General discussion

Erasmus University Rotterdam

Psychopaths and God Machines: A deeply provocative and puzzling debate

The end of human civilization is nearing, as a consequence of a toxic mix of rapid technological developments, an outdated moral psychology, and an evil minority that is capable to cause ever greater – even *existential* – harm. Our only hope in stopping climate change, preventing terrorism, solving global injustices, in short, in averting disaster, is to find ways to change no less than the moral character of humanity. In fact, there is such an abundance of evil, and the urgency is so great, that the development and implementation of state initiated, mandatory bioenhancement programs is justified.[24] Quietly awaiting people to acknowledge and change their flawed natures themselves is simply not an option.

This assessment of the apocalyptic character of (parts of) the debate on moral enhancement might seem a bit of a caricature. Still, in writing this thesis, many a time I have marvelled at the unworldly nature, and at times even silliness, of the debate. To borrow the words of Harris Wiseman:

> The scope of moral bioenhancement is great: presented by some in terms of pure fantasy, by others as hard-hitting real future prospects and as offering remedies for every last moral concern from all our petty and mostly harmless vices and to various addictions and all sorts of "undesirable behaviours" up to the ultimate fate of humanity and moral bioenhancement's apparently salvatory promise for humanity against its own inner biological evils. (Wiseman, 2017, p. 398)

My main motivation in writing this thesis has been to understand in what ways the debate on moral enhancement can be relevant for and could impact on existing or emergent human practices – and the other way around. Some authors express scepticism towards the potential real world relevance of the debate and have posed the question whether "moral enhancement is ever to gain relevance apart from merely theoretical interest" (Beck, 2015, p. 234)? In this final chapter, contrary to such scepticism, I will take the claim that a range of present and emerging practices already contain ele-

---

24 In Savulescu and Persson's God Machine thought experiment, such a large scale project is described as follows: "The Great Moral Project was completed in 2045. This involved construction of the most powerful, self-learning, self-developing bioquantum computer ever constructed called the God Machine. The God Machine would monitor the thoughts, beliefs, desires and intentions of every human being. It was capable of modifying these within nanoseconds, without the conscious recognition by any human subjects." (Savulescu & Persson, 2012, pp. 412-413) See the introduction of this thesis for a more extensive discussion of The God Machine thought experiment.

Erasmus University Rotterdam

ments of "moral enhancement by proxy" (Wiseman, 2016, p. 219) as a starting point. I continue to understand moral enhancement as formulated in the working definition in the introduction, to include biomedical and non-biomedical interventions, and assume that in principle, both of these merit equally intense ethical scrutiny.

As technological possibilities (and perhaps public and political willingness) to influence and alter moral behaviour and moral capacities increase, the responsibility to make explicit this moral enhancement dimension in our current practices, and to think carefully about the associated ethical issues, increase as well.

Evaluating moral enhancement

As discussed before, this thesis aims at formulating conditions for ethically justifiable moral enhancement practices, and focuses to a lesser degree on analysing particular moral enhancement technologies. This represents a choice, based on the observation that at present there are very few concrete moral enhancement technologies available – or at least, that there is little consensus on the criteria to decide which technologies would in fact constitute moral enhancement technologies. The choice is also based on the observation that there is little agreement on what would be a suitable evaluative framework to analyse moral enhancement, and what kind of elements should be included in it.

Before formulating conditions for ethically justifiable moral enhancement practices, first some questions need to be addressed and some distinctions made.

Do we know or simply assume the alleged positive and negative effects of potential moral bioenhancement technologies?

Commentators in this debate have, to a lesser or larger extent, assumed moral bioenhancement to take certain forms, be focused on certain capacities, to have certain effects; and from those implicit or explicit assumptions, have taken up a position towards moral bioenhancement. For example: Harris argues that moral bio-enhancement interventions will necessarily impede on human freedom, because he assumes that moral enhancement will bypass or distort reasoning, and therefore interfere with ethical analysis rather than enhance it. Likewise, Reichlin claims that biotechnological means will have serious negative effects on moral agency and capacity of authentic moral behaviour: "biotechnological ways of producing moral progress (…) risk having serious negative effects on our moral agency, by causing a substantial loss of freedom and capacity of authentic moral behaviour, by affecting our moral identity and by

imposing a standard conception of moral personality" (Reichlin, 2017). These assumptions about the effects of emerging or fictitious moral enhancement interventions are not necessarily warranted.[25]

It has been argued, for example, that neuromodulation techniques can have disrupting, restorative, or enhancing effects on autonomy, depending on the context, the exact effects, and the person in question (Schermer, 2015). Moreoever, there are no guarantees that traditional procedures, like talk therapy, to name one example of a nonbiomedical intervention, will necessarily foster autonomy or other morally relevant capacities – it could do the exact opposite. Rather than differentiating ethically appropriate from ethically worrisome means, it is preferable to explore the conditions under which a particular intervention is most likely to undermine human freedom or human agency – and whether there are measures that would minimize these risks – and, vice versa, the conditions under which an intervention can restore or enhance them. The potential – positive or negative effects – of moral enhancement interventions should not be assumed, but be considered as largely *open questions,* which can only be answered in relation to a specified context or practice.

Distinguishing moral self-enhancement from moral other-enhancement

Can we imagine there ever to be a commercial market for moral enhancers, resembling current (semi-legal) markets for cognitive enhancers? Is it likely that there will be do-it-yourself communities experimenting with moral *self* enhancement, similar to current communities that are experimenting with neurostimulation for self-improvement purposes (Wexler, 2015, 2017)? Are people in fact interested in morally enhancing themselves, or do they think that morally enhancing *others* should be prioritized?[26] These are, to a certain extent, empirical questions with regard to which the jury is still out – as is true for many questions surrounding the debate on moral enhancement.

Everyday experience suggests that behaving morally, and being motivated to do so, is not out of the ordinary. People are prone to reflect normatively, behave altruistically, and sacrifice themselves for what they consider good causes. For her book *Strangers Drowning: Grappling with Impossible Idealism, Drastic Choices, and the Urge to Help*, Larissa

---

25 Horstkötter and colleagues caution against undue "bio-exceptionalism" and the apparent gap between the intensity of ethical scrutiny life sciences approaches (to antisocial and criminal behavior) receive as compared to psychosocial approaches (Horstkötter, Berghmans, & de Wert, 2014).

26 Persson and Savulescu argue that those most in need of moral enhancement, are least likely to pursue it: "If safe moral enhancements are ever developed, there are strong reasons to believe that their use should be obligatory, like education or fluoride in the water, since those who should take them are least likely to be inclined to use them" (Ingmar Persson & Savulescu, 2008, p. 174).

Erasmus University Rotterdam

MacFarquhar portrays a range of people who, in different ways, go at great lengths to do good (MacFarquhar, 2015). Being motivated to *better* oneself morally is also not uncommon. To give but one example, the so-called Effective Altruism movement describes the most effective ways of doing right for people who want to do better but do not know how (Singer, 2015). The Giving What We Can charity and the 80.000 Hours movement encourage people to donate at least 10% – and then gradually more – of their income, and to choose the most high-earning career (not the profession one has a passion for) in order to be able to give away as much as possible.[27]

Research shows that people consider moral traits to be central to their identity, and suggests that they consider them equally or even more important than other mental faculties, such as (emotional and autobiographical) memory, lower-level cognition, and perception (Strohminger & Nichols, 2014). At the same time, studies show that most people (irrationally) believe themselves to be morally superior to the average person:

> Most people strongly believe they are just, virtuous, and moral; yet regard the average person as distinctly less so. (…) virtually all individuals irrationally inflated their moral qualities, and the absolute and relative magnitude of this irrationality was greater than that in the other domains of positive self-evaluation. (…) Taken together, these findings suggest that moral superiority is a uniquely strong and prevalent form of "positive illusion." (Tappin & McKay, 2016, p. 623)

If people care about their moral qualities, yet tend to overestimate them, this raises the question whether and to what extent people indeed consider themselves to be a candidate for moral enhancement. A study by Jason Riis and colleagues confirms that people consider morally relevant traits such as empathy and kindness fundamental to their self, but are least willing to pharmacologically modify these morally relevant traits compared to other traits they considered less morally relevant, such as wakefulness and reflexes, possibly precisely because these traits are so closely related to personal identity (Riis et al., 2008).

Chapter 6 of this thesis suggests that in general people are open to morally improving themselves (in terms of increasing their empathy for others), but that they care about (and change their preferences based on) the process, or means by which this change in their moral capacities comes about (Specker et al., 2017). Participants' willingness

---

27 See the website of the Oxford based Centre for Effective Altruism: https://www.centreforeffectivealtruism.org/; the website of Giving What We Can: https://www.givingwhatwecan.org/; and the 80.000 Hours Organization: https://80000hours.org/.

to participate in an empathy-enhancing program themselves was greater for the non-pharmacological as compared to the pharmacological program. Taken together, these findings suggest that people are open to morally improving themselves, but that widespread willingness to participate in moral self enhancement by means of biomedical interventions is lacking.

Arguably, the distinction between moral self enhancement and moral other enhancement is relevant when ethically evaluating (potential) moral enhancement practices (see chapter 3 and 6). As we have seen, a central and recurring question in the moral enhancement debate is what the need for and objectives of developing and implementing moral enhancement interventions are. These needs and objectives are often described in terms of potential advantages to others (reduction of harm) or society (significant overall decrease in criminal, violent, or otherwise antisocial behaviour).[28] Indeed, in his first article on moral enhancement, Douglas argued that because moral enhancement benefits others, it compares favourably to other forms of enhancement:

> Unlike the most frequently mentioned varieties of enhancement, enhancements satisfying this formula for moral enhancement could not easily be criticised on the ground that their use by some would disadvantage others. On any plausible moral theory, a person's having morally better motives will tend to be to the advantage of others. (…) One could not object to moral enhancement on the ground that it would systematically impose morally gratuitous disadvantage on others. (Douglas, 2008, p. 230)

With respect to moral self enhancement, the decision to pursue moral enhancement and to determine the goal and appropriate means of doing so, is to a large degree up to the person herself, provided a number of safeguards are put in place. It is the responsibility of those offering potential 'moral enhancement' interventions, programs, and courses, to safeguard that what they offer is safe and effective, and to provide ample opportunity for informed consent.[29] Doing so yields many challenges in and of itself, comparable to the challenges faced by anyone offering counselling, training, or therapy in the medical or semi-medical domain to people who come to them with a

---

28 DeGrazia provides the following rationale for why moral bioenhancement needs serious consideration: "because the status quo of moral behaviour is deeply problematic and traditional means of moral enhancement may prove inadequate to achieve needed improvements—notwithstanding the phenomenon of moral progress. The status quo is deeply problematic because there is such an abundance of immoral behaviour, with devastating consequences, and serious risk of worse to come." (DeGrazia, 2014, p. 362)

29 In principle, this is equally important for biomedical and nonbiomedical interventions and programs.

Erasmus University Rotterdam

request for help. It may also be comparable to the challenges faced by regulators and scientists who are trying to design sensible regulations and robust safety standards for direct-to-consumer personal genome testing (Bunnik et al., 2014) or home use of neurostimulation (Fitz & Reiner, 2013). For example, ethicists have called on scientists and journalists to balance their enthusiasm about the promises of new brain stimulation technologies with restraint, and to provide realistic information about potential harmful effects as well (Fitz & Reiner, 2013, p. 411).

Clearly, in the case of moral enhancement of *others*, there are additional and much larger responsibilities to justify the need, objectives, and procedure and the means employed, as well as the balancing of potential benefits and harms. This is true especially in the case of interventions that are initiated, incentivized, or mandated by a state or state body.[30] If we take proponents of moral enhancement at their word, at least some of their proposals would involve some sort of moral enhancement *program*. Mark Walker's proposal for a Genetic Virtue Program "to reduce evil in our world" by "engineering genetic virtue" might serve as an example (Walker, 2009, 2010).

Distinguishing between different target populations

As was discussed in chapters 3 and 4, in order to determine a suitable target population for moral enhancement programs *for others*, for example in terms of who will benefit most from a specific intervention, or of which individual or groups need to be targeted to optimize societal benefit, some form of screening will be necessary (Specker & Schermer, 2017). Moreoever, specific concerns are associated with implementing moral enhancement interventions in particularly vulnerable target gourps.

In children, there might be unique (but oftentimes unknown) effects on the developing brain (Cohen Kadosh et al., 2016)[31], challenges surrounding decision-making processes and consent (Focquaert, 2013; Maslen et al., 2014), and potential effects on a child's future (and developing) autonomy. Concerns about negative effects of screening and selection apply to educational contexts, with regard to programs aimed at lowering or preventing a particular kind of problem behaviour (e.g. the example of bullying in chapter 6). One potential risk is that children or youngsters, who are at risk and in need of help, but exhibit internalizing rather than externalizing problem behaviour,

---

30 Sarah Carter discusses ethical issues surrounding incentivizing programs for moral enhancement (S. Carter, 2015). In addition, considerable attention has been devoted to the distinction between voluntary and mandatory moral enhancement (Rakić, 2014b, 2014a; S. Carter, 2015; Baccarini & Malatesti, 2017; Rakić, 2017).

31 Again, in principle, this applies to potential effects on the developing brain of both biomedical and nonbiomedical interventions.

risk staying out of sight. In these contexts especially, the best interests of all children should be of concern of educators, social workers, and policy makers, and an exclusive focus on 'troublemakers' is not justified (Kaltiala-Heino & Eronen, 2015).

The group of forensic psychiatric patients is also vulnerable,[32] for example in terms of concerns over coersive measures (Nedopil, 2016) or so-called coercive offers, and the question whether people in forensic psychiatric care contexts can ever truly give their informed consent (McMillan, 2013, 2014; Ryberg, 2015; Adshead & Davies, 2016) (see chapters 3-5 for more extensive discussion of the specific challenges associated with the forensic field).

Conditions for ethically justifiable moral enhancement practices

Proponents of enhancement in previous debates have often defended and shown a strong, even libertarian, commitment to individual liberty and freedom of choice. However, in the debate on moral enhancement, a number of central commentators appear to depart from this commitment, when they advocate for mandatory, population-wide moral enhancement programs, aimed at solving societal problems like climate change, war, and social evil by changing individual biology.[33]

Those who advocate population-wide, mandatory moral enhancement programs would have to convincingly argue that the need for such large-scale programs is so great, that abandoning central and quite basic principles of liberal democracy is warranted.[34] In addition, they would have to show that their proposed solution – correcting individual moral deficiencies – in fact meets that urgent need adequately and effectively.

In addition, a range of regrettable historical as well as current examples caution against putting the full force of the state behind moral enhancement programs aiming to change a person in such a way, as to make her conform to (or refrain from) what in the eyes of those that order it, is (un)wanted or (im)moral. An often discussed example

---

32 "In the emotionally (and politically) charged context of crime, the imposition of unproven technical "fixes" on the always unpopular class of "criminals" seems quite plausible" (Greely, 2007, p. 1129).

33 Stefan Schlag argues that Persson and Savulescu misunderstand the nature of collective action problems: "The argumentative difficulties possibly arise because the authors misunderstand the basic problem of the tragedy of the commons. Social dilemmas are not caused by individual moral deficiencies but are rooted in the problematic structure of human interaction. Neuroscientific interventions into the biological basis of human behaviour are fundamentally inappropriate means to solve problems of this type" (Schlag, 2016, pp. 11-12).

34 One such a fundamental liberty is the right to privacy – which Persson and Savulescu recognize as a legal but not a moral right, and for that reason should not stand in the way of implimenting moral enhancement programs (Ingmar Persson & Savulescu, 2012; but see: Bublitz, 2016).

concerns homosexuality, which has long been criminalized (and still is in large parts of the world), and which has been classified by the American Psychiatric Association as a mental disorder until 1973 and as a sexual orientation disturbance until 1987 (American Psychological Association Task Force on Appropriate Therapeutic Responses to Sexual Orientation, 2009). Various sexual orientation change efforts are still applied today.

These kinds of examples show the importance of freedom and diversity of thought, civil disobedience, and 'praise for defiance' (Arnold, 2015; Ripley, 2016). Instead of imposing a particular comprehensive moral doctrine, a fundamental recognition of the worth of value pluralism and of moral disagreement and debate should be at the forefront of any moral enhancement program. From this recognition it follows that in line with the views from moral progressivists (chapter 3), an important 'guiding concern' in implementing potential moral enhancement programs is to focus on enhancing morally relevant capacities, not on moral conformity.

Moral conformity could be described as the conformity of human behaviour with social or moral norms (Civai & Ma, 2017). Instead of focusing on changing a person in such a way as to make her conform to social or moral norms, arguably a more promising approach is to focus on what it means to 'be better at being good', and what kind of capacities are needed to become 'better at being good.' Such a capacitarian and developmental approach focuses on the set of capacities that "mature moral functioning"[35] (Narvaez, 2010) or "moral intelligence"[36] (Tanner & Christen, 2014) presuppose, and asks what kind of environments would foster these capacities.

To conclude, a focus on (present and emerging) moral enhancement practices invites reflection on a range of questions that are largely missing from the current debate on moral enhancement: What goal/ purpose does the intervention serve? What political, sociological, institutional context are we dealing with? Who will likely benefit and in what way(s)? Who will risk harm, and what kind of harm? What kind of background theory of moral development and moral growth is referred to? Is the intervention justified in terms of benefits to the individual undergoing the intervention, or in terms of benefits to relevant others/ public safety/ common good – or both? Are all those

---

35 In short, mature moral functioning refers to "individual capacities for habituated empathic concern and moral metacognition—moral locus of control, moral self-regulation, and moral self-reflection—comprise mature moral functioning, which also requires collective capacities for moral dialogue and moral institutions" (Narvaez, 2010, p. 163).

36 Tanner and Christen define moral intelligence as "the agent's capacity to process and manage moral problems" (Tanner & Christen, 2014, p. 120).

involved aware of the attempt to morally enhance, especially the person undergoing the intervention; has the person given her *informed consent*? Do all those involved agree with the goal and the means, especially the person undergoing the intervention? Will the person undergoing the intervention as a result of the intervention most likely be better at being good, or will he more reliably do the good thing?

The reader might find these questions slightly banal and deceptively simple. Yet, in order to focus attention on the fundamental issues concerning individual freedom and right to self-realisation that are at stake in current or emerging moral enhancement practices, these are, I think, exactly the kinds of questions that need to be asked.

Strengths and limitations

As both the debate and the implications of the science of morality are far from settled, it is also far from clear what the central moral issues are. In imagining potential domains in which elements of moral enhancement are recognizable, the main goal of this thesis has been to identify ethical issues that are not central in the debate now.

In response to an article in which we suggested a number of domains that we think are relevant for the debate on moral enhancement (chapter 3), bioethicist Bert Gordijn has formulated a range of reservations with respect to our suggested set of practices (Gordijn, 2017). His main concern is that the success of our strategy "is predicated on the correctness of suppositions about future moral bioenhancement scenarios. Unfortunately, however, these assumptions are inherently speculative. It is fundamentally problematic to determine in which contexts moral enhancements might first be implemented. It is equally challenging to identify contemporary domains that might, in their central aspects, be comparable to future moral bioenhancement practices" (Gordijn, 2017, p. 427).

Gordijn is correct that we cannot know for sure whether the domains we have identified will indeed be relevant for future moral enhancement scenarios. Our choice of practices might indeed turn out to be irrelevant for future moral enhancement scenarios. However, Gordijn does not go into the question why our particular choice of practices is wrong, or unlikely to be relevant for moral enhancement in the future. We have argued that elements of moral enhancement/improvement are present within current forensic practices, depending on how one defines moral enhancement – Gordijn has not refuted that. Moreover, our central claim and position that moral enhancement

Erasmus University Rotterdam

should be analysed *contextually* and in relation to particular practices, still stands, even if our choice of practices turns out to be mistaken.

Without any doubt a range of other practices might be relevant and worth considering further. "Impartiality" moral enhancement, for example, could be offered to those who are in a position of power or fulfil an influential role, with the aim of nullifying implicit biases.[37] Douglas discusses studies of biological influences on fairness related behaviour (Douglas, 2015, p. 31). One often-cited study shows that judges' parole decisions vary dependent on extraneous variables, like for example how long ago the judge had a (food) break, letting the authors conclude that "justice is what the judge ate for breakfast" (Danziger et al., 2011). Neutralizing these extraneous influences would arguably make judicial sentencing fairer. Whether judges would in fact voluntary seek such impartiality enhancing interventions, or whether some sort of persuasion or coercion would be necessary (and justified) merits further discussion. Moreover, one would still have to determine who would be the fairer judge: the hungry of the satisfied one?

An additional potential concern is the fact that the working definition of moral enhancement as it was formulated in the introduction is very broad and remains silent on a number of arguably important distinctions. As the debate itself was the main subject of research, I did not want to exclude potential perspectives and practices in advance. However, such a broad interpretation has disadvantages as well. For example, a consequence of analyzing a diversity of practices through a moral enhancement lens might be that conceptualizing these practices as 'moral enhancement' practices could unintentionally promote their acceptance (chapter 5).

Concluding reflections and suggestions for further research

To enhance the debate, in my opinion it would be necessary to direct further research towards the following themes and questions:

Medicalization of immoral behaviour

Approaching violence (D. J. Williams & Donnelly, 2014) or other examples of arguably immoral behaviour from a health instead of a criminal justice lens, is an example of medicalization: quite literally, 'to make medical.' Medicalization is understood as the process by which previously nonmedical problems become defined and treated

---

37 One could think of judges, business and political leaders, parents, police and military personnel, programmers, American presidents, etc.

as medical problems, usually as diseases or disorders (Conrad, 2013, p. 196). Another example concerns the question whether addiction should be considered an illness or a moral failing (Wiseman, 2016). Although processes of medicalization often elicit a negative connotation the concept is, in principle, value neutral. There are both (sociological) studies *describing* processes of medicalization as well as critical analyses of instances of over- or under medicalization, and merely coining the term does not in itself differentiate between good and bad forms of medicalization (Conrad, 2013, pp. 1199-1200; Parens, 2013, p. 28; Horstkötter et al., 2015).

In the case of medicalization of violence, problematic, violent behaviour is defined, or redefined, as a medical problem (as a psychiatric disorder for instance). Medicalization of deviant or violent behaviour can have positive and negative effects. On the one hand, it can open opportunities for adequate care and treatment, where these were previously lacking. Instead of responding to this kind of behaviour with indignation and punishment, the objective of treatment is to address underlying illness or disorder and to limit chances of reoffending.

A possible worry is that by viewing (potential) offenders foremost from a medical angle, they are more likely to be viewed as basically incompetent, or at the mercy of forces beyond themselves:

> When we argue, say, against the medicalization of badness – e.g., against treating criminal behaviour as the symptom of a psychiatric disorder – we are arguing against the view of ourselves as objects at the mercy of forces beyond ourselves, and for the view of ourselves as subjects who can choose. (Parens, 2013, p. 29)

Another concern with respect to health approaches to violent behaviour concerns individualization: the risk that causes are (quite literally) sought within the individual, as a genetic, neurological, biological disorder, and to a far lesser extent in social, institutional, familiar, or other environmental circumstances. Further conceptual and ethical research is necessary here to consider the justification, and the benefits and risks of medicalization of 'badness'. Such research should ideally be performed by psychological and psychiatric forensic scientists and professionals, in close collaboration with ethicists, anthropologists, and sociologists.[38]

––––––––––––––––

38   The perspectives of offenders themselves should also be included. Jeremy Dixon (Dixon, 2018) (Glover, 2014)

Erasmus University Rotterdam

The roles and responsibilities of medical doctors. What should doctors do?

The debate on moral enhancement has proceeded without much attention for the kinds of institutional contexts in which moral enhancement would be implemented – or is already implemented – and the implications of those contexts in terms of professional responsibilities of those who would be responsible for the implementation.

On the one hand, perhaps biomedically trained professionals ('clinical technicians') could administer and monitor the technical aspects of certain interventions. At the same time however, focussing on the technical aspects exclusively, might come at the expense of attention for more comprehensive care aspects of the interaction. Perhaps a new sub discipline would emerge, in which hybrid doctors would move in between the domains of medicine (neuroscience, psychiatry), public health (prevention of health-related risks, safety, monitoring), and public policy (safety, crime prevention). Such a professional would be well-established in a diversity of domains, and would ideally be able to do justice to the specific responsibilities flowing from these diverse domains. There is a risk, however, that these diverse domains yield dual, multiple roles, and diffuse responsibilities (chapter 3 and 5). The merging of care and safety responsibilities warrants more fundamental reflection on the roles and responsibilities of medical professionals, as well as on the limits of those roles. This is true for forensic professionals in particular. Yet, forensic medical practice has thus far not received as much attention from medical ethics as would be warranted by the moral complexities of this practice.

Changing the individual, and/or changing the environment? Parallels with other debates

In order to stimulate people to act morally, or in accordance with social norms, a diversity of strategies can be employed. The debate on moral enhancement focuses on strategies that are aimed at changing the person herself (Klincewicz, 2016). Glenn Cohen for example tentatively considers the option of changing the biological makeup of human beings in such a way as to reduce serious human rights violations. He argues that "instead of merely crafting laws and setting up structures that get human beings such as they are to respect human rights, that the human rights approach should also consider embracing attempts to remake human beings (and more specifically human brains) into the kinds of things that are more respectful of human rights law" (Cohen, 2015, p. 1). In various other debates a variety of strategies that focus on changing the environment are discussed.[39]

---

39 These strategies could be used together, in concert, as is often done, for example in public health, where interventions are targeted at both the individual and the environment.

There have been proposals to change the so-called 'choice architecture' of the environment in order to 'nudg' people towards pro-social instead of egoistic choices (Capraro et al., 2017)[40], for instance by designing robots to serve as "moral nudgers" in order to foster "socially just" tendencies in humans (Borenstein & Arkin, 2016). Ismaili M'hamdi and colleagues have coined the concept of "other-regarding" nudges, in other to describe cases where "the principal but not necessarily sole beneficiary of the nudge is not the nudgee," but someone else (Ismaili M'hamdi et al., 2017, p. 702). Another example is the field of robot ethics, debating attempts to design "morally competent robots" (Malle, 2016).[41] Comparable to discussions about to who responsibility should be assigned to when a self-driving car ends up in an accident – the driver/passenger or the driver/car – one can ask who is responsible when a morally enhanced person goes astray. Who is responsible when the God Machine derails?

Another example concerns speculations about (and first experiments with) using so-called wearables to provide persons with insight as to biological processes underlying their activities and behaviour. The Empatica wristband for example supposedly can provide biofeedback about actual aggression and stress levels. This information can be used for self-regulation purposes, but could also be used for monitoring and surveillance, for example in secure prison contexts (de Kogel & Cornet, 2016; van Hintum, 2018). Such proposals should be reflected on in relation to existing efforts to monitor behaviour. To give one example, the Chinese government is experimenting with a social credit scheme through which citizens can earn (and lose) points and accompanying privileges based on online and offline behaviour (Creemers, 2017).[42] Data sources that are included in the current experiments include online behaviour, financial information, and behaviour in traffic and in public transport gathered by face-recognition

---

40 Nudges has been defined as "approaches that steer people in particular directions" (Sunstein, 2017, p. 4) while preserving freedom of choice.

41 In thinking about what a morally competent robot would look like, this field of robot ethics proves an interesting testing ground for thinking about the elements that are needed for "full" human moral agency.

42 "an ambitious proposed social credit scheme is intended to create a range of benefits and sanctions for online and offline behaviour. According to State Council plans, it is intended that social credit information will be connected with individuals' identity card numbers, creating unique and traceable files that can be used to facilitate citizens' access to financial and government services. At the same time, the plan called for the introduction of real-name identity-based appraisal and scoring of individual online comportment, as well as of blacklists for those perpetrating various kinds of fraud, deception and 'harm to others' lawful rights and interests'. In other words, it is not unlikely that undesired behaviour online may affect citizens' ability to gain a livelihood, find schools for their children or take out insurance." (Creemers, 2017, p. 97) See this webpage for a game that simulates the workings of the social credit scheme: https://app.nos.nl/social-credit-score/index.html.

cameras (Nieuwsuur, 2018). It is not unthinkable that future social credit schemes would include biomedical data as well.[43]

The common denominator in these discussions is reflection on a broadening range of technologies that are or can potentially be used to make people behave in accordance with social norms. In light of these parallel debates, one important direction for future research would be a more comprehensive ethical analysis of the diversity or "omnipresence"[44] of ways in which societal actors can stimulate, incentivize, nudge, or force persons towards certain behaviour, and the accompanying ethical concerns. All these emerging practices require moral scrutiny, and many of the relevant considerations will be common to all these debates. The 'moral enhancement debate' can function as a useful resource here.

Challenges and the relevance of the new sciences of morality

I find the findings of the new sciences of morality fascinating, in particular the ways this new and rich landscape of empirical studies on animal and human morality can challenge and enrich both philosophical and folk understandings of morality. These disciplines presuppose, paint, promote, and also undermine certain ideas of human nature. The debate on moral bioenhancement in a way represents an extreme example of this challenge: Is human morality to be understood on individual/biological level, and should neuro/bio-insight be taken to imply that our self-understanding needs an 'upgrade', and if so, in what ways?

It is notable that images of man and of human nature as they are usually being painted by proponents (e.g. transhumanist) of human enhancement are generally rather positive. They tend to focus on fundamental human capacities for imagination, self-creation, and empowerment. Proponents of moral enhancement on the other hand tend to depart from this positive depiction of human nature, and stress the ways human moral nature is fundamentally defective and for that reason needs a (preferably quick) 'fix'. The truth no doubt lies somewhere in the middle.[45]

---

43 See (Singh & Sinnott-Armstrong, 2014).

44 Such a comprehensive analysis is comparable to the argument that the ethical evaluation of health checks should go beyond weighing the harms and benefits of individual tests, but should take into account the potential beneficial and harmful effects of what has been termed an "omnipresence of health checks" (Stol et al., 2017), that is, the combined effects resulting from the multitude of tests on offer.

45 Are human beings inherently and thoroughly selfish, with morality only being an after-thought or thin, cultural layer of "veneer" around a corrupt core; or are human capacities for morality antique and part of our genetic inheritance (de Waal, 2009, pp. 7-12)?

We should do justice to human capacities for giving and exchanging reasons for actions (Sie, 2014), holding and giving responsibility, while at the same time "biting the bullet" by acknowledging the diversity of challenges (Levy, 2006). Despite a range of challenges, based on fascinating studies explaining the varieties of ways human moral behaviour, moral thought, and moral emotions are influenced by situationalist, environmental, genetic, and biological influences, we should be careful to not rush towards – debunking – conclusions.[46] Humans arguably have evolved into the kind of beings to whom 'moralizing' comes natural; we tend to evaluate ourselves and others in moral terms. As Harris writes; "We have certainly evolved to have a vigorous sense of justice and right, that is, with a virtuous sense of morality" (Harris, 2011, p. 104). To reconcile these different ways of thinking about human morality is a challenge for science and philosophy: for the explanatory models and their underlying presuppositions. But it is a challenge also for folk conceptions of the moral self/ moral identity. Along with providing new input or challenges to various conceptual and scientific *puzzles*, studies on human morality arguably can impact the way we view ourselves as moral agents as well.

To the extent that 'the new sciences of morality' indeed open up 'the black box' of human moral psychology, the question is how we want to use this new, evolving body of knowledge. If we truly are coming closer to understanding the many influences on human moral decision making and behaviour, the question is how this knowledge is interpreted and translated. What in my view is paradoxical is that in many discussions we quickly move in the direction of applications where the insight is not distributed more widely/ democratically, but is apparently limited to an unidentified 'we', a small group (scientists, moral philosophers, bioethicists), who discuss how that new knowledge can be used to change 'people'. Many of the applications that are discussed (nudging, using robots, environmental intervention, virtual reality, etc.) seem to be conceived, designed from 'above', with little consideration of the possible ways in which people themselves could (or would like to) benefit from this knowledge. Future research should focus on ways in which the new sciences of morality could indeed empower people to become better at being good.

---

46 Why is it that the realization that we typically/ often 'bypass reasons' often perceived as disturbing? Because, philosopher Regina Rini argues; "Doxastic embarrassment results from my awareness of a gap between the considerations that seem correct to me in my conscious thought, and the factors that actually drive my automated moral beliefs. What psychological research exposes is a form of disunity in my functioning as a moral agent. My conscious, reflective self is not appropriately unified with my automated, effective self" (Rini, 2016, p. 1449).