though her focus is on the criminal justice context and the notion of responsibility, whereas the focus here is more general moral enhancement and the implications of Mill’s plausible arguments against censorship.
6. As mentioned in footnote 1 above, argumentation is outside the scope of the present objections. Propaganda, though, characteristically bypasses reasoning processes to generate conformity, making it susceptible to the present

- critiques. Likewise, dogmatic moral education that hammers in principles without reflection will fall afoul of the objections, while a more circumspect approach that engages students in reasoned discussion would be preferable.
7. A proponent of moral enhancement might claim that moral failings come not from incorrect moral ideas, but rather failed motivations in light of those ideas. On this view, moral enhancement aims not at correcting ideas but simply motivating people to act in accordance with the ideas they have. However, it is extremely dubious that people are infallible regarding moral ideas; the existence of widespread disagreement over a whole host of substantive moral issues, including fundamental concerns like human rights and the value of life, is incompatible with universal infallibility. Moreover, an enhancement program aimed only at having people act in accordance with their considered moral views would actually be an indirect approach, avoiding the present objections. Akrasia reduction as moral enhancement is discussed in more detail in the section on indirect moral enhancement below.
 8. One might argue that certain programs of moral enhancement target conative states, leaving beliefs and ideas alone. However, conative states will inevitably have very powerful effects on people's judgments. Intuitive judgments, in particular, are very susceptible to conative influence and form a strong basis for many considered judgments.
 9. Purely external strategies such as punishment, which need involve no internalization of the enhancer's norms, could avoid these worries. To the extent that punishment leaves ideas or motives intact (i.e., we are not punishing the holding or dissemination of particular ideas or motives), such strategies are relatively immune to the concerns listed here. Strictly speaking, punishment would count as a form of direct moral enhancement. But, my argument does imply punishment should not be designed to operate by affecting people's ideas or motives themselves. This would not have overly radical implications for punishment in practice, as non-internalized incentives to avoid criminal activity and/or punitive retributivism will still be sufficient to justify imprisonment. Also, communicative retributivism (where punishment is a way to convince offenders of the wrongness of their acts) would be preserved, to the extent that punishment is actually a sort of rational argument. As mentioned in footnote 1 above, the present objections do not apply to argumentation. In any event, these issues are outside the usual scope of what people refer to as "moral enhancement"; the concern of this paper is more paradigmatic methods, which tend to operate at least partially by changing people's internal norms.

10. At times, Mill seems to say that, beyond the importance of reasoning, it is important for opinions to have a foil in order that ideas would have a “clearer perception and livelier impression of the truth” (Mill 1999, 59–60). The value of such “liveliness” is significantly less convincing (and less clear) than the value of coming to a view through reasoning and deliberation. And, it may just be that such liveliness is not meant to be an internal characteristic, but rather just one way of explicating the value of (lively) deliberation over ideas. In any event, I will set aside the “liveliness” conception of the value of disagreement for the present discussion.
11. This is not to say that the notion or value of reason is exhausted by such a process. Rational intuitionism, for instance, takes basic, intuitive, non-procedural grasp of certain concepts to be a part of reason, and arguably there is value in itself to properly grasping such basic concepts (Audi 2005). The present argument is consistent with such a view, so long as one leaves room for the existence and value a procedural component to reason.
12. One recent suggestion comes from Alison Hills (2009): there is value in moral understanding, which is usually not acquired via the moral testimony of others but through one’s own reflection and deliberation. Direct moral enhancement would be problematic for the same reason that accepting moral testimony is—it subverts or works around those personal deliberative processes. Reasoning, on this account, is not merely instrumentally valuable in bringing about moral understanding; it partially constitutes moral understanding.
13. It is of course possible to have too much discourse, in the sense that one spends too much time/resources debating when a decision should be made. However, the proper way to avoid inappropriately prolonged debate is not to directly induce agreement, but rather improve people’s ability to discern when a debate has run its course. This skill is a component of sound reasoning (as I will discuss later on), and improving it will be a form of indirect rather than direct moral enhancement.
14. The primary problem here is that reasoning processes are selected not because they are in themselves the best or most reliable procedures, but because they produce certain results. Below, I will suggest that improving reasoning can be an acceptable form of indirect moral enhancements, insofar as the opposite is true: reasoning processes are selected not because of the particular results they generate but because they are morally reliable in virtue of the processes themselves. The key difference is that the indirect moral enhancer will make no reference to specific, desired outcomes of a given reasoning process, letting the agent’s reasoning process itself—rather than the enhancer—be the real determinant of the agent’s beliefs, motives, and actions.

15. There have been some attempts in the literature on peer disagreement (e.g., Elga 2007) to justify continuing to hold onto one's own moral views in the face of disagreement. A possible strategy, then, is to favor the enhancer to the extent that one personally agrees with the enhancer's views. There is not space here to fully engage with this response, but suffice it to say that I do not find such arguments persuasive. As Simpson (2013) notes, such arguments have difficulty justifying the privileging of one's own views over others'; to do so seems dogmatic and self-serving.
16. This is admittedly somewhat close to the argument from the value of reasoning, above. Still, the emphasis here is somewhat different—on the agent's involvement per se, rather than the interference with the reasoning process itself.
17. Douglas (2013), though favorable towards moral self-enhancement, has recently offered support for a similar objection to third-party moral enhancement from a Kantian perspective: "Where A imposes a brute conformity enhancement on B, B's subsequent conduct might be thought to originate not in B's deliberation, but in A's, and this might be thought to detract from its moral worth" (9).
18. For a similar view, see Bublitz and Merkel (2009).
19. Though, notably, not all forms of compatibilism. See McKenna (2008) for a hard-line form of compatibilism that would not distinguish between cases (1) and (2) above.
20. These arguments speak against both novel biological moral enhancements as well as non-biological forms of widespread direct moral enhancement such as propaganda campaigns (insofar as they attempt to manipulate rather than persuade), and even some forms of dogmatic moral education. Still, as mentioned above, the way in which biological enhancements more directly interfere with mental (including reasoning) processes may make biological enhancements more problematic.
21. More avenues are, of course, possible. Not discussed here, for example, is the possibility of indirect emotional moral enhancement. The idea would be to tweak emotional processing in a way that reliably leads to more moral ideas, motives, and behavior. It is not immediately clear, however, how one would identify good and proper emotional functioning without specific reference to the morality of the output of that functioning. The risk with "indirect" emotional moral enhancement is that it in fact presupposes a certain normative framework and aims at bringing the emotions of enhancees in line with that framework, thus running afoul of the preceding concerns surrounding disagreement. The two methods discussed below, by contrast, have a relatively

compelling mechanism—considered judgments, which due to their structure can be generally trusted to at least tend towards the moral truth, just as they tend towards empirical truth.

22. Some reasons why we should trust considered judgments over more rudimentary inclinations will be adduced in the next section on akrasia.
23. Some of what is proposed here is compatible with third-party imposition of what Douglas (2013) has called brute conformity enhancements, in particular those brute enhancements that target reasoning and deliberative processes. So, a pharmaceutical intervention that improves reasoning capacities might be directly altering mental states, but it counts as an indirect enhancement insofar as it is not trying to get people to hold particular moral ideas, but help them think through those ideas more effectively.
24. One could even argue that similar means could be used to improve moral ideas. There may be a difference between some people's higher-order moral judgments and lower-order moral judgments. For similar reasons as in weakness of will cases, we may have reason to believe the higher-order moral judgments are more reliable and that therefore putting someone's lower-order moral judgments in line with higher-order judgments would count as an enhancement.
25. Some presently-available forms of reduction of weakness of will would be treatment of addiction and impulse control. These would count as indirect moral enhancements to the extent that they facilitate people acting on their all-things-considered judgments, which we can reasonably expect to be more morally reliable than the base addiction or impulse.
26. At a minimum, it would have to lead to moral improvement more often than leading to moral denigration. Whether a more significant margin of improvement (Three times out of five? Seven times out of nine?) is needed for real reliability will not be addressed here; it is indeed unclear how exactly one could determine the precise ratio needed.
27. That is not to say that limited uptake is necessary for self-enhancement to be acceptable. If there was great uptake, it would still avoid the objections adduced above.
28. It might appear that the indirect approach to moral enhancement does not avoid disagreement at all. Many moral debates are arguably based on irrationalities; with a more reasonable population, we can expect greater moral convergence in such areas. To be sure, this may lead to a reduction in disagreement; however, it does not threaten the *value* of disagreement, as analyzed in the present paper into the three dimensions of moral progress, reasoning, and individuality (each discussed below). This indicates that, while

disagreement is normally valuable, it is not essentially so—it is possible under certain circumstances (when it is solely the product of irrationality) to lose that value.

29. Even if this proposal does not necessarily inculcate the value of a certain form of reasoning, one might still worry that in practice it will lead to the promotion and promulgation of certain substantive ideas. Some ideas, after all, will be more coherent or unbiased than others. However, this should not be seen as a flaw of the approach; in order for any moral progress to be made, we should expect and hope for revision and rejection of some ideas. What makes the approach indirect (and preferable to a direct approach to moral enhancement) is that the content of such ideas are not presupposed.

REFERENCES

- Audi, Robert. 2005. *The Good In the Right*. Princeton: Princeton University Press.
- Bennett, Jonathan. 1974. "The Conscience of Huckleberry Finn." *Philosophy* 49 (188): 123–34.
- Bublitz, Jan Christoph, and Reinhard Merkel. 2009. "Autonomy and Authenticity of Enhanced Personality Traits." *Bioethics* 23 (6): 360–74.
- Douglas, Thomas. 2008. "Moral Enhancement." *Journal of Applied Philosophy* 25 (3): 228–45.
- . 2011. "Moral Enhancement via Direct Emotion Modulation: a Reply to John Harris." *Bioethics* 27 (3): 160–68.
- . 2013. "Enhancing Moral Conformity and Enhancing Moral Worth." *Neuroethics* 7 (1): 75–91.
- Elga, Adam. 2007. "Reflection and Disagreement." *Nous* 41 (3): 478–502.
- Fabre, Cécile. 2006. *Whose Body Is It Anyway? Justice and the Integrity of the Person*. Oxford: Oxford University Press.
- Harris, John. 2011. "Moral Enhancement and Freedom." *Bioethics* 25 (2): 102–11.
- Hills, Alison. 2009. "Moral Testimony and Moral Epistemology." *Ethics* 120 (1): 94–127.
- McKenna, Michael. 2008. "A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument." *Phenomenology and Philosophical Research* 77 (1): 142–59.
- Mill, John Stuart. 1999. *On Liberty*. Edited by Edward Alexander. Peterborough, Ontario: Broadview Press.
- Plato. 1997. *Plato: Apology of Socrates*. Translated by Michael C. Stokes. Warminster, England: Aris and Phillips.

- Persson, Ingmar, and Julian Savulescu. 2008. "The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity." *Journal of Applied Philosophy* 25 (3): 162–77.
- . 2010. "Moral Transhumanism." *Journal of Medicine and Philosophy* 35 (6): 656–69.
- Shaw, Elizabeth. 2014. "Direct Brain Interventions and Responsibility Enhancement." *Criminal Law and Philosophy* 8: 1–20.
- Shook, John R. 2012. "Neuroethics and the Possible Types of Moral Enhancement." *AJOB Neuroscience* 3 (4): 3–14.
- Simpson, Robert Mark. "Epistemic Peerhood and the Epistemology of Disagreement." *Philosophical Studies* 164 (2): 561–77.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Waldren, Matt Sensat. 2013. "Why Liberal Neutralists Should Accept Educational Neutrality." *Ethical Theory and Moral Practice* 16 (1): 71–83.
- Walker, Mark. 2009. "Enhancing Genetic Virtue: A Project for 21st Century Humanity?" *Politics and the Life Sciences* 28 (2): 27–47.