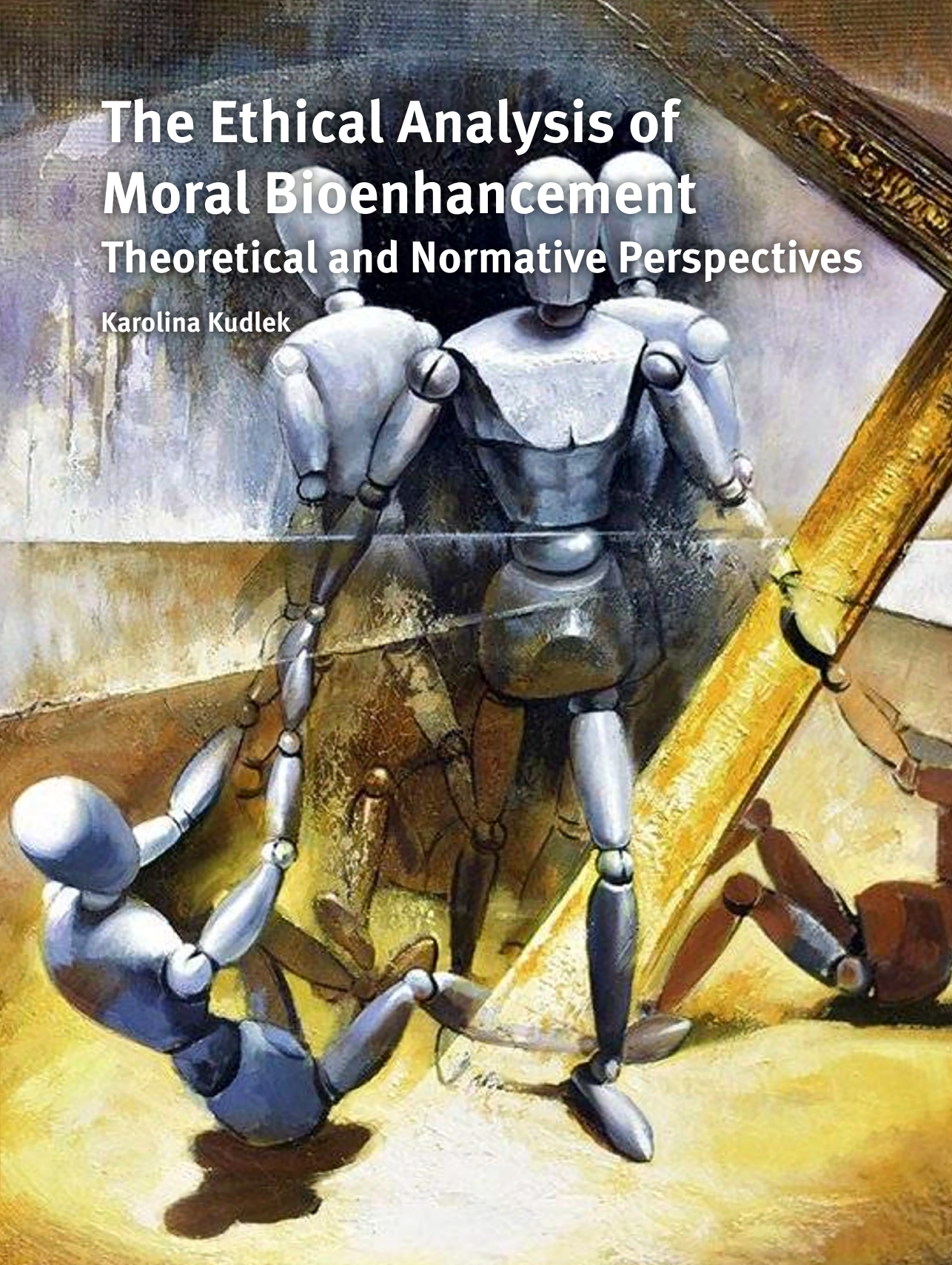


# The Ethical Analysis of Moral Bioenhancement Theoretical and Normative Perspectives

Karolina Kudlek



Simon Stevin Series in the Ethics of Technology



# **The Ethical Analysis of Moral Bioenhancement**

**Theoretical and Normative Perspectives**

Karolina Kudlek



# THE ETHICAL ANALYSIS OF MORAL BIOENHANCEMENT

## THEORETICAL AND NORMATIVE PERSPECTIVES

### DISSERTATION

to obtain  
the degree of doctor at the Universiteit Twente,  
on the authority of the rector magnificus,  
prof. dr. ir. A. Veldkamp,  
on account of the decision of the Doctorate Board  
to be publicly defended  
on Wednesday 15 June 2022 at 12.45 hours

by  
**Karolina Kudlek**

born on the 4th of July, 1986  
in Sisak, Croatia

This dissertation has been approved by:

Supervisor

prof. dr. ir. P.A.E. Brey

Co-supervisor

dr. M. Sušnik

© 2022 Karolina Kudlek, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

ISBN: 978-90-365-5393-3

ISSN: 1574-941X

DOI: 10.3990/1.9789036553933

Cover image: Aleksandros Karametas, *Morality* (2015)

Illustrations: Aleksandros Karametas

Copies of this publication may be ordered from the 4TU.Centre for Ethics and Technology, [info@ethicsandtechnology.eu](mailto:info@ethicsandtechnology.eu)

For more information, see <http://www.ethicsandtechnology.eu>

**Graduation Committee:**

Chair / secretary: prof. dr. T. Bondarouk

Supervisor: prof. dr. ir. P.A.E. Brey  
Universiteit Twente, BMS, Philosophy

Co-supervisor: dr. M. Sušnik  
Institute of Philosophy Zagreb

Committee Members: prof. dr. C. Aydin  
Universiteit Twente, BMS, Philosophy

dr. M.H. Nagenborg  
Universiteit Twente, BMS, Philosophy

prof. dr. J. Savulescu  
University of Oxford

prof. dr. M. Hauskeller  
University of Liverpool

prof. dr. B. Beck  
TU Berlin

This work was funded by the Croatian Science Foundation as part of the research project *Harm, intentions and responsibility* (HIRE) [Grant no. HRZZ-UIP-2017-05-4308], hosted by the Institute of Philosophy, Zagreb, Croatia.



# Contents

|   |    |
|---|----|
| Introduction  | I  |
| 1. What Is Moral (Bio)Enhancement and Why Do We Need It?                    | 4  |
| 2. Difficulties with MBE and the Enhancement Debate                         | 7  |
| 3. The Research Questions   | 8  |
| 4. Methodological Approach  | 11 |
| 5. Chapter Outline  | 13 |
| Chapter 1. Challenges in the Human Enhancement Debate:<br>A Critical Review | 19 |
| 1. Introduction   | 21 |
| 2. Conceptual and Taxonomical Difficulties                                  | 22 |
| 2.1. Traditional and Biomedical Aspects                                     | 24 |
| 2.2. Enhancing Well-Being   | 26 |
| 2.3. Enhancement-Treatment Distinction                                      | 27 |
| 3. Ethical Concerns   | 29 |
| 3.1. Concern: Changes in Human Biology, Genes, and Nature                   | 30 |
| 3.2. Concern: Playing God and the Quest for Perfection                      | 32 |
| 3.3. Concern: Exacerbating Social Issues of Equality and Justice            | 33 |
| 3.4. Concern: Loss of Spontaneity and the Mystery of Life                   | 35 |
| 4. Moving beyond Pro et Contra Exchange                                     | 37 |
| 5. Towards a Systematic Assessment of HET                                   | 39 |
| 6. Conclusion   | 43 |
| Chapter 2. Towards a Systematic Evaluation of Moral Bioenhancement          | 47 |
| 1. Introduction   | 49 |
| 2. The Need for Moral Bioenhancement  | 50 |
| 2.1. The Imperative for Moral Enhancement                                   | 50 |
| 2.2. The Permissibility of Moral Enhancement                                | 52 |
| 2.3. The Prospects of Moral Enhancement                                     | 53 |
| 3. Unfit for the Future   | 55 |
| 4. Concerns Regarding Moral Bioenhancement                                  | 57 |
| 4.1. Conceptuality and Feasibility Concerns                                 | 58 |
| 4.2. A Threat to Moral Values   | 60 |
| 4.3. Socio-Political Concerns   | 62 |
| 5. Response to Concerns: Moral Neuroenhancement?                            | 64 |
| 6. A Methodological Framework for the Assessment of Moral Enhancement(s)    | 67 |
| 6.1. Ethical and Theoretical Analysis                                       | 68 |
| 6.2. Application of the Criteria  | 70 |
| 7. Conclusion   | 72 |

|  |     |
|--|-----|
| Chapter 3. Is Human Enhancement Intrinsically Bad?                                   | 75  |
| 1. Introduction  | 77  |
| 2. Intrinsic Impermissibility Thesis, Intrinsic Value, and Enhancement               | 79  |
| 2.1. Conceptual Clarifications   | 80  |
| 2.2. Preliminary Argument  | 83  |
| 3. Concerns about the Intrinsic Badness of HET                                       | 85  |
| 3.1. Concern 1: The Unnaturalness of HET   | 85  |
| 3.2. Concern 2: Pursuing HET Is an Intrinsically Bad Disposition                     | 88  |
| 3.3. Concern 3: HET's Necessary Consequences   | 91  |
| 4. The Gap between Intrinsic Badness and Moral Permissibility?                       | 94  |
| 5. Conclusion  | 96  |
| Chapter 4. The Role of Emotion Modulation in the Moral<br>Bioenhancement Debate      | 99  |
| 1. Introduction  | 101 |
| 2. The Moral Bioenhancement Proposal   | 103 |
| 3. The Evolutionary and Moral Psychology of Emotions                                 | 107 |
| 4. Emotion Modulation and Its Problems   | 111 |
| 4.1. Emotion-Mechanism Correlation   | 112 |
| 4.2. Moral Emotions-Moral Reasoning Correlation                                      | 114 |
| 5. The Outcomes of Emotion Modulation  | 115 |
| 6. Conclusion  | 118 |
| Chapter 5. On the Uneasy Alliance between Moral Bioenhancement<br>and Utilitarianism | 121 |
| 1. Introduction  | 123 |
| 2. The Utilitarian Case for MBE  | 125 |
| 2.1. Making Better Utilitarian Agents?   | 125 |
| 2.2. Is MBE Optimific?   | 128 |
| 3. The Utilitarian Case against MBE  | 131 |
| 3.1. The Tension between Utilitarianism and Commonsense Morality                     | 131 |
| 3.2. The Problem of Partial Altruism   | 133 |
| 3.3. Prioritizing Just over Utility-Maximizing Actions                               | 135 |
| 4. Conclusion  | 137 |
| Chapter 6. The Kantian Promise and Peril of Moral Bioenhancement                     | 141 |
| 1. Introduction  | 143 |
| 2. MBE as a Generalized Skepticism towards Our Moral Agency                          | 145 |
| 3. Self-Paternalism and the Kantian Objection to MBE                                 | 149 |
| 4. MBE and Kantian Moral Psychology  | 153 |
| 5. The Kantian Promise and Peril of MBE  | 158 |
| 6. Conclusion  | 163 |

|   |     |
|---|-----|
| Conclusion  | 167 |
| 1. Contributions to the Debate                      | 169 |
| 1.1. Critical Review and Methodological Guidelines  | 170 |
| 1.2. Theoretical and Conceptual Analysis            | 171 |
| 1.3. Normative Ethical Evaluation                   | 173 |
| 2. Strengths, Limitations, and Further Implications | 176 |
| 2.1. Methodological Aspects                         | 176 |
| 2.2. Theoretical and Conceptual Aspects             | 178 |
| 2.3. Normative Aspects                              | 179 |
| 3. Socio-Political Ramifications                    | 180 |
| References  | 183 |
| Summary   | 197 |
| Samenvatting  | 199 |
| Sažetak   | 201 |
| Acknowledgments                                     | 203 |
| About the Author                                    | 209 |
| Simon Stevin (1548-1620)                            | 213 |



“We will never have a perfect world, but it’s not romantic or naïve to work toward a better one.”

— Steven Pinker



A. Karametas, *Unemployment* (2015)

# Introduction





# Introduction

Deficits in moral character can lead to some of the most dreadful atrocities in the world. These deficits come in various forms; from weakness of will, lack of altruism or empathy, to racial aversion and violent aggression. Since these traits are to a significant extent biologically grounded, the question is: Could we develop biomedical and biotechnological interventions to modify our biological traits? And if we could, *should* we use them to alter our moral character? These questions are at the heart of the debate on moral bioenhancement (MBE) – an emerging technology aimed at modifying capacities and traits relevant for moral deliberation and behavior. The debate is largely hypothetical and scattered over various conceptual, ethical, and practical difficulties raised by the emergence of human enhancement technologies (HET) such as MBE.

This dissertation is an attempt to shed some new light on the MBE debate. It offers an ethical analysis of several theoretical and normative challenges closely related to the moral permissibility of MBE. These challenges are addressed over six chapters, which are also stand-alone articles, and they, taken jointly, are an attempt at proposing a much-needed dialogue between ethical theory and bioethics that could give us insight into the moral status of MBE technologies, provide methodological guidelines, and help us clear the path for future research.

The motivation to engage in this intriguing and rapidly developing bioethical topic lies in the potential of enhancement technologies to improve the physical, cognitive and emotional capacities of humans – but also to pose substantial risks to valuable aspects of human life, such as freedom, autonomy, and self-growth. Accordingly, many scholars have taken part in this discussion advocating various approaches, from a radical ban to a moral obligation to enhance, and while there is much literature on these topics, uncertainties and disagreements remain. To this day, the state of the debate has not significantly advanced towards any consensus regarding the moral permissibility of human enhancements. Extreme polarization and unsystematic assessment criteria are important causes of lack of progress in this area.

This introduction provides context and an overarching narrative for the coming chapters in five main steps. First, I briefly introduce the thesis' research subject, the hotly debated emerging technology known as MBE. Second, I place MBE

within a broader ethical debate on HET and identify some difficulties with enhancements in general, as well as with the debate itself. Third, I turn to the research questions and identify aspects we should investigate to resolve the previously established issues. Fourth, I elaborate on the methodological approaches adopted in this thesis, i.e., how I go about answering the research questions. Finally, I present the structure of the dissertation by outlining the individual chapters and their particular aims in this study.

## 1. What Is Moral (Bio)Enhancement and Why Do We Need It?

Technically speaking, moral enhancement represents one of six main categories of human enhancement (alongside physical, cognitive, cosmetic, affective, and longevity enhancement), which are all meant to improve individual human capacities.<sup>1</sup> Moral enhancement primarily targets the improvement of moral bearing (such as agent's character, motives, or behavior), which should, ideally, bring someone's abilities above the species-typical baseline, i.e., making a person "better-than-normal".<sup>2</sup> For example, this means that a morally enhanced person would have more empathy, a better sense of justice or improved reasons-responsiveness compared to an average moral agent. However, moral enhancement could also have a restorative or preventive function by instilling capacities where there are not any or restoring lost capacities (Shook and Giordano, 2017). In latter cases, moral enhancement would overlap with therapeutic interventions, such as the treatment of psychopathic behavior or use of antilibidinal drugs in sex offenders. Moral enhancement could, in principle, be delivered directly (by correcting criminal or deviant behavior) or indirectly (by altering underlying capacities related to moral deliberation and decision-making). Its implementation could span biomedical/pharmaceutical methods, machine-based augmentation, and genetic engineering, and its interventions could be more or less invasive (e.g., via nasal

---

<sup>1</sup> In broad terms, human enhancement is a concept covering all sorts of interventions that can improve the quality of human life in almost any relevant respect. However, in the more recent bioethical discussion, human enhancement primarily refers to the idea of making better human beings by using biomedical interventions. I provide a detailed overview of definitions and demarcations in Chapter 1.

<sup>2</sup> This demarcation is not without its problems, since it implies that there will be "abnormal" and "below-normal" states. Still, it tentatively captures what enhancement comes down to. For a more detailed discussion see, e.g., Daniels (2000).

sprays or neuro-stimulation).<sup>3</sup> Moral enhancement that is delivered by biomedical methods is what we call biomedical moral enhancement, or MBE for short.

Within the bioethical framework, MBE is usually defined as any deliberate biomedical intervention that aims to (directly) improve morally relevant capacities or traits that most or all human beings typically have or, alternatively, to create some new morally relevant capacities (Buchanan, 2011a; DeGrazia, 2014).<sup>4</sup> For the most part, the contemporary discussion focuses on enhancing human motivation to act morally through the application of biomedical and pharmacological means (like drug treatment or genetic engineering), and relies heavily on the biological understandings of human morality. Moreover, the feasibility of such an intervention appears to be backed up by scientific findings, which show that manipulating the biological make-up can have morally desirable effects. Although the debate on MBE is largely hypothetical, there are suggestions in the literature for possible means of pursuing such enhancements – some of which have been widely used in biomedical practice. For example, the administration of powerful neurohormone oxytocin, which works as a neurotransmitter in the human brain, has been shown to promote “pro-social attitudes, like trust, sympathy and generosity” (Savulescu and Persson, 2012: 402). The alteration of serotonin or testosterone levels can “mitigate undue aggression while...ostensibly enhancing fair-mindedness, willingness to cooperate, and aversion to harming others” (Earp et al., 2018: 166). In real life, commonly prescribed types of anti-depressants (like Prozac), as well as methylphenidates (like Ritalin), have been shown to increase cooperation and reduce aggression when given to persons with particular disorders. There is also a possibility to apply newly developed brain modulation techniques, such as transcranial or deep brain stimulation in attempt to reduce impulsive tendencies (Earp et al., 2018).<sup>5</sup> Also, some personality disorders that predispose persons for immoral behavior have been linked to certain biological bases suggesting that if these conditions are better understood, interventions

---

<sup>3</sup> More detail is provided in Chapter 2. For a general overview of common demarcations, delivery methods, and applications across the human enhancement debate see, e.g., Jensen et al. (2018).

<sup>4</sup> By morally relevant capacities, we mean capacities that play important roles in moral deliberation and behavior, such as empathy, altruism, volitional capacities, reasons-responsiveness, etc. We could also talk about neurocognitive functioning that is instrumental to moral thought/behavior (Shook and Giordano, 2017).

<sup>5</sup> For a detailed overview of promising neurotechnological tools for pursuing MBE and primary scientific sources see Earp et al. (2018). More details can also be found in Chapter 2 of this dissertation.

might be developed to improve behavior (Persson, Savulescu: 2008; Douglas: 2008). Thus, proponents of MBE find it reasonable to assume that one could use biomedical means to improve moral motivation and moral decision-making even in healthy individuals.<sup>6</sup>

A highly influential proposal for MBE was developed by Ingmar Persson and Julian Savulescu (2008; 2012) to bridge the gap between scientific and technological progress achieved by humans during the past few centuries and address the alleged absence of comparable progress in what we can describe as our species-typical moral psychology.<sup>7</sup> Persson and Savulescu argue that human beings are not naturally equipped with a moral psychology that would empower them to adequately cope with the moral problems associated with new life conditions (such as threats from weapons of mass destruction, climate change, and environmental degradation). Drawbacks of human moral psychology manifest primarily in the lack of moral motivation and might cause ultimate harm, i.e., endanger human life on earth. Since the mismatch in question presents a severe threat to human survival, MBE should offer some sort of “evolutionary short-cut” or “artificial upgrade” of moral psychology to improve human adaptability to the current environment.

For the most part, this dissertation takes Persson and Savulescu’s (2008; 2012) account of MBE as its point of departure (unless stated differently), but it also takes notice of its evolution (e.g., Schaefer 2015; Earp et al. 2018). The reasons for relying on Persson and Savulescu’s account are mainly the influential nature of this proposal and its high level of sophistication. Also, the dissertation is not thoroughly concerned with MBE’s defining features (what is its “correct” definition) and its biotechnological feasibility (is it, in fact, scientifically feasible). Instead, the focus is mainly on how we should reason about it *if it were to become feasible* (i.e., should we allow it from an ethical perspective). Finally, when referring to MBE, I primarily have in mind less invasive biomedical interventions such as pharmaceuticals rather than highly invasive interventions such as genetic engineering.<sup>8</sup> Although the latter types of interventions might turn out to be more effective,

---

<sup>6</sup> However, this is not to say that there is a necessary causal connection between the two.

<sup>7</sup> I explain this entire background in more detail in Chapter 2 and repeatedly in Chapter 4.

<sup>8</sup> Also, biomedical moral enhancement should be distinguished from traditional moral enhancement, brought about by non-biomedical means such as upbringing, education, training, etc. For the most part, traditional means of moral enhancement are not the primary research subject of this dissertation.

pharmaceutical interventions are generally further along, more feasible, and more likely to become marketable in the foreseeable future.

## 2. Difficulties with MBE and the Enhancement Debate

Given its controversial nature, MBE quickly garnered attention and sparked a great deal of scholarly disagreement. Not only did it raise a number of ethical issues, but the discourse turned out to be rather unsystematic, polarized, and lacking common ground. Although this dissertation attends to many substantive, content-related issues with MBE as an intervention, one underlying goal is also to propose methodological guidelines that can mitigate said difficulties in the debate. Hence, in this section, I briefly outline some difficulties with MBE and the debate surrounding it before proposing how to make progress more concretely.<sup>9</sup>

MBE raises various concerns on conceptual, normative, and socio-political levels. These challenges range from determining what and how to enhance to ensuring proper and harmless application. First, some worry that MBE is conceptually unsound or unfeasible, and that, even if it proves to be feasible, it may have undesirable effects (Harris, 2016; Jotterand, 2011; Hauskeller, 2017). Second, some worry that even if MBE proves to be a coherent and feasible idea, it will still be morally impermissible because it threatens fundamental moral values like freedom and autonomy (Harris, 2012; Sandel, 2007; Kass, 2003). Third, some worry that MBE would exacerbate existing socio-political problems, such as inequality, injustice, discrimination, exploitation, etc., or cause new problems, such as the attainment of higher moral status by the enhanced individuals (Sparrow, 2014a, 2014b; Triviño, 2013; Specker et al., 2014).

MBE quickly became a self-standing debate but it is inextricably linked to the broader bioethical debate on HET and should be situated within it. Although MBE has distinctive features separating it from other enhancements, the discourse is carried out similarly as with HET. Arguably, MBE may be causing even more disagreement than other human enhancement interventions. Positions on MBE range from its approval and prioritization over other modes of human enhancement given its possibly crucial role for the continuation of human life (e.g., Persson and Savulescu, 2008; 2012), to claims that it threatens the very nature of morality and should be outright rejected (e.g., Harris, 2013, 2016; Agar, 2015a; Hauskeller, 2013).

---

<sup>9</sup> An extensive state-of-the-debate is provided in Chapter 1 and Chapter 2.

Perhaps what makes MBE especially interesting and peculiar is that, unlike most other enhancements, moral enhancement should benefit both enhanced (oneself) and unenhanced (others).<sup>10</sup> Also, it is quite common to think that moral progress or ‘more morality’ are good things *per se*, making the opposition to moral enhancement somewhat dubious. Unless in some extraordinary circumstances, a world with more empathy and less crime would surely be a better world than the one we have. However, things are not that simple and, regardless of good outcomes, we might not be willing to pay the price of living in such a world. But, as with every new technology, we are not exactly sure what this price would amount to and whether it would in fact be reasonable to dismiss the benefits of moral enhancement for merely maintaining the status quo. Although caution is often justified, excessive caution can be counter-productive and hinder much-needed moral progress. It is my view (which also motivates this study) that promising technologies such as moral enhancement are worthy of our attention and thorough ethical (and in due time empirical) research.

### 3. The Research Questions

The concern pervading this debate is the moral permissibility of these interventions. It surpasses other concerns because we tend to believe that things can be legal, feasible, or safe, and still be morally impermissible.<sup>11</sup> Given that the moral permissibility of HET, including MBE, is such a fundamental normative issue in a rather disperse and unsystematic debate, an explicit dialogue with ethical theories and employing ethical analysis could provide a more promising, systematic normative framework. Thus, the main research question guiding this project is:

*How can we use ethical analysis (relying on normative ethics and metaethics) to improve the bioethical debate on MBE and our understanding of what is or is not morally permissible?*

---

<sup>10</sup> This holds if we assume that being moral is an intrinsically valuable trait, contributing to a fulfilling, satisfying life – which is generally true. A possibility of taking advantage of enhanced individuals is a legitimate concern but it should not undermine the intrinsic value of the moral character itself.

<sup>11</sup> This claim became known in the debate as the bioconservative thesis because it primarily reflects views of prominent bioconservatives such as Kass, Sandel, and Fukuyama. The term ‘bioconservative thesis’ as I understand it here was introduced by Thomas Douglas (2013).

Hopefully, this will minimize normative ambiguity and guide the moral assessment of different enhancements and their applications. I attempt to show how ethical and metaethical analyses may benefit the discussion and that the application of these methodological interventions yields valuable insights into the permissibility of MBE. For example, I will show that HET (and by extension MBE) are not intrinsically wrong; that MBE is not so obviously supported by utilitarianism or opposed by Kantianism; and on net, its permissibility will likely be contextual, but we have no reason to think it cannot be permissible under realistic scenarios. The main research question can be broken down into three sub-questions that are each addressed throughout two chapters.

Following the rationale to assess the debate and make progress on it, it is essential to first ask: *What are the main challenges in the MBE debate and how can we move towards a systematic evaluation of enhancement projects?* Addressing this question entails systematic study, presentation, and analysis of the most critical points within the human enhancement (Chapter 1) and moral enhancement (Chapter 2) debates.<sup>12</sup> The detailed taxonomical and conceptual distinctions relevant to the overall discussion are primarily based on an extensive literature review. Although many of the problems raised in the human enhancement debate can be reduced to simple commonsense objections against it, others are more complex and require careful conceptual elaboration. The possible misconceptions and oversights associated with human and moral enhancement need to be emphasized to clear the path for more fruitful discussion. Also, the strength and credibility of central *pro-et-contra* arguments, which became most common in the contemporary literature, need to be evaluated. Based on the described analysis and evaluation of main concerns, a proposal on how to make progress in future research can be developed. The thesis makes use of the identified concerns by turning them into methodological guidelines for a systematic assessment of human enhancement projects.

After I define the broad theoretical background and the main conceptual framework, it is necessary to further assess MBE's theoretical implications. The discussion then turns to the second sub-question: *What does a close theoretical examination of enhancement's intrinsic and extrinsic properties imply about its coherence and feasibility and how is this relevant for its moral permissibility?* This somewhat complex question involves looking into theories that are external but integrally

---

<sup>12</sup> On the whole, Chapter 1 sets the stage for Chapter 2, as MBE needs to be situated within a broader context of the human enhancement discussion.

linked to human enhancement and examining whether certain assumptions made within the debate are warranted. I explore one such possibility by looking into philosophical theories of intrinsic value to examine whether HET are intrinsically bad and what implications this may have for their moral permissibility (Chapter 3). In this way, Chapter 3 sets the stage for Chapter 4, as well as the other chapters dealing exclusively with MBE, since it tackles an essential concern which would, if warranted, severely limit the ethical discussion on HET.<sup>13</sup> However, the intrinsic value of HET is not the only relevant subject here; their effects play a major role too. Therefore, the second approach I take is investigating whether a concrete enhancement (in this case, the enhancement of moral motivation via direct emotion modulation), would in principle have the anticipated effects (Chapter 4). These approaches offer valuable theoretical input about the coherence and feasibility of enhancement interventions.

If MBE became biotechnologically feasible, many concerns regarding its effects on valuable aspects of human life would be raised. The third sub-question is thus: *How does MBE align with predominant moral norms and does it conflict with basic moral values?* In order to assess the moral permissibility of MBE, its relationship with predominant moral views needs to be established. Since normative ethics offers specific accounts of rightness and justifiability of a particular act or choice, ethical input seems indispensable in evaluating the moral permissibility of new and emerging technologies. This aspect of the thesis makes use of the normative analysis not only by making evaluative judgments about the moral permissibility of human and moral enhancement, but also by employing normative ethical theories to arrive at these evaluative judgments. In this regard, the thesis is concerned with the relationship between MBE and two predominant but fundamentally different ethical theories – utilitarianism and Kantian ethics. More precisely, Chapters 5 and 6 aim to identify conditions under which MBE may or may not undermine some of the most fundamental moral norms, as viewed by these two ethical doctrines.

---

<sup>13</sup> I take a broader perspective on HET in this chapter because the position I am attacking is directed against all biomedical enhancements, including MBE. Given that MBE is inextricably linked to the broader debate on HET, a more general approach does not take away from my focus on MBE.



#### 4. Methodological Approach

The debate on human enhancement has been as engaging for moral philosophers as it has been for bioethicists. Although some say that ethical theory appears to be unilluminating, unappealing and indeterminate regarding bioethical issues, they also acknowledge that many issues in bioethics are fundamentally problems that should be handled by the methods of moral philosophy (Beauchamp, 2004). I believe human enhancement is one such domain, where moral philosophy can offer an external standard of justification through rigorous conceptual and moral analysis that will positively reflect on the bioethical approach, and eventually, the practice of bioethics itself. Bioethical issues emerging throughout the enhancement debate cannot be autonomously or single-handedly solved via applying overly broad and vague bioethical principles<sup>14</sup> because such approaches lack content and capacity to guide actions, as well as the ability to provide any definite answers to moral problems (Holm, 1995: 337). Therefore, we should not disregard ethical theory, since it can make up for the limitations of the formulaic framework, i.e., without engaging with ethical theory on a deeper level, we miss something morally important (Sorell, 2011).

As mentioned previously, some of the most pertinent issues within the enhancement debate are related to conceptual ambiguities and the lack of in-depth theoretical analysis, making it perfectly reasonable to focus on untangling these ambiguities by employing ethical theory. In the broadest sense, this thesis sets out to explore several theoretical and normative justifications of the moral permissibility of HET, while taking MBE as its distinct research subject, and with the ultimate goal of making progress on the stagnant debate.

In broad terms, the thesis challenges how we reason about MBE and investigates the beneficial effect of including high-level normative and metaethical theories and concepts into the enhancement discourse. Although this work includes a considerable amount of ethical analysis of human and moral enhancement, it is not a full-scale ethical analysis of these technologies. Instead, it aims to provide a *proof of concept* of the broader point on how a dialogue with our best (meta)ethical and philosophical theories can fill the gaps, provide a vantage point, and improve the discourse in the human enhancement debate. This proof of concept is an attempt to show that we *can*, in fact, make progress in our analysis and arrive at

---

<sup>14</sup> I am primarily referring here to basic ethical principles identified by Beauchamp and Childress (1994), intended as universal guidelines for the biomedical approach: respect for autonomy, nonmaleficence, beneficence, and justice.

some conclusions with a higher degree of confidence. Thus, it brings us closer to resolving some basic questions.

In more technical, methodological terms, this dissertation belongs to the tradition of analytic philosophy, which means it deals with concrete (and relatively narrow) problems in a piecemeal fashion, typically adhering to thorough analysis and clear language. In addition to bringing ethical theory into a direct dialogue with the bioethical debate on MBE as a “meta-methodological” solution, the present work proceeds through three main methodological steps: (i) literature review and discourse analysis; (ii) theoretical (conceptual) analysis; and (iii) normative (ethical) analysis.

First, the debate is “diagnosed” by reviewing the most pertinent issues and concerns and by analyzing the surrounding discourse. Since there are extremely polarized and under-theorized points, we need to seek out a methodological framework to tackle the issues at hand. This is why I first develop a broad methodological framework for systematic analysis of the moral permissibility of different HET projects and for moving beyond the pro-ET-contra exchange. Second, I demonstrate the feasibility of this methodological framework (proof of concept) by examining the relationship between enhancement and several external but integrally linked theories (such as the theory of intrinsic value, evolutionary psychology of emotions, utilitarianism, etc.). I am mostly using results of conceptual analysis to clarify particular concepts and terms (such as “intrinsic badness”), which, in turn, enables me to test particular hypotheses (such as “HET are intrinsically bad”). Theoretical analysis and testing common assumptions about enhancement is also supplemented by connecting the theory with some empirical research (such as neuroscientific understanding of emotions). These approaches can help us gain insight into the coherence, plausibility, and feasibility of enhancement projects, and reach more general conclusions about how this reflects upon their moral permissibility. This step in the discussion also forms interdisciplinary connections with, for instance, neuroscientific theories and synthesizes data from related fields, such as bioethics and metaethics. Finally, normative analysis is needed to form evaluative judgments about the moral permissibility of MBE and to provide an external standard of justification. This part of my analysis mostly focuses on testing the moral permissibility of MBE by means of normative ethical evaluation – namely, utilitarianism and Kantianism. I examine, using a top-down approach, whether common assumptions about MBE align with the prevailing ethical principles (such as maximizing wellbeing, agent neutrality, or respect for autonomy/agency).

On the whole, all mentioned methods are used to test and refine some initial intuitions about MBE against the theoretical backdrop with the ultimate goal of balancing out the results (this resembles the reflective equilibrium method applied on a smaller scale).

*The table below summarizes how particular methodological approaches tie to research questions, goals, and individual chapters.*

| RESEARCH QUESTION   | MAIN METHODOLOGY   | GOALS   | CHAPTER   |
|---|--|---|-----------|
| <i>What are the main challenges in the MBE debate and how can we move towards a systematic evaluation of enhancement projects?</i>              | - Literature Review  | - Diagnosing the debate   | Chapter 1 |
|   | - Discourse analysis   | - Development of a systematic framework and assessment guidelines | Chapter 2 |
| <i>What does a close theoretical examination of enhancement's intrinsic and extrinsic properties imply about its coherence and feasibility?</i> | Testing hypotheses by means of:  | - Resolving conceptual ambiguities                                | Chapter 3 |
|   | - Conceptual analysis<br>- Logical analysis<br>- Comparative analysis                  | - Assessing theoretical feasibility and effectiveness             | Chapter 4 |
| <i>How does MBE align with predominant moral norms and does it conflict with basic moral values?</i>  | Conducting normative ethical analysis by consulting:                                   | - Assessing compatibility with moral norms                        | Chapter 5 |
|   | - consequentialism (direct & indirect utilitarianism)<br>- deontology (Kantian ethics) |   | Chapter 6 |

In the following section, I elaborate on the more specific aims of each chapter.

## 5. Chapter Outline

In addition to the introductory and concluding chapters, this dissertation consists of six main chapters which jointly provide a coherent whole but are also intended as stand-alone contributions to the debate. Since all chapters are individual

articles, they capture specific problems and address them more or less concisely. This also explains why some overlap between them was necessary.

**Chapter 1** is specifically concerned with identifying the main challenges in the human enhancement debate and developing a framework for the systematic evaluation of human enhancement projects. This chapter fulfills several purposes. It offers the “state-of-the-art” overview of the human enhancement debate and presents the most relevant points in the discussion. But most importantly, it draws attention to some problems with the discussion where there is extreme polarization and lack of dialogue, and discusses how these deadlocks can undermine progress in this domain. Finally, this chapter outlines a methodological framework by specifying a number of guidelines that could help us make progress in future research. In particular, I propose that we should pay special attention to the following: (i) examining whether a particular enhancement project is plausibly coherent, feasible, and effective; (ii) whether it conflicts with fundamental moral values and norms; and (iii) whether it is compatible with or facilitates socio-political goals of equality and justice. This approach should help us minimize normative ambiguity and facilitate the moral assessment of different enhancements and their particular applications.

This chapter is forthcoming as: Kudlek, K. 2022. Challenges in the human enhancement debate: a critical review. *Techné: Research in Philosophy and Technology*, 26 (2).

**Chapter 2** narrows the research subject down from human enhancement to the specific case of MBE to demonstrate, on a more concrete example, the similar points made in Chapter 1. This chapter thoroughly reviews the debate on moral enhancement, explores the main challenges specific to this domain, and incorporates the same methodological framework developed previously in Chapter 1, with the aim of a systematic evaluation of moral enhancement. Exploring how MBE corresponds to important moral standards identified in Chapter 1 should give us better understanding of its coherence, plausibility, and desirability, which, in turn, reflect its overall moral permissibility.

This chapter is forthcoming as: Kudlek K. 2022. Towards a systematic evaluation of moral bioenhancement. *Theoretical Medicine and Bioethics*.

**Chapter 3** explores a pertinent concern in the human enhancement debate that HET are intrinsically bad and, hence, morally impermissible, in order to show

how careful application of relevant theory can help us dismiss or overcome potentially unwarranted claims in the human enhancement debate. The chapter challenges the claim about the intrinsic badness of HET by looking into philosophical theories of intrinsic value. It investigates how well-established conceptions of intrinsic value map onto typical bioconservative arguments about HET's intrinsic badness. My analysis shows that the debate on intrinsic value places serious constraints on claims about the intrinsic badness of HET. In other words, bioconservative arguments are, for the most part, inconsistent, misconceived, and overly speculative. Enhancement interventions cannot be bearers of intrinsic value on any of its plausible understandings, and, even if we could grant such a possibility, there are no compelling reasons to presume that the intrinsic value of HET would be necessarily negative. As a result, claims regarding their moral impermissibility are unwarranted.

This chapter is published as: Kudlek, K. 2021. Is human enhancement intrinsically bad?, *Medicine, Health Care and Philosophy*, 24, 269-279.

**Chapter 4** further challenges some claims about the feasibility and effectiveness of MBE by emotion modulation by looking into neuroscientific and evolutionary psychological theories of emotions. Similar to Chapter 3, this chapter demonstrates that applying external but integrally linked theories to the enhancement debate can help us reach important conclusions about the coherence, feasibility, and effectiveness of moral enhancements. It estimates the theoretical and conceptual soundness of the central argument underlying the MBE project, i.e., the claim that emotions can serve as vehicles of moral enhancement. Through careful examination of the principles endorsed and ideas proposed by MBE proponents, this part of the dissertation reveals potential problems that might ensue for MBE (as initially conceived). It is argued that MBE indeed seems to be biotechnologically feasible, but its piecemeal implementation might have (morally) undesirable outcomes.

This chapter is published as: Kudlek K. 2019. The role of emotion modulation in the moral bioenhancement debate, *Topoi* 38, 113-23.

**Chapter 5** is one out of two chapters that aim to demonstrate how applying normative ethical theories can help us further illuminate the moral permissibility of moral enhancements. It investigates the effects of MBE for fundamental moral norms on the example of utilitarian morality. In particular, this chapter challenges the common assumption that MBE rests on consequentialist principles and is

justified on utilitarian grounds. It seems that MBE could modify moral agents to accord with main utilitarian demands and facilitate the adoption and realization of utilitarian prescriptions. Although MBE would, in principle, create preconditions for achieving utilitarian ends, this chapter shows that there are certain limits to this claim.

This chapter is published as: Kudlek, K. 2022. On the uneasy alliance between moral bioenhancement and utilitarianism, *Bioethics* 36, 210– 217.

**Chapter 6** continues to examine MBE's relationship with predominant ethical theories, but in contrast to Chapter 5, it takes Kantian ethics as its primary focus (given that consequentialism and deontology are thought to stand in opposition). More specifically, this chapter challenges the common assumption that MBE is at odds with deontological doctrines. It looks into deontologically permissible conditions and constraints on how and why we might engage in MBE from the perspective of Kantian ethics. The analysis shows that a careful engagement with Kantian moral psychology does provide space for MBE, but it also describes a potential danger brought about by MBE.

This chapter is published as: Kudlek, K., Smith, P.T. 2022. The Kantian promise and peril of moral bioenhancement, *Journal of Applied Philosophy*, <https://doi.org/10.1111/japp.12575>.

**Conclusion** of this dissertation summarizes the results of my research by highlighting the chapter's individual and joint contributions to the debate. Next, it elaborates on some strengths and limitations of my approach and emphasizes what my methodology allowed us to conclude about the moral permissibility of human and moral enhancement. Finally, I discuss some broader implications of my work and outline the potential avenues for future research.





A. Karametas, *Democracy* (2015)



# Chapter 1

## Challenges in the Human Enhancement Debate: A Critical Review

**Abstract:** The discussion on human enhancement technologies has been mostly focused on exchanging views about the dangers and benefits of these technologies. Although the ongoing debate has been exceptionally stimulating, it is defined by extreme polarization and lack of consensus about what constitutes human enhancement, do we want it, and should we allow it. We need a systematic attempt to move beyond pro et contra exchange. Thus, in this paper, I capture and analyze some issues with the human enhancement debate and I outline a set of methodological guidelines that could help us make progress in future research. I propose that we should pay special attention to the following conditions: (i) examining whether a particular enhancement project is plausibly coherent, feasible and effective; (ii) whether it conflicts with fundamental moral values and norms; and (iii) whether it is compatible with or facilitates socio-political goals of equality and justice. This approach should help us minimize normative ambiguity and facilitate the moral assessment of different enhancements and their particular applications.

This chapter is a version of the article: **Kudlek, K. 2022. Challenges in the human enhancement debate: a critical review. *Techné: Research in Philosophy and Technology*, 26 (2) [forthcoming].**



# Challenges in the Human Enhancement Debate: A Critical Review

## 1. Introduction

The discussion on human enhancement technologies (HET) has been mostly focused on exchanging views about the dangers and benefits of these technologies. Although the ongoing debate has been exceptionally stimulating, it is defined by extreme polarization and lack of consensus about what constitutes human enhancement, do we want it, and should we allow it (Clarke, 2016).<sup>15</sup> We need a systematic attempt to move beyond pro et contra exchange. Thus, the purpose of this paper is to capture and analyze some issues with the human enhancement debate and to outline a set of methodological guidelines that could help us make progress in future research.

The debate seems to be broadly concerned with the technological feasibility, practical implementation, and moral permissibility of new and emerging technologies. The last question, however – that of moral permissibility – surpasses other concerns because we tend to believe that things can be legal, feasible, or safe, and still be morally impermissible.<sup>16</sup> The moral permissibility of HET is, therefore, the most important underdeveloped normative concern requiring careful in-depth examination. Hence, I propose that we, amongst other things, place the debate in more explicit dialogue with ethical theories. This can help us minimize normative ambiguity and facilitate the moral assessment of different enhancements and their particular applications.

The first part of this article engages various conceptual and ethical disagreements in the human enhancement debate by looking into general concerns and direct responses from the pro-enhancement literature. This review of typical pro et contra arguments primarily captures how the discourse has been carried out

---

<sup>15</sup> Since human enhancement is best understood as an intervention brought out by biotechnological means, I will use “human enhancement” and “human enhancement technologies” more or less interchangeably, taking that both expressions roughly amount to the same thing.

<sup>16</sup> This claim became known in the debate as the *bioconservative* thesis because it primarily reflects views of prominent bioconservatives such as Kass, Sandel and Fukuyama.

(but it is not meant to be exhaustive or provide a conclusive evaluation of the raised arguments). Although I focus on constructive internal criticism of the existing debates, this does not foreclose more critical perspectives on the literature.<sup>17</sup> Some views on the strengths and weaknesses of the presented arguments are expressed throughout the article.

The second part aims to identify issues with the current debate and show that there may be other, perhaps more constructive, ways to make progress. My suggestion is that we should pay special attention to the following conditions: (i) examining whether a particular enhancement project is plausibly coherent, feasible and effective; (ii) whether it conflicts with fundamental moral values and norms; and (iii) whether it is compatible with or facilitates socio-political goals of equality and justice. I believe these aspects should guide our examination of HET and jointly they may provide valuable input about their permissibility.

The article proceeds as follows. Section 2 begins by laying out the overall conceptual and taxonomical disagreements such as the ones surrounding the enhancement definition; traditional and biomedical aspects (2.1); enhancing well-being (2.2); and the enhancement-treatment distinction (2.3). Section 3 engages in various ethical disagreements, addressing concerns about changes in human biology, genes, and nature (3.1); playing God and the quest for perfection (3.2); exacerbating social issues of equality and justice (3.3); and loss of spontaneity and the mystery of life (3.4). In Section 4, I identify issues with the enhancement discourse and reflect on the plausibility of presented arguments. In Section 5, I propose a set of methodological guidelines that could help us investigate the matters of moral permissibility of HET in a more systematic manner.

## 2. Conceptual and Taxonomical Difficulties

The human enhancement debate is facing many difficulties concerning the definition of its central term and the demarcation of its conceptual features. The concept of human enhancement is exceptionally indeterminate and, accordingly, applicable to all sorts of improvements in quality, value, or extent of specific capacities. In the broadest sense, “to enhance human beings is to expand their

---

<sup>17</sup> By “internal” criticism I here understand that my primary focus is to capture the state of the discourse – taking the debate at face value – and not to engage in conclusive external evaluation of which side in the debate is right. However, I do offer some views about plausibility of presented arguments, but these evaluations do not affect my main goal.

capabilities – to enable them to do what normal human beings have hitherto not been able to do” (Buchanan, 2011a: 38). For example, inventions such as tools, clothes, or the wheel can all be considered human enhancements. Numeracy and literacy are surely two of the biggest and most significant cognitive enhancements in history – enabling the development of other enhancements such as computers. The agrarian revolution of the 18<sup>th</sup> century caused radical positive changes in physical and cognitive capacities, increasing human longevity to boot. Cultural institutions, medical achievements, scientific and technological discoveries – these can all be counted as enhancements. Particular examples such as vaccination, eye-glasses, calculators, smart-phones, shoes, clothes, even certain foods and drinks are often included on the lists of widely used enhancements.

Within the bioethical framework, enhancements can be understood as any kind of intervention that improves some characteristics or capacities that human beings ordinarily have, or more radically, that bring about new characteristics or capacities (Buchanan, 2011a: 5). Enhancements typically fall into six chief categories: (i) physical enhancements (improving characteristics such as speed and strength); (ii) cognitive enhancements (improving characteristics such as memory and reasoning); (iii) cosmetic enhancements (improving cosmetic traits); (iv) affective enhancements (improving emotion control, motivation or temperament); (v) moral enhancements (modifying morally relevant dispositions); and (vi) longevity enhancements (increasing lifespan). Early signs of such interventions are already around us (anti-depressants, doping, cosmetic surgery, etc.), while others are still only speculated about (such as the genetic engineering of embryos or nano-technological implants).<sup>18</sup>

The term human enhancement can be approached from various angles; we could emphasize its socio-pragmatical and, therefore, relative construction (Parens, 1998; Canton, 2003), contrast it with treatment or medicine in general (Juengst, 1998; Pellegrino, 2004), or focus on enhanced functions (Daniels, 2000; Engelbart, 1962).<sup>19</sup> More recent attempts to define human enhancement hone in on the value of well-being or goodness of one’s life, defining enhancement as an improvement or a change in a person’s life which is good (Savulescu, 2006; Kahane et al., 2011). It could be said that, in a sense, enhancements begin where *normal* capabilities end, enabling us to overcome the limitations of our current

---

<sup>18</sup> A similar categorization of enhancements can be found in Bostrom and Savulescu (2009) and Buchanan (2011a).

<sup>19</sup> For a systematic overview and further references see Savulescu (2006).

species-typical capacities.<sup>20</sup> The “threshold” of normality has been raised throughout human history and many examples of what is currently normal were once enhancements. To name one, the life expectancy in the Stone Age was 20-34 years, whereas in some parts of the world today it is estimated at over 80 years, primarily owing to social and technological developments (Bostrom and Roache, 2008).<sup>21</sup> While some advocate further development of human enhancement technologies, others are concerned that improving ourselves beyond our current normal functioning could lead to creating something other-than-human. For example, some believe that biomedical enhancements could produce posthumans (beings of a different species from *Homo Sapiens*) or that they could produce postpersons (beings with a higher moral status than that of “current” persons) (see, e.g., DeGrazia, 2012; Rakić, 2015). The possibility of overstepping our “boundaries” and creating human beings substantially different from what is familiar causes much commotion in contemporary discussions. One of the primary sources of disagreement is whether there is any relevant difference between biomedical enhancements and ordinary human enhancements.<sup>22</sup>

## 2.1. Traditional and Biomedical Aspects

It is common to distinguish between human enhancement that results from applying, on the one hand, non-biomedical methods, and, on the other, biomedical methods. Examples of non-biomedical or “traditional” methods of enhancement are education (both formal and informal), training, upbringing, nurture, and other similar ways of *indirect* influence on characteristics and capacities. Conversely, biomedical methods include the application of science to enhance specific capacities

---

<sup>20</sup> See, for example, Daniels (2000).

<sup>21</sup> This example indicates that enhancement/treatment distinction is relative to an era’s technology.

<sup>22</sup> Newer forms of biomedical enhancements are usually thought to be radically different from the older forms of enhancements. On the other hand, it can be claimed that there is no relevant difference between new and old enhancements. This argument is also known as “avoiding biomedical enhancement exceptionalism” and it is used to clarify the lack of any substantial, morally significant difference between biomedical and traditional enhancements – or otherwise – to emphasize that great historical enhancements have the same morally problematic features as the new ones (Buchanan, 2011a).

more *directly* (e.g., by pharmacological or genetic means).<sup>23</sup> This differentiation provides the basis on which traditional (non-biomedical) human enhancement and biomedical human enhancement are standardly morally evaluated. It is implied that traditional enhancements are harmless and permissible, whereas biomedical enhancements are usually perceived to be harmful and impermissible. Some ethicists find this distinction unjustified by objecting that most biomedical enhancements, in their view, are not different from the traditional ones in any relevant respect.

Therefore, assessments of the ethical permissibility of enhancements often reflect common intuitions associated with traditional and biomedical enhancements. The easiest way to resolve the dilemma is to deem traditional methods of enhancement acceptable and biomedical methods unacceptable. These intuitions are widespread in the enhancement debate. Traditional enhancements seem unproblematic to the extent that there is no need to label them as enhancements – instead, the enhancement label is almost exclusively pinned onto biomedical techniques.

There are two extreme positions in the enhancement debate – the *bioconservative* and the *transhumanist* view. Bioconservatives advocate extreme caution and oppose the development of virtually all human enhancement technologies, claiming they might be harmful to individuals and society in various ways. Transhumanists hold that we should seek to develop technologies and make them available in order to protect and expand valuable aspects of individual lives such as health, cognition, and emotional well-being (Bostrom and Roache, 2008).<sup>24</sup> There are many moderate positions on the spectrum between these two extremes, holding that enhancements should not be banned across the board, but approached with great caution and evaluated on a case-by-case basis.<sup>25</sup>

Many proponents of bioenhancement hold a more nuanced view: they are neither against biomedical enhancements wholesale nor unconditionally supportive of them. This subtle difference is emphasized by Buchanan (2011b), who sorts the participants in the enhancement discussion into “anti-enhancement” and “anti-anti-enhancement” camps. Anti-enhancement authors hold that enhancements

---

<sup>23</sup> It is important to emphasize that unlike biomedical, traditional methods do not involve direct application of science. So, we cannot infer that non-biomedical enhancements such as exoskeletons fall under traditional enhancements.

<sup>24</sup> This position can be also taken in a more extreme way suggesting that we should fully “transcend” our humanity by technological means.

<sup>25</sup> For moderate enhancement views see, e.g., Mukerji et al. (2014).

ought to be avoided across the board. Some of their arguments against enhancements are highly general, opposing all enhancements in all circumstances (e.g., Kass, Sandel, President's Council on Bioethics). Their opponents from the anti-anti-enhancement category are not opposed to all kinds of enhancements – even though they may have serious reservations (e.g., Glover, Savulescu, Agar, Bostrom, DeGrazia, Sandberg, etc.). Buchanan finds an anti-anti-enhancement position more reasonable because it is not monolithic: “it includes some who are enthusiastic about a rather wide range of enhancements and some who are much more cautious and skeptical” (2011b: 14). Anti-anti-enhancement proponents are generally unconvinced by arguments that enhancement as such is immoral or always too risky, and yet they are ready to acknowledge the possibility of serious risks.

## 2.2. Enhancing Well-Being

Why are biomedical enhancements thought to be so problematic and risky? A possible explanation is the concern that enhancements will not necessarily make us better off overall. Biomedical enhancement, by definition, involves direct application of biomedical science to one's body or brain with the goal of improving existing or creating new capacities (Buchanan, 2011b: 23). These biomedical enhancements can be categorized in two ways: according to the *type* of capacity that they aim to improve (e.g., cognitive function, physical strength, mood, temperament, or longevity) and according to the *mode* of intervention (the technology used to improve the capacity) (Buchanan, 2011a: 5). This distinction helps us to avoid a common mistake: the presumption that enhancement makes one better off by definition. Buchanan emphasizes that enhancement is an improvement of some particular capacity, but not necessarily something that makes us better off all things considered – in other words, we should talk about enhancing *capacities* rather than enhancing *people*. An apparent example is enhanced hearing in a noisy environment, which could easily make one worse off (2011b: 23). Similarly, an enhanced capacity such as enhanced cognition or enhanced mood does not necessarily equal enhanced well-being (Landeweerd, 2011). The nature of enhancements is complex and it is important to bear in mind that enhanced capacity does not necessarily benefit the quality of a person's overall life.

One way to get around this concern comes in the form of a more precise definition of human enhancement called the *welfarist account*. The need for a clear and shared concept of enhancement that will enable us to answer many ethical



questions raised by the emerging possibility to enhance normal human capacities has been raised in the discussion (Savulescu, 2006; Kahane et al., 2011). By reviewing several accounts of enhancement, Kahane et al. identify its two key senses: “*functional* enhancement, the enhancement of some capacity or power (e.g., vision, intelligence, health) and *human* enhancement, the enhancement of human being’s life” (2011:3). The welfarist account appears to be most relevant to the ethical debate and is specified as follows: “Any change in the biology or psychology of a person which increases the chance of leading a good life in the relevant set of circumstances” (2011: 7). As the example of improved hearing in a noisy environment shows, enhancing a specific capacity can lead to a worse-off life overall. The welfarist account of enhancement, unlike many other attempts, is inherently normative – it ties enhancement to the value of well-being:

The welfarist approach distinguishes ways in which some treatment might benefit a person from other relevant values, such as justice. It thus allows us to say that although some treatment is enhancement (i.e. contributes to individuals’ well-being), it might be bad overall, because its employment in the current social context will lead to far greater injustice. (2011: 7)

Although the welfarist approach has its advantages, it leaves some questions open. For example, it is not specified what welfare entails, so we lack adequate criteria for discerning right or permissible action. It can be claimed that well-being “is either too subjective to define neutrally and use as an ethical assessment tool, or too broad since to name everything people value ipso facto a “promoter of well-being” does not provide us with a useful criterion to discern what counts as an enhancement and what as the opposite” (Landeweerd, 2011: 215). The welfarist account could be improved by applying more concrete normative guidelines – such as those we can find in normative ethical theories. However, this is something I will address in more detail in Section 5.

### 2.3. Enhancement-Treatment Distinction

Distinguishing enhancements from (medical) treatments has been a persisting issue in the enhancement debate. It is usually considered that enhancements begin where medicine ends or, in other words, enhancements are understood as interventions that go beyond the restoration of health (Juengst, 1998; Pellegrino, 2004). They improve functioning that is not considered necessary to sustain or restore good health. Some authors find the “not-medicine” approach to

enhancement indeterminate because of the highly complex definitions of treatment (Kahane et al., 2011: 4-5). They even doubt that a consistent and useful distinction between treatment and enhancement is possible. One option is to use the familiar model that statistically determines the standard of normal species-functioning and to argue that society must help reach it (Kahane et al., 2011). Under the normal species-functioning model, a conception of disease is: “Any state of a person’s biology or psychology which reduces species-typical normal functioning below some statistically defined level.” (Kahane et al., 2011: 5). This would implicate that “enhancement can be thus defined as improvement in human functioning that goes beyond what is needed for medical treatment” and it brings us to the normal species-functioning definition of enhancement as “[a]ny change in the biology or psychology of a person which increases species-typical normal functioning above some statistically defined level” (Kahane et al., 2011: 5).

However, as Kahane et al. (2011) rightly point out, this model can be very insensitive towards some people’s needs. For example, if a person’s condition is only slightly above this statistically defined threshold and classified as an enhancement, the person might be denied access to treatment covered by basic health insurance. Although this may seem like an artifact of bureaucracies, it points to greater conceptual difficulties with defining what is normal and on what basis do we distinguish between necessary and unnecessary treatments. The welfarist account claims to avoid this issue because it makes no use of the distinction between medical treatment and enhancement: “on the welfarist account, common medical treatments are enhancements, or more precisely, a subclass of enhancements” (Kahane et al., 2011: 8). Instead, it focuses on increasing the value of a person’s life. The question of whether and when we should enhance becomes: when should we increase human well-being? On the welfarist account, the permissibility of intervention depends on the following:

1. The account of well-being we employ.
2. Whether the modification is expected to increase the chances of the person in question leading a good life in the circumstances likely to be obtained.
3. Whether there are reasons to prefer modifications of the natural or social environment.
4. Whether the modification will harm others or create or exacerbate injustice (2011: 16).

Nevertheless, blurring the line between therapy and enhancement can be considered an issue in itself (e.g., Sandel, 2007; Fukuyama, 2002). This problem became

known in the context of cognitive and mood enhancement as the *medicalization problem*. Many authors express concern that “normal” emotional and social problems will soon become regarded as medical (e.g., Berghmans et al., 2011; Schermer et al., 2011; Brülde, 2011). For example, the medical treatment of extreme shyness or social anxiety might be an unnecessary extension of the limits of medicine. This may be reason enough for keeping the distinction between treatments and enhancements, along with making it more transparent and better regulated.

In the context of human enhancement, conceptual difficulties are often entangled with ethical concerns, making it challenging to resolve the former in isolation from the latter. In what follows, I identify and analyze key ethical concerns associated with biomedical enhancements, followed by some common responses to those concerns.

### 3. Ethical Concerns

Apart from conceptual difficulties, enhancements raise numerous ethical issues. They are often perceived threatening to some of the most valuable aspects of human life. Given that enhancements come in various forms and that their impact cuts across all aspects of modern human life, it comes as no surprise that concerns regarding these technologies are heterogeneous.<sup>26</sup> The concerns usually regard a range of negative effects they may cause or point to the controversial nature of enhancement technologies. For opponents, enhancements are either expected to bring about bad consequences or are bad in themselves. Apart from threatening to undermine societal values such as freedom, autonomy, dignity, equality, fairness, or safety, enhancements may face significant practical challenges such as ensuring fair regulation, access, distribution, implementation, or avoiding state imposition. In what follows, I will identify and analyze four broad categories of concerns about human enhancement, followed by responses or counter-arguments to these concerns. These include: (1) changes in human biology, gene pool, or nature; (2) playing God and the quest for perfection; (3) exacerbating socio-political issues of equality and justice; and (4) the loss of spontaneity and the mystery

---

<sup>26</sup> I decided to use the term “concern” instead of “objection”, based on A. Buchanan’s (2011b: 71) distinction between these two concepts. “Objection” is a conclusive argument showing that something is undesirable, while “concern” is merely a consideration that may not add up to some conclusion. In the human enhancement discussion, we mostly encounter the latter.

of life.<sup>27</sup> The overall goal of this section is to demonstrate that typical ethical concerns regarding HET often lack argumentative strength and philosophical soundness. These shortcomings are negatively impacting the overall quality of the debate.

### 3.1. Concern: Changes in Human Biology, Genes, and Nature

There is a general concern that biomedical enhancements could lead to irreversible and harmful changes in the human genome, biology, and nature – possibly forever destroying some of the most valuable aspects of human life – and therefore such interventions should be avoided or banned.<sup>28</sup> Various formulations of this concern are encountered in the contemporary literature on enhancement. It is sometimes expressed as the concern about the “unnaturalness” of enhancements (Kass, 2003). For example, enhancement skeptics worry that biomedical interventions might cause irreparable damage to the human gene pool. This fear seems to have been so great that the human gene pool has already been proclaimed “common heritage” of mankind and, as such, a normative plea has been made for its preservation.<sup>29</sup> For example, Juengst is concerned that biomedicine poses a danger to our ‘integrity’ as a species and humanity as a whole (2009). A more radical version of this approach comes from Annas (2002), a human rights lawyer, who is known for his relentless quest to ban all species-altering research. Annas even requested that such research is proclaimed “a crime against humanity” in a convention designed to protect human rights in the age of genetic technology.

#### Response

Although such concerns and propositions are relevant, many scholars find them to be exaggerated and failing to distinguish between essential concepts. For example, Harris (2007) and Buchanan (2011a, 2011b) both wonder whether it is

---

<sup>27</sup> This list is not meant to be exhaustive, but it should cover some of the most pertinent issues in the debate. For alternative categorization of ethical concerns in the human enhancement debate see, e.g., Allhoff et al. (2010); Bostrom and Roache (2008); or Miah (2012).

<sup>28</sup> As one of the reviewers rightfully pointed, this concern has a more scientifically sophisticated version that “we cannot do just one thing” when changes to the genome and DNA are concerned.

<sup>29</sup> The notion that the human genome must be preserved as part of the common heritage of humanity was first included into the UNESCO’s Universal Declaration on the Human Genome and Human Rights from 1997. The same idea was proclaimed by the Council of Europe in 2001.

erroneous to talk about the ‘preservation’ of the human species and the human genome.<sup>30</sup> Firstly, the human gene pool is not static. It is continuously changing without deliberate human intervention (e.g., genetic mutations, natural selection, and even the agrarian revolution have all been causing radical changes in the human gene pool). Secondly, we should not assume that changing the gene pool is always wrong – on the contrary, some changes, such as curing nasty genetic diseases, should be welcomed.<sup>31</sup> Also, it is important to note that only one kind of biomedical enhancement involves changing genes: the genetic engineering of human embryos. Hence, the fear of causing damage to the human gene pool is not even applicable to considerations of whether biomedical enhancements are permissible.

Similar concerns are raised for the sake of human nature and biology. It is usually assumed that “our biology is what is natural and that the natural is what is good” (Buchanan, 2011a: 20). The natural and the good are two distinct conceptual categories – the natural is not always good and the unnatural is not always bad (Kamm, 2005). There are many aspects of human nature as well as biology that would be worth changing, so we should not assume that every change is adverse. Apart from this, it is crucial to keep in mind what the notion of human nature in fact means. Some creatures’ nature is determined as a set of traits which are explanatory of what interests us about it (the traits are the results of a range of environmental and genetic variations) (Daniels, 2009). Daniels specifies conditions that would have to be met if we want to think seriously about changing human nature: it is a concept that applies to populations, and it is dispositional and selective. In order to change human nature, our interventions must extend to the whole population of humans, and target a trait central to that nature (e.g., emotionality).<sup>32</sup> The intervention would also have to be radical to the extent that it

---

<sup>30</sup> An interesting remark that can be found in the referenced texts is that, paradoxically, the only method that could keep the human genome intact is in fact human reproductive cloning. Meanwhile, sexual reproduction keeps tampering the genome by bringing about random combinations of genes (Harris, 2009). In his recent writings, Juengst seems to be aware of this paradox and mildly modifies his views.

<sup>31</sup> The premise here is, of course, that curing a disease makes a person better off overall and that such enhancement is not enforced in any way.

<sup>32</sup> The example given by Daniels (2009: 35) describes an isolated island populated by humans who discover a powerful mind-altering drug which initially relieves tension, but also irreversibly flattens their emotional responses. They become completely emotionally unresponsive and even transfer the trait to their offspring. Since emotionality is one example of a central human trait, destroying it at the population level would represent changing human nature.

would modify the explanatory role of that trait (the prevalence of the select traits needs to be shifted dramatically). According to this view, modifying human nature is not impossible but “much that is envisioned in the realm of genetic intervention does not involve modifying human nature (even if it sometimes modifies an individual’s nature)” (2009: 26).

### 3.2. Concern: Playing God and the Quest for Perfection

Because of its implication that any deliberate intervention in what is considered human nature would amount to playing God and could lead to unpredictable shifts in the *natural* order (President’s Council on Bioethics, 2003), this concern is often closely associated with the previous one. An additional worry is that we lack sufficient knowledge to engage in biomedical enhancements and reliably predict possible risks. Thus, we should remain humble and let natural processes take their course. Playing God by designing ourselves and our children is considered by some to be the quest for perfection and mastery or as “the anxious excess that misses the sense of life as a gift” (Sandel, 2009: 82). This view is usually combined with the belief that God, nature, or evolution are “Master Engineers” who intelligently designed their products – making them perfect just the way they are. Present-day organisms that are results of the extremely long process of evolution are often considered to be ‘finely balanced’ and ‘delicately integrated’ (Kass, 2003).<sup>33</sup> Therefore, any intervention that would tamper with the design of a “master engineer” is considered dangerous.

#### Response

First, in line with their response to the previously mentioned concern about interference with the natural order, advocates of enhancement often respond that people have been interfering with nature and the natural order ever since they started curing diseases. It is inconsistent, therefore, to claim that the natural course should be preserved, while holding onto the achievements of modern medicine. It could also be claimed that developing and implementing enhancements is part of a natural course – another step in the course of evolution. Ultimately, the only rationale behind the “do not play God” slogan might be a simple warning not to forego caution or show unjustified confidence in our ability to control

---

<sup>33</sup> This particular view is taken from the President’s Council on Bioethics report, *Beyond Therapy: Biotechnology and the Pursuit of Happiness* (2003).

technologies (Buchanan, 2011a). Buchanan holds that warnings such as not to play God or seek perfection and mastery can be understood as a general plea of caution that pertains to all technologies, not just biomedical ones. Most pro-enhancement authors are not opposed to extra cautious approach to new and emerging technologies, on the contrary, they encourage it.

Second, scholars note that there is also something odd and erroneous about the master-engineer metaphor, regardless of whether it uses God or natural selection as its reference point (Buchanan, 2011a). If we take evolution seriously from the scientific point of view, we have to acknowledge that nature *does not* exhibit flawless intelligent design, instead it often produces suboptimal products driven by reproductive fitness.<sup>34</sup> Therefore, conservative bioethicists who employ this analogy might be revealing a gross misunderstanding of how evolution works. Most of us are suffering the consequences of our suboptimal design daily and it is difficult to see why it would be impermissible to improve this design if possible. Even if evolution made us successful organisms from biological point of view (in terms of reproductive fitness), we might want more than that. In fact, most people do believe there's more to life than basic survival and reproduction, and enhancements could increase the quality of particular aspects of our lives which are not promoted by evolution.

### 3.3. Concern: Exacerbating Social Issues of Equality and Justice

The possibility that biomedical enhancements will exacerbate existing socio-political issues and/or create new ones is one of the most pressing concerns in the enhancement debate. The problems of inequality and injustice could be deepened by a gap created between enhanced and unenhanced individuals or societies (Sparrow, 2014a; 2014b). Such discrepancies can facilitate elitism, discrimination, domination, exploitation, etc. (Specker et al., 2014; Triviño, 2013). Enhancement sceptics often pose questions such as: Won't biomedical enhancement further deepen the gap between the rich and the poor (assuming enhancements will be expensive and available only to the highest social echelons)? Would biomedical enhancements make certain types of discrimination more likely (e.g., preference for enhanced individuals in the employment process)? Could biomedical

---

<sup>34</sup> Some examples of suboptimal design in humans, are: the inability to synthesize vitamin C; poor drainage in sinuses; the dual usage of the human pharynx – air intake and food intake, etc. (Buchanan, 2011a: 30; 2011b: 157).

enhancements be interpreted as unearned advantages – causing a lower appreciation of personal achievements? Isn't it unethical to develop technologies that we know will increase unfairness? Will the enhanced enjoy a higher moral status compared to the unenhanced – especially if it is possible to create beings that are something other than human?

The concerns about fairness and equality are strongly related to positional goods – goods whose value is dependent on others not having them (Sandberg, 2011: 83). A positional good like height, for example, confers a relative advantage if not everyone has it (Savulescu, 2006). Cognitive and physical enhancements are often associated with fostering positional goods – like increasing IQ, height or strength. It is often suggested that such a state of affairs can provide grounds for abuse of the resources in question by directly advantaging the enhanced at the expense of the unenhanced, especially in competitive environments like business or sports.<sup>35</sup> Sandel claims that we need to endorse the notion that natural talents that some people have are undeserved – a result of the genetic lottery (2009). Otherwise, Sandel thinks, the successful will become even more likely to view themselves as self-made or wholly responsible for their success, while those at the bottom will be viewed as simply unfit in such an unforgiving environment.

## Response

The enhancement advocates respond to these concerns by pointing to the relative value of enhancements, i.e. to cases where they can help to reduce or prevent inequalities caused by natural lottery. This idea rests on the generally accepted view that we have an obligation to intervene in the natural lottery for the sake of equal opportunity (Savulescu, 2006).<sup>36</sup> It is commonly argued that enhancements are not new, and neither are the problems that they bring to the fore. By simply looking at past experiences, we see that what in the beginning seemed dangerous or unavailable to many soon became accepted and widely accessible (e.g., certain drugs, medical treatments, or technologies such as computers and cell phones). In light of these past enhancements, we notice that some are already more

---

<sup>35</sup> Positionality is here best understood as a continuum rather than as yes/no: some goods are more positional or even have positionality dominate, but that doesn't mean they are solely positional. IQ, height or strength are valuable in themselves, but they can provide comparative advantage in competitive environment. For more about positional and non-positional goods, see Buchanan et al. (2000). For more about enhancement, fairness, and sport, see Savulescu (2006).

<sup>36</sup> A more detailed discussion on this topic can be found in, for example, Buchanan et al. (2000), and Harris (2007).



enhanced than others in the world we inhabit. For example, most children born in affluent countries are healthier, have higher IQs, have fewer psychological problems, and lead much longer lives (Buchanan, 2011a). Such discrepancies constitute *social lottery*, which in combination with the natural lottery, makes enhancements seem like a potential remedy.

Enhancements are often wrongly associated with fostering positional goods because most goods promoted by enhancements have intrinsic value. For example, most cognitive functions are not purely positional goods: “their immediate value to the possessor does not entirely depend on other people lacking them” (Bostrom and Sandberg, 2009: 329). Even if we still fear that biomedical enhancements will be distributed unjustly, we must take into account that this is not a legitimate reason to ban their development. For example, Harris (2007) argues that the mere fact that we cannot fulfill every need for an organ transplant in the world does not imply we should not fulfill the ones that we can. It does not follow that just because something cannot be instantly accessible to everyone it should not be available to some. Life is replete with examples of unfairly distributed goods: top education or expensive drugs are still not available to all, and sadly, the world will hardly ever run out of such examples. It is impossible to erase all injustice and inequality, but that does not imply that goods such as biomedical enhancements are in themselves bad or unjust. Enhancements hold potential for us to cope with some of the most pressing issues of the modern world like famine, poverty, or climate change. The issues of fair distribution or access are not plausible arguments against these technologies themselves, but a matter of social and political regulation.

#### **3.4. Concern: Loss of Spontaneity and the Mystery of Life**

I mentioned previously that certain anti-enhancement philosophers, such as Sandel, believe that biomedical and genetic enhancements represent an aspiration to remake human nature and take total control of our lives. The flip side of that aspiration, Sandel claims (2009), lies in the possibility of destroying appreciation of the gifted character of human powers and achievements. Genetic enhancements will “undermine our humanity by threatening our capacity to act freely, to succeed on our own and to consider ourselves responsible – worthy of praise or blame –

for the things we do and for the way we are” (2009: 78).<sup>37</sup> Apart from undermining our humanity, Sandel also fears that enhancements will put us at risk of losing spontaneity and “openness to the unbidden”. This means that enhancements might eliminate the element of surprise or unpredictability from significant life events such as the birth of a child. Not knowing what to expect or what life will bring can be considered valuable. Removing the element of mystery could make our lives dull and uninteresting. Another aspect that is usually associated with this concern is the loss of meaning or, to put it differently, a loss in value that accompanies living a safe and predictable life.

### Response

Although most of us are likely to admit that spontaneity and mystery of life have their appeal, they can hardly outweigh the appeal of a safe, disaster- and disease-free life. Either way, we will want to decide for ourselves whether we value surprise or predictability more. Human enhancements should be seen as an opportunity to finally remove many of our biopsychological limitations, rather than as constraints to our freedom:

Our biological and psychological nature as individuals represents a barrier to our own wellbeing, to moral behaviour and to love. This nature constrains our freedom. (...) Human Liberation is the concept that our biology and psychology present impediments to wellbeing, social justice, economic productivity, morality, human relationships, and the existence of humanity. We will soon understand these impediments better and may be able to liberate ourselves from the constraints imposed on us by our biology and evolutionary origins (Savulescu, 2010:4.14-15).

The ability to live with imperfections or even grow fond of them does not remedy the wrongness of not improving the quality of life if possible (Savulescu, 2009). Savulescu argues that not only should enhancements be *permissible*, but we should have a *moral obligation* to pursue them if they become safe and available (2009; 2010). For example, not using the opportunity to give your child the best possible

---

<sup>37</sup> For example, Sandel (2007) and Habermas (2003) argue that genetic enhancements of children would represent a threat to their freedom – forcing them to endorse their parents’ desires as their own. However, this view is flawed, since genetic enhancement implies engineering the genotype and not designing the child.

life can be considered selfish and irresponsible. In the absence of some good countervailing reason, it would be wrong not to act on such an opportunity.<sup>38</sup>

Enhancement critics may argue that “normal” life is not so bad and that even a disabled life can be satisfying and meaningful. However, it is not the value of normal life that is being questioned here but the omission to develop its full potential. Most importantly, enhancements do not guarantee success or happiness; they do not determine personality, identity, or fate. Therefore, they do not threaten diversity or authenticity. Enhancements should be understood as a means for ensuring the best possible start, for removing impediments to our freedom, and for providing the enhanced with more opportunities for a better, more meaningful life. They are merely dispositions that can be developed in various ways – leaving values such as personal achievements (worthy of praise or blame) intact.

#### 4. Moving beyond Pro et Contra Exchange

This overview of conceptual and ethical issues in the enhancement debate presents some of the most recurring challenges raised by the very idea of human enhancement, but it also captures the state of the debate itself. By the state of the debate, I here understand how this academic discourse has been carried out. On the one hand, there are issues with the discussion, such as extreme polarization of the debate, lack of dialogue, vague rhetoric, lack of mutual consideration, etc. For example, there is no shared concept of human enhancement, and participants are exchanging monologues instead of a constructive dialogue. On the other hand, there are dialectical issues, such as the lack of argumentative depth and philosophical soundness, messy criteria, underdeveloped and misconceived arguments. Although pro et contra exchange about the dangers and benefits of HET is a valuable tool, we might need a systematic attempt to move beyond such discourse.

A closer look at the pro et contra exchange reveals certain asymmetry in strength and plausibility, and by extension, the relevance of particular concerns. For instance, we first saw that the harshest criticism of biomedical enhancement seems to emerge from misconceiving some of the basic concepts and ideas. Once the main concepts closely related to the enhancement enterprise are clarified,

---

<sup>38</sup> This position may strike some as imposing on patient autonomy because there are cases when people do not want enhancements. It is a standard criticism of the enhancement communities that they engage in privileged ideals that are insensitive to actual human concerns (see Christensen and Leigh (2002)). I believe the moral imperative Savulescu is talking about is on making these technologies available to those who might want them.

many of the aforementioned problems become less worrying. For example, the concern about enhancements causing irreparable damage to human nature and biology depends on a fixed understanding of the notions in question, as well as what enhancements would, in fact, amount to. It can be argued that not all changes are necessarily bad and many aspects of human nature would be worth improving. Second, given that human “design” is currently far from optimal, improving some of our maladjusted biological traits need not be interpreted as a desire for perfection or mastery. Third, enhancements could reduce biological barriers and inequalities caused by the natural and social lottery. Fourth, they can hardly introduce absolute control over our lives (assuming that would be a bad thing), let alone eliminate all their unpredictability and mystery. I find most anti-enhancement arguments previously discussed as not particularly compelling or pointing to easily resolvable issues. Although both sides in the discussion have room for improvement, it is my impression that the criticism (anti-enhancement position) has been prevailingly vague and unconstructive.<sup>39</sup>

This is not to say that pro-enhancement side of the discussion is without its weaknesses. Although pro-enhancement authors often make effort to provide more structure to their arguments, much is still left to speculation. For instance, there are many further conceptual challenges within responses to the raised concerns. The appeal to biological definition of “human nature” seems to smuggle in a lot of assumptions about what is normal, what is necessary for health, etc. Many responses also trade on an ambiguity over what is natural: is manipulating the natural order itself something natural or something artificial? Some views come across as overly simplistic or overly confident in scientific progress – not taking important ethical and empirical limitations seriously. The attitude some pro-enhancement authors take can appear insensitive to particular communities or values. Important example are discussions on disabilities that especially challenge strong pro-enhancement positions about moral obligation or imperative to enhance. Namely, the debates in disability studies over treatments like cochlear implants, suggest that not only do people object to treatment sometimes, they further object to the characterization of deaf people as being in need of treatment at all (Christiansen and Leigh, 2002; Campbell, 2009). It could be, therefore, said that

---

<sup>39</sup> This, of course, applies to the arguments presented in this article. Clearly, there are many other concerns and responses, as well as responses to responses one could consider, and potentially gain a different impression.

enhancement advocates engage in privileged ideals that are insensitive to actual human concerns.<sup>40</sup>

Given that concerns raised in the debate vary in terms of argumentative strength and objective, they will require tailor-made (case-by-case) solutions rather than a “one-size-fits-all” policy. If we accept the *normalization* of enhancement – meaning that at a fundamental normative level, there is nothing special about them – we have to accept they should be evaluated using the same criteria we employ in other areas of practical ethics (Bostrom and Savulescu, 2009: 4). Furthermore, we can notice that the concerns previously described as less plausible or underdeveloped typically target substantially different aspects than the concerns described as more compelling. Namely, most bioconservative concerns target some intrinsic features of human enhancement technologies, such as their very nature. Other concerns target extrinsic or contingent features of enhancements, such as the potential to undermine societal values like fairness or equity.

So, even if the debate on HET has not been especially satisfactory in terms of resolving pressing issues, its examination provides specific methodological guidance – i.e., it is indicative of important (moral) standards that HET will have to satisfy. I believe we can use these indications to improve the quality of the debate and resolve some of the most fundamental questions. All of the mentioned hindrances – such as polarization, lack of dialogue, underdeveloped arguments, variety of intertwined concerns – point to a need for a more systematic and transparent evaluation. In the next section, I propose a set of methodological guidelines that can help us move the debate in a more fruitful direction.

## 5. Towards a Systematic Assessment of HET

Various concerns about HET discussed in this paper and broader seem to indicate certain conditions enhancements would need to satisfy to become permissible. It is my view, that there are three broad categories we should pay special attention to: (i) plausible coherence, feasibility, and effectiveness of a human enhancement project; (ii) absence of conflict with fundamental moral values and norms; and (iii) compatibility with or facilitation of socio-political goals of equality and justice. I

---

<sup>40</sup> I am thankful to the anonymous reviewer for these suggestions. It is also worth noting that the concern here is about general permissibility of HET or are there contexts in which at least some enhancements are permissible.

believe these aspects should guide our examination of HET and jointly they may provide valuable input about their permissibility.

Firstly, we need to determine whether a particular human enhancement project is internally consistent (in terms of theoretical/conceptual soundness) and whether the potential problems ensuing from a close theoretical examination are relevant for determining its moral permissibility. Take, for example, the previously mentioned concern about the intrinsic wrongness of HET. Namely, concerns that target intrinsic features of HET (such as their very nature) suggest that there is something fundamentally wrong with enhancements and no amount of extrinsic value would morally justify the use of such technologies. If such concerns are warranted, it seems that enhancements are morally impermissible regardless of their effects, and empirical questions such as risk or safety become irrelevant in accounting for their moral permissibility. Hence, my view is that we need to clearly distinguish between intrinsic and contingent concerns and thoroughly investigate whether there is something intrinsically wrong with enhancements. Further on, it is relevant to establish how intrinsic value affects moral permissibility, i.e., are intrinsically bad things always morally impermissible. With this approach, we will clear the way for a more fruitful discussion focused either on evaluating the contingent or intrinsic value of HET.<sup>41</sup>

Another way to tackle consistency, plausibility, and feasibility concerns is to look into relevant theories and examine whether certain assumptions are warranted. The above concern about the intrinsic wrongness of HET would then require looking into philosophical theories of intrinsic value and the implications this may have for the moral permissibility of HET. We could also take a specific human enhancement proposal and examine its internal soundness or plausibility on a more particular level. For example, moral enhancement is often said to be self-defeating and unrealizable because improving morality by artificial means could defeat the purpose of morality itself, not to mention the risk and safety concerns. In such a case, we need to examine how the very proposal, say moral bioenhancement via emotion modulation, aligns with the evolutionary psychology of emotions or neuroscientific research in general, which are the perspectives adopted by moral enhancement proponents.<sup>42</sup> This approach would also offer valuable input about the feasibility of such a project, at least in principle.

---

<sup>41</sup> I discuss this in more detail in Kudlek (2021), here Chapter 3.

<sup>42</sup> I discuss this in more detail in Kudlek (2019), here Chapter 4.

Feasibility alone, however, does not imply permissibility so, following the second condition, we will want to make sure that any human enhancement project is not in conflict with our fundamental moral values and norms. Since predominant ethical theories roughly reflect our moral norms and values, examining how enhancements correspond with them is a logical starting point. Normative ethics, in general, offers specific accounts of rightness and justifiability of particular acts and decisions, so it should be an adequate source for investigating moral permissibility of HET. For instance, moral permissibility of human enhancement will depend on the particular normative criteria that we decide to apply, such as the accordance with the utilitarian principle of maximizing utility or the conformity with deontological rules. Although such an approach will not offer conclusive reasons for or against human enhancement, it will provide valuable insight into whether enhancement is, and under what conditions, justified from the perspective of normative ethics.

Some attempts at providing normative input already exist in the debate, but they have certain limitations. One example is the previously elaborated *welfarist account* (see Section 2.2) by Kahane et al. (2011). Although the welfarist account is an inherently normative approach (instructing us to enhance when the intervention increases overall well-being), it does not offer clear criteria for discerning right or permissible action. We are left with an “empty” concept of well-being which we can “fill” in line with our theoretical preferences. Resolving whether a particular enhancement is permissible requires a less arbitrary approach and a more in-depth examination. This is why the discussion can benefit from a normative model that employs well-established stipulations for discerning permissible from impermissible actions. Examining whether HET present a threat to valuable aspects of human life involves zooming into particular applications of HET and inspecting whether they collide with things we find valuable, whether they can potentially cause harm in interactions with other things, and whether there are circumstances in which particular applications of HET can be morally permissible. The assessment of moral permissibility of HET from the perspective of predominant normative frameworks is a big leap towards untangling this debate and improving its quality.

This is not to say that utilitarians, deontologists, virtue ethicists, and others do not discuss concerns about HE from their perspectives, but this is rarely done explicitly or systematically. Hence, my suggestion is to focus on placing the current debate in a more explicit dialogue with ethical theories. This would involve the application of well-established ethical principles to specific instances of

human enhancement. Take, for example, utilitarianism. Enhancements are often associated with the consequentialist, especially utilitarian framework, in a way that suggests that they would be permissible insofar as they maximize the overall utility (Persson and Savulescu, 2019). But we have not yet examined whether enhancements would, in fact, align with utilitarian principles, would they be good for utilitarianism as a moral theory (in terms of promoting utilitarian ends), nor how would they affect utilitarian reasoning. A similar initiative has been taken by some scholars who mapped a number of consequentialist arguments in the human enhancement debate (Heinrichs and Stake, 2018), while others discussed the potential negative effects of enhancements for utilitarianism (Agar, 2015b). I strongly believe the debate would benefit from investigating more thoroughly aspects such as whether enhancements make us better utilitarian agents and whether enhancement *itself* is an optimific action (such that it yields greater benefits over drawbacks).<sup>43</sup>

Similarly, the fiercest objections against enhancements usually rest on intuitions typical for deontological doctrines (such as undermining freedom or autonomy). This often drives the conclusions about HET's impermissibility on deontological grounds.<sup>44</sup> However, these concerns are very nuanced and require a thorough investigation to understand whether enhancements are at odds with deontological principles. Recently, there have been several such attempts to provide a concrete evaluation of HET from a deontological, especially Kantian perspective. Interestingly enough, many of those indicate that Kantians would not necessarily be opposed to some enhancements, as we may have originally thought (see, e.g., Carter (2017); Bauer (2018); Clewis (2017); Hickey (2017)). Additionally, I believe we should pay special attention to concerns about human enhancement's inherent instrumentalism and whether it amounts to using ourselves as mere means. We should examine the problematic ways in which enhancements relate to one's autonomy/agency and outline deontologically permissible conditions or constraints on how and why we might engage in human enhancement.<sup>45</sup>

---

<sup>43</sup> I discuss this in more detail in Kudlek (2022a), here Chapter 5.

<sup>44</sup> For example, not only do people worry that enhancement will infringe on our autonomy, they also worry that "it would amount to using ourselves as mere means, that it violates human dignity, or that it does not respect our natures or identities" (Hickey, 2017: 165). Some find it difficult to reconcile enhancements' inherent instrumentalism with the ideals of moral autonomy and dignity, cherished in the Kantian outlook (Vedder, 2019).

<sup>45</sup> This is discussed in more detail in Kudlek and Smith (2022), here Chapter 6.



Finally, even if a particular human enhancement has sufficient normative support (i.e., if it does not contradict prevailing norms), its practical implications represent the next challenge. We need to investigate whether human enhancement would exacerbate existing problems of social equality and justice, or it could instead serve as a tool for reducing the negative effects of the natural and social lottery. Such and similar questions can be addressed with the help of well-established theories of social justice and equality. We will want to examine whether such enhancement would be in the individual's and society's interest. Would it provide grounds for exploitation and discrimination, or would it reduce social inequality and injustice? Evaluation of these and similar issues will depend on the specific implementation policies such as whether the enhancement is voluntary or involuntary, applied to children or adults, socially desirable or undesirable, constitutes criminal behavior, etc. This part of the evaluation will largely depend on empirical data that can only be provided by science in due time.

In summary, employing the proposed criteria for assessing the permissibility of HET can significantly improve the discussion. A closer examination of these aspects could result in plausible views about theoretical soundness and biotechnological feasibility, compatibility with fundamental moral norms, and compatibility with or facilitation of socio-political goals.

## 6. Conclusion

This article critically examined some of the most pertinent issues surrounding human enhancement in order to make progress on the stalled debate. First, it captured some aspects of the current debate by engaging various conceptual and ethical disagreements and reflecting on their cogency/plausibility. Second, it showed that the debate should move beyond pro et contra exchange by implementing a more systematic normative approach. As to the first aim, it appears that the debate generally lacks fruitful dialogue and careful deliberation of arguments. It was argued that most of these shortcomings apply more strongly to arguments against enhancement – i.e., criticism is prevalently unconstructive and vague. The second aim suggested that the debate could benefit from following a set of methodological guidelines that could help us make progress in future research. The conditions I find especially important to examine include: (i) plausible coherence, feasibility, and effectiveness of a human enhancement project; (ii) the

absence of conflict with fundamental moral values and norms; and (iii) compatibility with or facilitation of socio-political goals of equality and justice.





A. Karametas, *Democracy* (2015)

## Chapter 2

# Towards a Systematic Evaluation of Moral Bioenhancement

**Abstract:** The ongoing debate about moral bioenhancement (MBE) has been exceptionally stimulating, but it is defined by extreme polarization and lack of consensus about any relevant aspect of MBE. This article reviews the discussion on MBE, showing that a lack of consensus about enhancements' desirable features and the constant development of the debate calls for a more rigorous ethical analysis. I identify a list of factors that may be of crucial importance for illuminating the matters of moral permissibility in the MBE debate and which could help us move beyond the current lack of consensus. More precisely, I propose three important theoretical and normative standards that MBE should satisfy if we wish to mitigate the concerns about its utter impermissibility. Systematically assessing MBE interventions across the presented categories should provide valuable conclusions about its theoretical soundness and feasibility, its compatibility with fundamental moral norms, and its compatibility with or facilitation of socio-political goals of equality and justice. On the whole, MBE interventions should facilitate our understanding of what is or is not permissible.

This chapter is an expanded version of the article: **Kudlek K. 2022. Towards a systematic evaluation of moral bioenhancement. *Theoretical Medicine and Bioethics* [forthcoming].**



# Towards a Systematic Evaluation of Moral Bioenhancement

## 1. Introduction

Biomedical enhancements have been a subject of intense discussion over the past decade, with moral bioenhancement (MBE) perhaps causing even more scholarly disagreement than other forms of bioenhancement.<sup>46</sup> Given its controversial nature, MBE is susceptible to various interpretations. For example, views on MBE range from advocating it be given priority over other modes of human enhancement because its role could be crucial for the successful continuation of human life (e.g., Persson and Savulescu, 2008; 2012), to the claims it goes against the very nature of morality and it should be downright rejected (e.g., Harris, 2012, 2016; Agar, 2015a; Hauskeller, 2017). Thus, although the ongoing debate has been exceptionally stimulating, it is defined by extreme polarization and lack of consensus about what moral enhancement is, whether we want it, and whether we should allow it.

In this article, I review the discussion about MBE, showing that a lack of consensus about enhancements' desirable features and constant development of the debate call for a more rigorous ethical analysis. I identify a list of factors that may be of crucial importance for helping us understand what is or is not morally permissible, allowing us to move beyond the current lack of consensus. More precisely, I propose three important theoretical and normative standards that MBE should satisfy if we wish to, at least, mitigate the concerns about its utter impermissibility – if not even determine its moral status more clearly.<sup>47</sup>

These guidelines are influenced and shaped by the most pertinent concerns raised in the discussion because those concerns roughly indicate the requirements that an MBE proposal should ideally satisfy. These requirements include (i)

---

<sup>46</sup> For simplicity, I will henceforth use the term moral bioenhancement (MBE) to refer to a variety of biomedically induced moral enhancements, as opposed to non-biomedical moral enhancements such as upbringing, socialization, etc.).

<sup>47</sup> This work is an extension and application of the same methodological framework developed in Kudlek (2022b, *forthcoming*), here Chapter 1.

plausible coherence, feasibility and effectiveness of the enhancement project; (ii) the absence of conflict with fundamental moral values and norms; and (iii) compatibility with or facilitation of socio-political goals of equality and justice. This is not to say that meeting these requirements would guarantee the moral permissibility of MBE, but it would, at least, seriously challenge the common assumption about their outright impermissibility. Additionally, systematic examination and evaluation of an MBE project on these three levels should improve future discussions' quality by providing valuable insight into a particular MBE project's coherence, feasibility, and overall permissibility.

I proceed to summarize the current state of the debate by reviewing some of the most influential moral enhancement proposals (Sections 2-3; Section 5) and identifying three main categories of concerns raised explicitly in the context of MBE (Section 4). Finally, I introduce the above-described methodological framework and explain the purpose and application of the proposed requirements for a systematic evaluation of MBE's moral permissibility (Section 6).

## **2. The Need for Moral Bioenhancement**

Two articles kicked off the debate on MBE. One was by Persson and Savulescu (2008) and the other was by Douglas (2008). In a nutshell, Persson and Savulescu argued that MBE could ensure safe and ethical development and implementation of other human enhancement technologies, whereas Douglas used MBE to challenge the bioconservative view about moral impermissibility of *all* human enhancement technologies. I will now explain each view in greater detail.

### **2.1. The Imperative for Moral Enhancement**

Persson and Savulescu (2008) argued that human enhancements, such as cognitive enhancement, present a threat unless we develop effective ways of improving humanity's moral character. They elaborate the potential perils of “a further expansion of scientific and technological knowledge — let alone an acceleration of this expansion by novel means” — claiming that it “is problematic because we are already on the brink of acquiring — if we have not already acquired — knowledge which enables small groups, or even single individuals, to kill millions of us” (2008: 166). The especially problematic means of killing millions are weapons of mass destruction and our ability to deploy them (nuclear and biological weapons, terrorist attacks, misuse of supercomputers, environmental pollution, to name a



few). Prompted by such existential threats, which are imposed by both biomedical and traditional cognitive enhancements, the argument for the urgent imperative to morally enhance ourselves was developed.

Moral enhancement can be accomplished through traditional and biomedical means. Traditional and cultural means are socialization, education (formal and informal), and nurture – i.e., any “transmission of moral instruction and knowledge from earlier to subsequent generations” (Persson and Savulescu, 2008: 168). The moral progress that the human species achieved throughout history was accomplished by traditional, cultural means and, from where we stand today, remarkable progress was made in morally relevant achievements. For example, significant advancements were made in the recognition and promotion of fundamental human rights: abolition of slavery, prohibition of racism, equal rights for women, minority rights, freedom of speech, organized humanitarian aid, etc. Although these changes unquestionably indicate that humanity has made great strides in recognizing morally unacceptable behavior, the contemporary world still fails to put many of these findings into practice. The question, therefore, arises: is traditional moral enhancement effective (or effective enough)?

Persson and Savulescu argue that moral enhancement by traditional means is neither effective enough nor quick, especially compared to other traditional enhancement techniques (e.g., traditional cognitive enhancement). Although a certain level of moral progress has been achieved, “we do not necessarily do what is right and good as soon as we gain knowledge of what this is” (2008: 168). Another reason why moral progress, they believe, has been less impressive than the scientific one is that moral knowledge is harder to come by, not to mention its complex nature and the plurality of moral perspectives. Furthermore, in order for traditional moral enhancement to be effective, it is required that we are already morally motivated to a significant extent (2008: 168). Persson and Savulescu conclude that traditional means work very slowly, given how little moral progress there has been in the last 2,500 years. By adding dramatic scientific and technological progress to the equation, the threat to the survival of the human species becomes serious. Authors suggest that, under these circumstances, it seems reasonable to focus on developing biomedical and genetic means that “may be much more effective in terms of both how thoroughly and quickly they could improve everyone in need of improvement” (2008: 168).

## 2.2. The Permissibility of Moral Enhancement

Douglas (2008) invoked the example of MBE to oppose the view that biomedical enhancement is always morally impermissible. His paper focuses on the possibility of using biomedical technology to enhance ourselves morally. He notes that “there are various ways in which we could understand the suggestion that we morally enhance ourselves” –we could make ourselves more virtuous, more praiseworthy, more capable of moral responsibility, or we could simply bring ourselves to act or behave more morally (2008: 229). However, he understands moral enhancement differently. He suggests that we may cause ourselves to have morally better motives. Motives in this context should be understood as “psychological — mental or neural — states or processes that will, given the absence of opposing motives, cause a person to act” (2008: 229). Douglas also emphasizes that these enhancements do not necessarily amount to more moral persons, more moral characters or even that agents will *act* more morally. What he offers is a more neutral definition of moral enhancement:

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

Douglas’ account has several noteworthy features: it compares sets of motives, rather than individual motives; it focuses on “whether an alteration may reasonably be expected to result in the agent having morally better motives”; and it allows for moral enhancement to be achieved by non-biomedical means (2008: 230). Another essential characteristic of this proposal is the inapplicability of the standard objection against biomedical enhancements stating that enhancements would benefit the enhanced, but harm others. As Douglas explains: “On any plausible moral theory, a person’s having morally better motives will tend to be to the advantage of others” (2008: 230). He claims that there are no good objections to MBE, assuming it is performed under certain conditions.

Douglas believes that the following qualifies as moral enhancement (regardless of means being traditional or biomedical):

My thought is that there are some emotions — henceforth, the counter-moral emotions — whose attenuation would sometimes count as a moral enhancement regardless of which plausible moral and psychological theories one accepted. I have in mind those emotions which may interfere with all of the putative good motives (moral emotions, reasoning processes, and combinations thereof) and/or which are themselves uncontroversially bad

motives. Attenuating such emotions would plausibly leave a person with better future motives, taken in sum. (2008: 231)

Examples of such counter-moral emotions, the attenuation of which could count as moral enhancement, are (i) *a strong aversion to certain racial groups* and (ii) *the impulse towards violent aggression*. Although these emotions in most situations count as morally bad motives, there are rare circumstances in which reducing them would interfere with otherwise good motives. For this reason, Douglas ends up proposing a slightly weaker claim: “there are some emotions such that a reduction in the degree to which an agent experiences those emotions would, under some circumstances, constitute a moral enhancement” (2008: 231). Although he firmly believes that his account of moral enhancement should sometimes be permissible, Douglas is aware of possible obstacles to its realization within a medium-term time span: it is usually argued that some aspects of our moral psychology cannot be altered through biomedical interventions and that our understanding of the complex neuroscientific foundation of moral psychology is insufficient to allow such interventions (2008: 233). However, it is possible to respond that the two previously mentioned counter-moral emotions (racial aversion and violent aggression) belong to the more familiar domain of behavioral genetics and neuroscience, implying more than adequate understanding of their biological underpinnings. Douglas concludes on an optimistic note: “Given this progress in neuroscience, it does not seem unreasonable to suppose that moral enhancement technologies which operate on relatively simple emotional drives could be developed in the medium term” (2008: 233).

### 2.3. The Prospects of Moral Enhancement

Both Persson/Savulescu and Douglas argue that human morality has some biological underpinnings. On these grounds, they suggest that improving it should be within reach of biomedical and genetic treatment. In order to support their claims, they point to some relevant findings. Neuroscientific research and pharmacological and psychological experiments offer an invaluable contribution to understanding moral reasoning and moral behavior.<sup>48</sup> It therefore comes as no surprise that advocates of MBE often rely on the emerging science of morality when they argue for developing the means of enhancing moral dispositions. Such

---

<sup>48</sup> For a comprehensive overview and further references, see e.g. Persson and Savulescu (2012).

interventions' feasibility seems to be backed by some scientific findings showing that the manipulation of biological make-up can have morally desirable effects.

For example, the powerful hormone oxytocin, which works as a neurotransmitter in the human brain, has been shown to promote trust (e.g., Insel et al. 2004; Kosfeld et al. 2005).<sup>49</sup> Selective serotonin reuptake inhibitors (SSRIs) – commonly prescribed anti-depressants – have been shown to increase cooperation and reduce aggression (e.g., Tse and Bond, 2002; Wood et al., 2006). Methylphenidate – more familiar as Ritalin – seems to reduce violent aggression when given to persons with Attention Deficit Disorder (e.g., Ginsberg et al., 2013; Margari et al., 2014). The possibility of influencing human choices and behavior through the manipulation of biological traits like trust, aggression, and empathy can have morally relevant effects (Persson and Savulescu, 2012). Also, some personality disorders standardly associated to immoral behavior have been linked to certain biological bases, suggesting that if we were to arrive at a better understanding of these conditions, interventions might be developed to improve behavior (Persson and Savulescu, 2008; 2012; Douglas, 2008).<sup>50</sup> Proponents of MBE find it reasonable, therefore, to assume that biomedical means can and should be used to influence human choices and moral decision-making even in healthy individuals.<sup>51</sup>

To summarize, advocates of MBE believe that there is an urgent need to develop and implement biomedical enhancers of human morality. They hold it has the potential to mitigate many pressing issues of the modern world. There are reasons to believe such enhancement could be achievable via sophisticated scientific methods and overall morally permissible. A more detailed account of MBE followed and it quickly became the subject of a heated bioethical discussion. I will now look more closely into that proposal.

---

<sup>49</sup> It is important to note here that some further research showed that oxytocin could also reduce pro-social behavior towards out-group individuals (e.g., de Dreu et al. 2010), and the original oxytocin research does not seem to replicate and may be attributable to file drawer effect (e.g., Lane et al. 2016; Nave et al. 2015).

<sup>50</sup> In addition to these examples, there are other, non-pharmacological variants of MBE such as transcranial-magnetic stimulation, deep brain stimulation, and genetic engineering. Given that, as of yet, such options are in a highly speculative phase, I will mostly focus on pharmacological possibilities of MBE. For a systematic overview and further references, see Persson and Savulescu (2008; 2012); Earp et al. (2018).

<sup>51</sup> This only suggests there are biomedical prospects for developing moral enhancers. It is not to claim that any efficient techniques already exist or will become available soon.

### 3. Unfit for the Future

Persson and Savulescu ground the urgent need for MBE on two main reasons. First, they make the case that increased scientific and technological progress can be dangerous for the survival of the human species. Second, they argue that our moral psychology is maladapted to our modern environment (which places us at an as-of-yet unprecedented risk of catastrophe). As a result, MBE should be purposed to compensate for the discrepancy between scientific and technological progress achieved by the human species over the past few centuries and the allegedly lagging progress in our species-typical moral psychology. Human moral psychology has been adapted to the living conditions of the distant past – such as small societies based on kinship and direct reciprocity, where only primitive technology was used. Nowadays, most people live in large communities with access to advanced technologies, enabling them to exercise influence over great distances and far into the future (2012). Persson and Savulescu argue that human beings are not currently equipped with a moral psychology that would empower them to adequately cope with the moral problems that these new conditions create – e.g., threats from weapons of mass destruction and climate change and environmental degradation:

We hypothesized, based on evidence from evolutionary biology and psychology, that the moral psychology of humans is adapted to the former conditions, which have obtained for most of the time the human species has existed. This mismatch is a serious matter because humans now have at their disposal technology so powerful that it could bring about the destruction of the whole planet if misused. (Persson and Savulescu, 2012: 1)

These drawbacks of our moral psychology are manifested in certain behavioral tendencies and biases that do not seem to be a good fit for the moral requirements of the modern world. For example, it seems that we are disposed to care much more about what happens in the near future and to those near and dear to us, than about the suffering of distant, unknown individuals and larger collectives (Persson and Savulescu, 2015a: 338). We are often inclined to morally undesirable behavior such as nepotism, xenophobia, and groupishness.<sup>52</sup> Psychological myopia also causes inadequate responses in the context of care for future people, the

---

<sup>52</sup> The term ‘groupishness’ originates from Jonathan Haidt’s book *The righteous mind: why good people are divided by politics and religion* (2012). This stands for the psychological phenomenon of group or social selfishness – and it manifests in selfish motivations such as loyalty, which promote survival and interests of the group one belongs to.

environment, and non-human animals. On a more general level, one may conclude that outdated moral psychology mostly manifests as the absence of sufficient and adequate moral motivation. Given that such moral shortcomings present serious threats to the survival of the human species, proponents of MBE offer an “artificial update” or “evolutionary short-cut” to our moral psychology in order to make us better “morally adapted” to our current life conditions.

As I mentioned earlier, Persson and Savulescu argue that our moral dispositions have biological foundations and are shared with many non-human animals. Moral dispositions that constitute the core of morality are the dispositions to *altruism* (to sympathize with other beings) and to a *sense of justice* or *fairness*. It is assumed that these dispositions originate from typical reciprocally-based behavioural patterns, the so-called tit-for-tat emotions:

[T]here is a set of dispositions from which the sense of justice or fairness originates. The most basic of these dispositions are, we believe, the ones that have been called ‘tit-for-tat’. Evolutionary theorists have found that collectives in which this pattern of reciprocal reactions is widespread are most successful in terms of survival and reproduction. (Persson and Savulescu, 2008: 169)

If these core moral dispositions are grounded in our biology (and we imagine that biomedicine has advanced to a sufficiently sophisticated state) then they could be modified by biomedical means.<sup>53</sup> More precisely, biomedical means could be used for reinforcing our altruism and making us more just or fair. Authors stress that altruism and tit-for-tat emotions need to be appropriately tuned to be maximally useful because: “more altruism is likely to initiate more tit-for-tat exchanges, though too much altruism may be an obstacle by making us turn the other cheek when tit-for-tat requires retaliation” (Persson and Savulescu, 2008: 168-9). Persson and Savulescu firmly believe their proposal is in principle feasible and present a reasonable solution to the problems described earlier: “if you increase the altruistic motivation of people, you decrease the risk that they will negligently fail to consider the possible harmful effects of their behaviour on their fellow-beings” (2008: 172). However, the practical dimension of MBE seems to be more complex and challenging.

---

<sup>53</sup> Just because something is grounded in biology does not necessarily mean it can be modified by biomedical means. The intention here is to say that, all things considered, this may, in principle, be possible for MBE.

As Persson and Savulescu point out (2012), human beings will not by themselves spring into action necessary to avoid global collapse. Traditional moral enhancement and enhanced powers of reason are tremendously important, but they are insufficient when the span of necessary concern is rapidly expanding:

Education or instruction about what is morally good is not sufficient for moral enhancement because to be morally good involves not just knowing what is good, but also being so strongly motivated to do it that this overpowers selfish, nepotistic, xenophobic, etc. biases and impulses. (...) The fact that being moral is not just a matter of possessing some knowledge is (...) the reason why the big chasm between our moral and technological capacity has opened up. (2012: 117)

MBE seems to be largely a matter of motivating ourselves to do what we already know to be right – of overcoming our moral weakness of will – but the question is how to ensure a wise and proper application of MBE techniques (2012: 123). Although this is an extremely relevant question, many other ethical and empirical concerns have emerged in the debate. In the following section, I identify three broad sets of concerns commonly raised in the context of MBE.

#### **4. Concerns Regarding Moral Bioenhancement**

More than a decade later, there is virtually no consensus about any aspect of enhancing morality by biomedical means. Difficulties include the disagreement about what counts as MBE (what constitutes morality and what it means to act morally) and the ethical assessment of permissibility and desirability of such interventions. Over time, the absence of consensus gave rise to various approaches to MBE, but the common denominator among critics is a significant amount of scepticism about its prospects. In this section, I present and categorize the most persisting sets of concerns in the MBE debate. I identify three broad categories. First, some worry that MBE is conceptually unsound or unfeasible, and that, even if it proves to be feasible, it may have undesirable effects. Second, some worry that even if MBE proves to be a coherent and feasible idea, it will still be morally impermissible because it threatens fundamental moral values like freedom and autonomy. Third, some worry that MBE will exacerbate existing socio-political

problems such as inequality and injustice or cause new ones such as the attainment of higher moral status by the enhanced individuals.<sup>54</sup>

#### 4.1. Conceptuality and Feasibility Concerns

Like the broader discussion on human enhancement, there is a wide spectrum of approaches and enhancement categories regarding the types of capacities and the kinds of improvement in question. This diversity makes concerns of conceptuality and feasibility the most extensive category. Approaches diverge on what and how to enhance, what it means to enhance morally, and how to ensure wise and proper application. It is possible to distinguish between several types of capacities usually associated with improving moral judgment and behavior. On the one hand, we can try to improve affect, emotion or motivation (Buchanan, 2011a). On the other hand, MBEs can be generally designated as “interventions that are intended to improve our moral capacities such as our capacities for sympathy and fairness” (DeGrazia, 2014: 361). DeGrazia also distinguishes between three relevant kinds of improvement: (i) motivational, (ii) cognitive and (iii) behavioral. Motivational improvement concerns having better motives, character traits, and overall motivation to do what is right. Improved insight refers to a better understanding of what is right (when making a decision). Behavioral improvement should result in greater conformity to appropriate moral norms and, therefore, in more reliably carrying out the right actions (DeGrazia, 2014: 362-3).

Similarly, Karim Jebari distinguishes between behavioral, emotional, and dispositional biomedical moral enhancement (2014). While behavioral enhancement might be the least appealing of the three, due to its intrusive nature in restricting certain acts, Jebari argues that a combination of emotional and dispositional enhancement is the best option we currently have at our disposal. Emotional enhancement, as described by Douglas and Persson/Savulescu, directly changes the way we feel about specific behaviors (our emotions), whereas dispositional enhancement aims at reinforcing dispositions such as empathy and a sense of fairness; these dispositions ground the propensity to respond in a socially appropriate

---

<sup>54</sup> This is not to suggest that these levels exhaust all possible ethical implications of the MBE project. However, they seem to adequately cover a wide range of pertinent issues raised in the MBE debate, and thus require closer examination.



way in a specific context (Jebari, 2014: 256). This variety of approaches indicates how challenging it is to determine the conceptual and theoretical features of MBE.

The plurality of approaches prompted some scholars to deal strictly with defining and conceptually analyzing MBE. Raus et al. (2014) developed a taxonomy of existing definitions and uses of the concept of MBE. They showed that we could differentiate according to broad or specific interventional means; according to whether the focus is on the individual or the intervention itself; whether the intervention in question represents moral treatment or MBE; whether the intervention is oriented towards capacities or behavior, etc. They conclude that “[w]hile some differences, have relatively limited implications, perhaps simply in skewing the normative debate, other differences have more far-reaching implications and are revealing of underlying normative theories or positions” (Raus et al. 2014: 271).

But what does it truly mean to “morally enhance”? Anna Pacholczyk (2011) tackled the meaning of “moral” in “moral enhancement”. She suggested that it could be understood in three different ways: (i) moral enhancement as an ethically desirable enhancement of any capacity, i.e., any kind of enhancement that will (other things being equal) result in a better world; (ii) moral enhancement as a change in some aspect of morality that results in a morally better person, and (iii) moral enhancement as a beneficial change in the sphere of morality, i.e., the meaning of the term moral can be descriptive and refer to a certain aspect of human cognition (2011: 252-4). The second understanding is closest to what Persson and Savulescu had in mind, since it includes “making people more likely to act on their moral beliefs, improving their reflective and reasoning abilities as applied to moral issues, increasing their ability to be compassionate, and so on” (Pacholczyk, 2011: 253).

Birgit Beck (2014) also pointed to several conceptual puzzles regarding the meaning, prospects, and ethical evaluation of MBE. She argued, mostly referring to Douglas’ definition of ME, that “there are competing concepts of morality and moral conduct which complicate a definition of moral enhancement and the assessment of suitable target features for moral enhancement measures to be applied” (2014: 236). Beck points out that there is little chance, at least in terms of the near future, for coming to an agreement on the proper understanding of morality, not to mention the questionable appropriateness of a single ethical theory. She believes that rival (meta-)ethical theories have different views on what features should be targeted through enhancement; if we want to enhance moral conduct, we need to provide a unanimous answer (2014: 234).

Depending on the preferred approach to MBE, it will be possible to differentiate between several kinds of MBE in practice. For example, one can categorize MBE in terms of the capacities that are being enhanced (such as affect, emotions, motivation, cognition, dispositions, etc.) or in terms of the techniques/interventions being used to arrive at MBE (such as pharmaceuticals, deep brain stimulation, genetic selection or engineering). Even if we manage to agree on the content of MBE, other difficulties of a practical nature will arise. For example, who would be responsible for making decisions about MBE? Should MBE be voluntary or mandatory? Should we target only certain groups – such as convicted criminals, morally corrupt individuals, and the mentally ill – or should we apply it widely to children, like vaccination? What would be the advantages and disadvantages?

Some believe that the absence of agreement on such a fundamental level indicates that MBE is not very likely to be made sense of or realized in the medium-term future (Beck, 2014). When the discussion turns in this rather sceptical direction, it often brings about skepticism in other ways: Are we in *real* need of MBE? Isn't there enough evidence that humans are moral? Who is likely to profit from MBE and has an interest in its implementation? It is not readily conceivable that people would be eager to give biomedical means a try (2014: 238). However, while some do not see the prospect or need for MBE for the reasons given, others believe it is possible to reach an overlapping consensus about moral matters and to develop many kinds of effective MBEs (Shook, 2012). Even if we are optimistic about the prospects of MBE, in terms of reaching an agreement on what to enhance and how, it remains to examine the moral and ethical permissibility of such interventions.

#### 4.2. A Threat to Moral Values

If MBE turned out to be coherent and biotechnologically feasible, many concerns regarding its effects on valuable aspects of human life would be raised. Therefore, the most challenging philosophical question in the MBE debate is probably: Even if we had sufficient scientific knowledge and understanding of biology, neuroscience, and moral psychology to engage in this project, would MBE endanger some of the most valuable aspects of human life? It is often suggested that MBE will interfere with our freedom, identity, and autonomy – by making our acts morally desirable, but automatic and unwitting. Many believe that any compulsion (or absence of choice) in the moral domain directly undermines the praiseworthiness and blameworthiness of acts, as well as the agent's moral responsibility. Along

these lines, Fabrice Jotterand (2011) argues that moral neuro-enhancement is not possible because we can only become better, morally improved people through the careful, reflective exercise of our moral agency, not through neural manipulation. Another problematic aspect in this context is the fear of misfiring or causing moral decline. What if MBE has unintended side-effects – like destroying psychological tendencies that are actually good for us? What about diminishing the diversity of moral positions and eliminating healthy disagreement that is essential for moral progress?

Some scholars challenge the assumption that MBE necessarily means making people morally better. For example, Pacholczyk (2011) points to two constitutive parts of this idea: the factual and the normative claim. The factual claim refers to the notion that the enhancement in question in some way affects the moral sphere, while the normative claim is about whether the intervention creates a morally better person. The author claims that the structure of our discussion depends crucially on whether we take a combined factual-normative claim or only a factual (non-normative) one. In her opinion, this distinction makes the discussion clearer; besides, discussing MBE from a strictly non-normative point of view can be interesting and fruitful. Paula Casal (2015) points out that moral enhancers might induce many benefits to be produced for us, but they do not provide us with the truth about what is morally right. If moral enhancers only bring us to conform to the duties we think we have, things can go wrong (e.g., religious fanatics and terrorists often strongly believe their duties are moral). Another point made by the author concerns greater altruism and empathy as the means of achieving moral progress. She notes that “[d]oing the right thing, however, is not always a matter of greater altruism or empathy” (2015: 340) (e.g., terrorists and suicide bombers can be altruistic and empathetic in devoting their lives to their causes).

The example of the radical terrorist indicates that we can often judge the same action as moral or immoral, depending on the interpretation or moral system we adhere to (Hardcastle, 2018). Hardcastle emphasizes that lone wolf terrorists look to their communities to define what constitutes the good; they believe that they are pursuing laudable goals. Humans use their reference groups to mirror good and bad choices (2018: 282). By enhancing moral motivation to do what *we* believe is right can misfire in the case of a morally corrupt individual such as a terrorist. “Whose morality – and the “facts” that underlie that morality – we should adopt are fundamental questions that must be answered before we start any sort of moral enhancement project.” (Hardcastle, 2018: 284) Deep and fundamental clashes of moral perspectives are often drivers of cultural violence. So, the point

here is that: “Whatever individual moral enhancing (or clinical treatment) goes on must be in combination with changes to the social and cultural environment of the individual as well” (ibid.: 287), otherwise any attempt to morally enhance will miss its mark.

John Shook similarly argued that, unlike in the case of cognitive enhancement, classifying MBE must take into account environing social contexts: “moral intuitions, virtues, and rules are not identical around the world; changing the social context can switch a classification of a moral enhancement into a moral deficit” (2012: 3-4). However, he does not hold that this implies the reign of relativism or complete skepticism regarding MBE, and warns that *moral* enhancement should not be confused with *ethical* enhancement: “Modifying conduct in line with what people already regard as ordinary moral behavior can never replace thoughtful adjudication among conflicting moral duties or adjusting our social norms for improving the human condition” (2012: 12). Finally, Thom Brooks claims that MBE may undermine the existence of reasonable pluralism in modern liberal societies: “The question is not only whether moral enhancement might lead to only one moral judgment, but also whether moral enhancement might benefit some reasonable moral, philosophical, and religious doctrines over others” (2012: 29).

A multitude of concerns outlined here indicates an apparent lack of theoretical and normative foundations in the MBE debate. There is a tendency to believe that MBE will undermine various moral values, but it is not clear what values MBE would improve if applied, nor is there a consensus on what values ought to be improved. It is questionable whether a unanimous answer can be offered to this question since any answer by itself may threaten a healthy diversity. The issue with different individual standards of morality is gradually transcribed into a broader discussion about social norms and, by extension, to the impacts of MBE on a socio-political level.

### 4.3. Socio-Political Concerns

The third set of concerns revolves around the social and political effects that the implementation of MBE could bring about. The standard questions raised in this area include: Would enhanced individuals and societies acquire some kind of higher moral status than the unenhanced ones? Would this provide grounds for elitism, discrimination, domination, exploitation, or other undesirable tendencies? Could this exacerbate the already existing problems of inequality and injustice? Some even infer that MBE could have posthuman effects (similarly as in the

case of human enhancement). The previously raised question: Who would benefit from MBE – an individual or a society as a whole? – is also a vital part of this discussion. Political philosophers frequently address such concerns. For example, Robert Sparrow worries that the science of bioenhancement might lead to arbitrary inequalities in access to political power and would, in effect, facilitate the “unjust rule of authoritarians” (2014b: 20). He also considers the risk that MBE represents in terms of “reinvigorating dangerous ideas about the extent of natural inequality in the possession of the moral faculties” (2014b: 20). Several authors considered the possibility of abusing MBE. For example, Specker et al. (2014) discuss the idea that MBE might induce free-riding, like in a prisoner’s dilemma (i.e. the virtuous could become exposed to the exploitation by the vicious). This problem might also appear on the broader level in the form of domination or exploitation of enhanced societies by the unenhanced ones (Triviño, 2013).

A matter closely related to that of social equality concerns social justice. Some worry that MBE might not only foster abuse but also exacerbate injustice. For example, Tom Beauchamp is concerned “whether the use of bioenhancements, especially by officials in political states, would create a more just world, or rather a less just world” (2015: 347). He believes in the possibility of MBE exacerbating, rather than diminishing, existing social prejudices and distributive unfairness (2015: 347). It is often assumed that morally enhanced individuals would acquire some kind of higher moral status, which would then spiral into elitism and discrimination. For example, Alfred Archer argues that “[t]he widespread use of moral enhancement would raise the standards for praise and blame worthiness, making it much harder for the unenhanced to perform praiseworthy actions or avoid performing blameworthy actions” (2016: 501). If this were the case, the morally unenhanced would be harmed because it would be “more difficult for [them] to perform morally praiseworthy acts or to avoid performing blameworthy acts. This is because there is a good reason to think that what people are praise or blame worthy for is at least partially determined by the level of moral behavior that most people manage to obtain.” (2016: 502) It seems that, in order for MBE to be successful and abuse-free, it would have to be undertaken collectively – applied on a global level. It is, however, doubtful that people will be willing to expose themselves to such interventions voluntarily on a massive scale. If the level of participation in society is inadequate (people are not willing to undergo MBE procedure voluntarily), we will face the problem of compulsory MBE, which would restrict or completely obliterate individual freedom (Glannon, 2018).

In summary, MBE evokes various concerns on conceptual, normative and socio-political levels, and from various professional, individual, and societal perspectives. These challenges range from determining what and how to enhance to ensuring proper and harmless application. There is no consensus about what it means to enhance morally or which standard of morality we should adhere to. Will MBE interventions threaten valuable aspects of human life? Is it likely that problems of social inequality and injustice will be exacerbated by MBE? These are some of the central questions in the discussion. Sadly, a satisfying or unanimous answer does not seem to be in sight. One attempt to respond to many of the listed concerns is the development of a more sophisticated account of MBE. In the following section, I present one type of MBE which (arguably) escapes many of these worries.

## 5. Response to Concerns: Moral Neuroenhancement?

Although we can use the term moral neuroenhancement (MNE) to cover a wide range of biomedical interventions, used to improve morally relevant capacities, MNE also refers to a particular account of moral betterment.<sup>55</sup> Proponents of this account rely strongly on neurotechnological findings and focus on a new set of tools that may foster moral enhancement. These tools are broadly described as “neurotechnologies”. Neurotechnologies are meant to directly alter brain states and neural functions in such a way as to bring about desired moral improvement (Earp et al., 2018: 166). All along, advances in neuroscience have been a promising ground for the development of enhancement technologies. There are many examples of neuroscientific research and findings that could be useful and relevant in “correcting” immoral behavior. These findings range from the administration of neurohormones such as oxytocin (which promotes pro-social attitudes like trust, sympathy, and generosity), to the manipulation of serotonin and testosterone levels (which can mitigate undue aggression, and increase fair-mindedness, willingness to cooperate, and the aversion to harming others). Brain stimulation devices have also proven useful in morally relevant areas by reducing impulsive tendencies in psychopaths, treating addiction, and improving self-control (Earp et al., 2018: 167). Advocates of MNE are fully aware that such measures might be ethically and conceptually controversial, as well as morally impermissible, but in their

---

<sup>55</sup> Please note that here, MNE refers to a more theoretically sophisticated account of MBE, but practically speaking, MNE and MBE belong to the same category of enhancements.

opinion, this should not exclude the possibility of, nevertheless, using some moral neuroenhancers that have already been developed (Earp et al., 2018: 167).

An *agential* conception of MNE may help us understand how moral betterment is to be achieved:

Moral neuroenhancement: Any change in a moral agent—effected or facilitated in some significant way by the application of a neurotechnology—that results, or is reasonably expected to result, in the agent being a morally better (i.e., more moral) agent. (Earp et al., 2018: 168)

They leave open the matter of what should count as moral “betterment”, indicating that this list might include various examples, such as “increased moral worth or praiseworthiness of the agent, increased moral excellence of the agent or increased moral desirability of the agent’s character traits” (Earp et al., 2018: 168). Earp et al. complement the MNE account by developing the *functional-augmentative approach* to enhancement.<sup>56</sup> Since they are aware that having “more” of some morally relevant function or capacity does not always constitute betterment, they argue that a morally adept agent should be able to “respond flexibly to different situations, and to employ or tap into different cognitive and emotional resources as necessary to arrive at the motives, decisions, and behaviors that are morally desirable given the context” (2018: 169). A narrow focus on boosting specific moral capacities will not do the job entirely (e.g., increased empathy can lead us astray when it comes to making certain moral judgments), so in order to produce better moral agents, we must distinguish between lower- and higher-order moral capacities. For example, *empathy* or a *sense of fairness* represent lower-order capacities – basic features of human psychology relevant for moral motivation and behaviour. Second-order capacities are in fact *abilities* to know or identify when it is morally desirable to feel and act upon lower-order capacities:

So it wouldn’t be just “more empathy” (tout court) that would be expected to lead to the improvement of a moral agent, qua moral agent, but rather an increase in what might roughly be described as a kind of second-order empathic control – an ability to (1) know or to identify, whether consciously or unconsciously, when it is morally desirable to feel empathy and/or allow it to shape one’s outward behavior (and in what way), as well as (2) to be

---

<sup>56</sup> This approach aims to show that “[i]nterventions are considered enhancements ... insofar as they [augment] some capacity or function (such as cognition, vision, hearing, alertness) by *increasing the ability of the function to do what it normally does*” (2018: 168).

able to feel such empathy, or if necessary, suppress such feelings (or their effects on behavior), in accordance with (1). (Earp et al., 2018: 170)

In other words, enhancement of higher-order capacities would involve reflective, flexible, reasonable, and context-dependent modulation of moral responses and for this reason, it would be more reliable than the previously described approach of “blindly” increasing or diminishing specific moral capacities. The idea of higher-order MNE can be contrasted with Owen Schaefer’s distinction between direct and indirect moral enhancement (2015). Schaefer argued that direct moral enhancement is problematic because it aims at bringing about particular ideas, motives or behavior – making people commit to them without much room for rational deliberation. Indirect moral enhancement would, on the other hand, aim at enabling people to produce morally correct ideas, motives, and behaviors without having them commit to their content (2015: 261). MNE resembles indirect moral enhancement in many aspects and it seems to be a solid *prima facie* case.

Ideally, indirect moral enhancement (or similar techniques of MNE) could be resilient to many objections standardly raised against direct forms of moral enhancement, such as restriction of freedom or fear of unintended bad consequences. If indirect moral enhancement would, as Schaefer anticipates, improve the reasoning process itself, without committing people to adopt specific beliefs, this would eliminate the threat to healthy disagreement essential for moral progress. It would also avoid potential problems of insufficient moral understanding (raised in the case of direct moral enhancement), as well as the problem of making morality “too easy” or “disconnected” from the world (Harris, 2011; 2012). If indirect means would leave enough room for reflective reasoning, effort, and engagement, this would also reduce the potential threat of direct moral enhancement to authenticity, autonomy, and rational decision-making:

[S]ince moral lessons, abilities, dispositions, etc., that are achieved or developed with the help of a neurotechnology—as opposed to directly caused by it (thereby preserving space for conscious reflection, effort, and engagement)—could be seen as posing less of a threat to such important issues as authenticity, autonomy, and rational deliberation... (Earp et al. 2018: 174-5)

Earp et al. envision “a *facilitating*, rather than *determining* role for any drug or neurotechnology” (2018: 175). This means that neuroenhancers should be administered as part of a richly contextualized process of moral learning, rather than in a vacuum.



Although MNE escapes many standard moral concerns, the overall state of the debate leaves some room for skepticism. Namely, a lack of consensus and the continuous evolution of the debate necessitate the devising of widely applicable methodological guidelines and evaluation criteria. For example, one may doubt whether sophisticated interventions such as MNE are at all feasible. This is, however, an empirical question that will be unequivocally answered by technology and science in due time. Apart from empirical realizability, there are also open philosophical questions about the conceptual soundness of this project and what should count as moral betterment. Along these lines, socio-political (or practical) implications of MNE (or any other proposal of enhancement) may also lack a normative assessment.<sup>57</sup>

These open concerns, together with lack of consensus and the debate's continuous evolution, necessitate the devising of widely applicable methodological guidelines and evaluation criteria. More precisely, the debate on MBE, as it stands, is insufficiently theoretically informed. It seems to be caught up in details of fictional scenarios and implementations and their outcomes, whereas there is still fundamental disagreement at the conceptual and normative level. Although these ongoing discussions are valuable in their own right, we may benefit from ascending to a more general level of ethical analysis and by setting out a methodological framework for a systematic assessment of moral permissibility.<sup>58</sup>

## 6. A Methodological Framework for the Assessment of Moral Enhancement(s)

The previously detailed sets of concerns represent the toughest challenges in the MBE debate and, on the whole, are the best reasons in favor of MBE's moral impermissibility. These concerns, however, may be the only common ground in the enhancement debate and roughly correspond to important theoretical and normative standards MBE should satisfy to, at least, *not* be considered downright unacceptable or impermissible. Accordingly, three general levels of requirements

---

<sup>57</sup> For example, one of the latest developments in the debate has been a proposal to shift from MBE to AI-based moral enhancement by introducing an "artificial moral advisor" (e.g., Giubilini and Savulescu, 2018; Lara and Deckers, 2020). Virtual reality could also play a role in enhancing human empathy (e.g., Rueda and Lara, 2020). I believe the conclusions reached in this paper would apply to such "digitized" versions of MBE, too.

<sup>58</sup> I have also developed and proposed this methodological intervention at the human enhancement level in Kudlek (2022, *forthcoming*) [Chapter 1] and here I want to make use of it on a more concrete example of moral enhancement.

prove to be especially important across the debate on the permissibility of MBE techniques:

- (i) plausible coherence, feasibility, and effectiveness of an MBE project;
- (ii) the absence of conflict with fundamental moral values and norms;
- (iii) compatibility with or facilitation of socio-political goals of equality and justice.

This is not to say that these requirements directly correspond to the above concerns or that satisfying these requirements would automatically guarantee moral permissibility. Instead, it means that exploring these conditions could give us a better understanding of what would or would not make MBE permissible.

In what follows, I explain the implementation of my approach in terms of ethical and theoretical analyses, and then give an example of how this approach can be applied, assuming that MBE becomes a possibility. The former also outlines why these criteria are needed, what questions they attempt to answer and how they are intended to be answered, and what contributions the investigation of these three perspectives would bring to the debate.

### **6.1. Ethical and Theoretical Analysis**

The first requirement entails determining whether MBE is (in some or all conceivable forms) a consistent and plausible proposal, and, in principle, feasible and effective. We want to answer here: Whether MBE would have the effects anticipated by its proponents? Which, in turn, requires asking: How does the proposal align with external but integrally linked theories (such as cognitive neuroscience, evolutionary psychology of emotions, etc.)? Are common assumptions made in the discussion warranted (such as that MBE is self-defeating)? In addition, this would also require addressing how potential problems ensuing from a close theoretical examination affect MBE's moral permissibility. Establishing theoretical soundness and internal consistency of a particular MBE proposal can be achieved by tools such as thorough conceptual analysis, comparative analysis, and the application of available and relevant scientific data. If, for example, the theoretical analysis indicates that MBE by emotion modulation would misfire because emotions cannot reliably guide human moral decision-making, then this would impact

our understanding of the proposal's coherence and effectiveness, as well as its moral permissibility.<sup>59</sup>

Second, in order to assess whether MBE conflicts with fundamental moral norms and values, we must investigate its relationship with predominant moral views. Since normative ethics offers specific accounts of rightness and justifiability of particular acts and decisions, it should be an adequate point of departure to answer questions such as: Does MBE present a genuine threat to fundamental norms and values? Is MBE harmful? Can it produce good consequences or, perhaps, will it undermine intrinsically valuable things such as freedom, autonomy, and moral responsibility? Illuminating the moral permissibility of MBE requires looking into particular accounts of moral rightness and permissibility if we want the assessment to be sufficiently theoretically informed.<sup>60</sup> Therefore, the moral permissibility of MBE can be assessed through its accordance with the utilitarian principle of maximizing utility or the conformity with deontological rules.<sup>61</sup>

For example, the ongoing debate leaves a general impression that MBE would likely be justified on consequentialist grounds – as far as it maximizes the overall utility (Persson and Savulescu, 2019) and that it would be intrinsically (or deontologically) wrong as far as it undermines our freedom, autonomy or responsibility (Harris, 2011). I believe these matters are worth digging into because they will, consequently, provide insight into whether MBE is, in fact, a threat to specific values and under which circumstances, if any, it could be acceptable.<sup>62</sup>

---

<sup>59</sup> I discuss the role of emotions in the MBE debate in more detail in Kudlek (2019), here Chapter 4.

<sup>60</sup> There have been attempts to discuss enhancements from different normative perspectives, but this is rarely done explicitly and systematically. For example, the welfarist account (Kahane et al., 2011) is an inherently normative approach instructing us to enhance whenever the intervention increases overall wellbeing, but it does not provide us with clear criteria or content for how to understand and evaluate wellbeing. These gaps can be filled by applying familiar normative models that give more straightforward instruction on discerning permissible from impermissible acts.

<sup>61</sup> This is not to say that we should base MBE on preferred ethical theories, but rather that we should use ethical theories to inform and illuminate matters of moral permissibility in the enhancement debate.

<sup>62</sup> Whether MBE is justified on utilitarian grounds is further discussed in Kudlek (2022a) [Chapter 5], and some conditions and constraints on how and why we might engage in MBE on Kantian grounds is further outlined in Kudlek and Smith (2022) [Chapter 6]. Some scholars have also mapped a number of consequentialist arguments in the human enhancement debate (Heinrichs and Stake, 2018), while others discussed the potential negative effects of moral

Finally, even if a particular MBE technique is morally justified (i.e., if it does not contradict prevailing norms), its social implications represent the next challenge. This requires examining the previously discussed puzzling questions and concerns that arise from MBE's socio-political impacts. Some of those include answering: Whether MBE would deepen social inequality and injustice, or whether it could reduce the negative effects of the natural and social lottery? Whether MBE would be in an individual's or in society's interest? Would it provide grounds for exploitation and discrimination? Methods for tackling these concerns, in terms of conceptual and theoretical clarifications, would be the application of well-established theories of social justice and equality. For example, the Rawlsian maximin principle would require that any redistribution of social goods, say enhancements, benefits those who are the worst off.<sup>63</sup> Hence, my third methodological guideline would provide a means for elucidating justice and equality matters in the enhancement debate and how they affect the overall moral permissibility of discussed technologies.

## 6.2. Application of the Criteria

Let me illustrate a preliminary application of the proposed criteria on a concrete, practical example.

Imagine Jane, who is inclined to become a lone wolf terrorist and intends to commit a terrorist act. However, Jane decides to engage in MBE by emotion modulation – where she will have her altruism (empathy) increased and her sense of justice strengthened. Now, according to the first criterium, assessing whether this concrete MBE intervention is coherent, feasible and effective will, in practice, depend upon whether it will have the desired effects. The desired effects would involve Jane: (i) recognizing that her intentions (and, consequently, actions) are morally impermissible and (ii) refraining from committing the act of terrorism. By contrast, undesirable effects would consist of MBE making Jane even more

---

enhancement for utilitarianism (Agar, 2015b). Deontological analysis of moral enhancement has also been receiving some attention, see, e.g., Carter (2017); but more has been said about human enhancement in this regard, (see, e.g., Bauer (2018); Clewis (2017); Hickey (2017)). Most of these studies indicate that enhancements would not necessarily undermine deontological values, as we may have originally thought.

<sup>63</sup> This approach mirrors the previously discussed application of normative theories to the matters of moral permissibility of MBE. After all, theories of social justice and equality are to some extent a subset of normative theories. I thank the anonymous reviewer for leading me onto this point.

devoted to her cause, which would present the backfiring problem and render MBE ineffective (as it would not have the anticipated effects).

The assessment of MBE's compatibility with fundamental moral values and norms (the second criterium) would, in practice, involve assessing whether Jane's actions, *following* the enhancement intervention, are compatible with a set of widely accepted and uncontroversial moral norms and values. For example, let us imagine that after engaging in MBE, Jane decides to refrain from committing the terroristic act she previously intended. Since harming others is morally wrong on almost any moral account, including commonsense morality, Jane's change of heart would be more than welcome. This is to say that Jane's engagement in MBE would be morally right from, for example, consequentialist and deontological viewpoints. However, deontologists and virtue ethicists could further question whether her actions are praiseworthy or whether MBE has undermined her moral responsibility.

Finally, assessing whether MBE is compatible with or facilitates socio-political goals of equality and justice would require assessing the broader socio-political effects of Jane's actions. Namely, Jane's decision to engage in MBE and, by extension, to refrain from terrorism, is obviously a socially desirable endeavor because it aligns with principles of respect for fundamental human rights (e.g., the right to life and to not be harmed). Also, by not committing this crime, Jane would avoid incarceration, which is desirable individually and socially. However, if Jane would not engage in MBE voluntarily or if she would be stigmatized for engaging in it, some concerns about social justice and equality could be raised. For example, her individual rights could be infringed if MBE is coerced or if she is discriminated against for undergoing the enhancement intervention.

At this point we are facing the question of *how* we will tackle various issues arising in Jane's example, which leads us back to the previously described methodological suggestions for applying ethical analyses to the bioethical debate on MBE. Namely, assessing MBE interventions across the three presented categories (and implementing the previously described methods) should provide conclusions about its theoretical soundness and feasibility, compatibility with fundamental moral norms, and compatibility with or facilitation of socio-political goals. On the whole, it should facilitate our understanding of what is or is not permissible.

## 7. Conclusion

This article narrowed down the list of factors that may be crucial for elucidating what may or may not be morally permissible in the debate on biomedical moral enhancement. To identify these relevant factors, I reviewed the debate and showed that it requires a more rigorous ethical analysis. I outlined some of the most influential proposals of MBE and reviewed their noteworthy features and the role they are supposed to play in handling the pressing issues of the contemporary world. Despite its potential, MBE evokes substantial ethical concerns that seem to overshadow almost every aspiration that this project will ever become tenable. Although different accounts of MBE, such as MNE, escape many standard moral concerns, a lack of consensus about these interventions' desirable features still prevails. Such a state of affairs necessitates clear and precise methodological guidelines. I proposed three important theoretical and normative standards that MBE should satisfy if we wish to mitigate the concerns about its utter impermissibility. These include (i) plausible coherence, feasibility, and effectiveness of the MBE project; (ii) absence of conflict with fundamental moral values and norms; and (iii) compatibility with or facilitation of socio-political goals of equality and justice. Exploring how MBE corresponds to these requirements will give us a better understanding of whether and under what conditions it is permissible, or at least what would make it permissible. It will also provide a theoretical and normative basis for a more fruitful discussion.



## The Ethical Analysis of Moral Bioenhancement



A. Karametas, *Egoism* (2014)



## Chapter 3

### Is Human Enhancement Intrinsically Bad?

**Abstract:** A pertinent concern in the human enhancement debate is that human enhancement technologies (HET) are intrinsically bad and, hence, morally impermissible. This article evaluates the claim about the intrinsic badness of HET by looking into philosophical theories of intrinsic value. It investigates how well-established conceptions of intrinsic value map onto typical bioconservative arguments about HET's intrinsic badness. Three predominant variants of these arguments are explored and found wanting: (i) HET are intrinsically bad owing to their unnaturalness; (ii) the pursuit of HET reveals intrinsically bad character (“the desire for mastery”); and (iii) HET will necessarily undermine intrinsically valuable things (e.g., human dignity). My analysis shows that the debate on intrinsic value places serious constraints on claims about the intrinsic badness of HET. More specifically, the analysis shows that bioconservative arguments are, for the most part, inconsistent, misconceived, and overly speculative. Enhancement interventions cannot be bearers of intrinsic value on any of its plausible understandings, and, even if we could grant such a possibility, there are no compelling reasons to presume that the intrinsic value of HET would be necessarily negative. As a result, claims regarding their moral impermissibility are unwarranted.

This chapter is published as: Kudlek, K. 2021. Is human enhancement intrinsically bad?, *Medicine, Health Care and Philosophy*, 24, 269-279. DOI: <https://doi.org/10.1007/s11019-021-10003-w>



# Is Human Enhancement Intrinsically Bad?

## 1. Introduction

Some authors in the human enhancement debate hold that human enhancement technologies (HET) are intrinsically bad and hence morally impermissible.<sup>64</sup> For example, prominent bioconservatives typically claim that we might agree about improvements, safety, the fairness of distribution etc., and still be morally hesitant about the permissibility of HET (Fukuyama, 2002; Kass, 2003; Sandel, 2004). This unease about biotechnologies points to something of ethical significance, pertaining to the essence of the activity itself (Kass, 2003). Namely, if objections suggesting the intrinsic badness of HET are plausible, empirical questions (such as risk and safety) would become less relevant in accounting for the moral permissibility of HET. Bioconservatives may take an even stronger stance by adopting the so-called *bioconservative thesis*, which states that if HET are intrinsically bad, such interventions should never be permitted, even if they are safe, reliable, and justly distributed.<sup>65</sup> The claims about HET's intrinsic badness, together with implications for their moral permissibility, are robust and influential – impacting the overall discussion and perception of these technologies. Therefore, they require a deeper examination.

Although bioconservatives explicitly claim that HET are wrong in a nonconsequentialist sense, they do not explicate this claim on a deeper philosophical level – such as by adhering to a theory of intrinsic value. It seems that their understanding of intrinsic badness is commonsensical, but much is left to speculation. In order to reduce speculation and improve the discussion, we need to apply more

---

<sup>64</sup> Although bioconservatives claim that human enhancement *technologies* are intrinsically bad, it seems odd to think of technologies like medical machines or pills as bad in themselves. Hence, I will assume that bioconservatives believe human enhancement as an *action*, brought about by biotechnological means, is intrinsically bad. Henceforth, I will use the terms HET and human enhancement as roughly amounting to the described intervention.

<sup>65</sup> Some scholars may think that this formulation of the bioconservative thesis does not represent bioconservative views accurately. Milder interpretations of their views include “a general plea for caution” or simply a suggestion that “implementing HET is unwise”, but with no repercussions for its moral permissibility (e.g., Hauskeller, 2013). I, however, believe that bioconservative views are too robust and influential to be taken lightly.

rigorous standards. My analysis proceeds in two distinct but interconnected steps. First, I examine the claim that HET are intrinsically bad by looking into philosophical theories of intrinsic value. I bring forth some pertinent understandings of what it means for a thing to have intrinsic value (Moore, 1993; Korsgaard, 1983), what are the primary senses (Kagan, 1998), as well as different valences of intrinsic value (Zimmerman, 2001). Second, I investigate how these views map onto typical bioconservative arguments about intrinsic badness and moral permissibility of HET. My analysis shows that the debate on intrinsic value places serious constraints on these claims, i.e., they appear unwarranted.<sup>66</sup>

The article explicitly looks into three variants of bioconservative arguments about the intrinsic badness of HET: (i) HET are bad purely in virtue of the way they are – owing to their unnaturalness; (ii) the pursuit of HET (“the desire for mastery”) indicates an intrinsically bad disposition to act; and (iii) HET will necessarily undermine intrinsically valuable things. I challenge each of these variants by leaning on some of the well-established views in the debate on intrinsic value. In particular, I identify intrinsic and relevant nonintrinsic (relational) properties of HET to determine whether one could (and should?) plausibly ascribe negative intrinsic value to HET. Although some bioconservative concerns are relevant in their own right, I argue they are, for the most part, inconsistent, misconceived, and overly speculative to prove HET are problematic in virtue of their intrinsic properties.

Concerning the first variant, I show that we cannot consistently claim across cases that unnaturalness is necessarily a bad property, and if we are concerned about the *disruption* of the natural, the claim about the intrinsic badness of HET is still not decisive. I then reject the idea that the desire for mastery has intrinsic or relational properties that are able to ground the negative intrinsic value of HET, thus undermining the second variant. Finally, contrary to the third variant, necessary consequences that affect intrinsically valuable things cannot ground the (negative) intrinsic value of HET because ‘necessary’ and ‘intrinsic’ are distinct

---

<sup>66</sup> My analysis is limited to the evaluation of intrinsic badness of HET from the perspective of philosophical theories of intrinsic value and the evaluation of internal consistency of bioconservative arguments. As the anonymous reviewer for this journal rightly noted, there are other approaches that could be taken here. Many terms used in the intrinsic value discussion (e.g., nature, dignity, intrinsic badness, etc.) are borrowed from the rich tradition of natural law reasoning or Aristotelian metaphysics school (see, e.g., Murphy, 2019). Although natural law theory may offer valuable insights on this subject, my mode of argument follows the cited authors and is analytical philosophy. For a discussion on biotechnologies and natural law see, e.g., Anderson and Tollefsen (2008); for some theological aspects see, e.g., Boer and Fischer (2013).

concepts, and we lack theoretical as well as empirical support that HET will in fact undermine any of the valuable aspects of human life.

My overall conclusion is that enhancement interventions cannot be bearers of intrinsic value on any of the plausible understandings and, even if we could grant such a possibility, there are no compelling reasons to suggest that the intrinsic value of HET would necessarily be negative. In addition, even if HET had negative intrinsic value, this would not necessarily entail that they are morally impermissible, since moral permissibility need not entirely depend upon intrinsic value. In effect, the assumption that HET are intrinsically bad does not seem to warrant the moral impermissibility thesis. If HET are indeed not morally impermissible for intrinsicality reasons, we may have cleared the way for a more fruitful discussion. Perhaps HET's moral permissibility may further depend on other things, such as their contingent properties.

First, I resolve some conceptual ambiguities and draw upon the debate on intrinsic value to clarify what it means for a thing to be intrinsically bad. I also show how this maps onto the enhancement discussion in the form of a preliminary argument. In the third section, I conduct a three-part analysis of key bioconservative arguments, identifying intrinsic (and some nonintrinsic) properties that may be relevant for establishing the (negative) intrinsic value of HET. Fourth, I reflect on the relationship between intrinsic value and the moral permissibility of HET and offer some suggestions for further research of this subject.

## 2. Intrinsic Impermissibility Thesis, Intrinsic Value, and Enhancement

We can think about the moral impermissibility of human enhancements in at least two senses – enhancements may be considered impermissible regardless of their effects or because of their effects. This is underpinned by the distinction between two types of value, intrinsic and contingent (extrinsic).<sup>67</sup> Namely, things are intrinsically bad purely in virtue of the way they are, and they are extrinsically bad in virtue of the way they interact with the world. Thus, if we believe HET are bad in themselves, and therefore morally impermissible regardless of their effects, we endorse what I will call the *intrinsic impermissibility thesis*. If, by contrast, we hold

---

<sup>67</sup> The traditional distinction between intrinsic and extrinsic value maps onto my concerns about intrinsic and contingent reasons for or against human enhancement. I prefer using the term 'contingent' because I do not think any knock-on consequences or implications are necessary except for those that follow from intrinsic properties.

that the moral impermissibility of HET is derived from bad consequences, we endorse what I will call the *contingent impermissibility thesis*.<sup>68</sup> The intrinsic impermissibility thesis is a stronger claim that corresponds with the aforementioned bioconservative view that HET are morally impermissible even if they turned out to be technologically feasible, legal, and safe.<sup>69</sup> More precisely, the intrinsic impermissibility thesis states that HET are morally bad, and therefore impermissible, *because* of the specific properties that they have intrinsically or necessarily. My main focus is the examination of HET's supposed intrinsic badness, which has direct implications for the plausibility of the intrinsic impermissibility thesis.<sup>70</sup> Before turning to this task, I resolve some terminological ambiguities pertaining to key concepts, and examine how this impacts the discussion on human enhancement.

## 2.1. Conceptual Clarifications

To test the intrinsic impermissibility thesis, we need to examine whether the practice of human enhancement, on its most plausible understanding, can be a bearer of *negative* intrinsic value.<sup>71</sup> This ought to be preceded by establishing conceptually sound notions of human enhancement and intrinsic value, as well as identifying some of their fundamental features. This also involves identifying what exactly is the subject of this ethical evaluation. In bioethics, human enhancement is typically understood as an action or process, “a deliberate intervention, applying biomedical science, which *aims to improve* an existing capacity that most or all normal human beings typically have, or to create a new capacity, by acting directly on the body or brain” (Buchanan, 2011b: 23; emphasis added). This definition has several closely and causally related parts that could all serve as subjects of our analysis. Conceivably, we could analyze separately the biomedical technologies

---

<sup>68</sup> By ‘consequences’ I understand a broad range of effects that are necessarily or contingently brought about by enhancement. They include affecting societal values like justice, rights, and virtue, practical aspects like distribution and regulation, as well as welfare/states of affairs. I will touch upon the relevance of the contingent impermissibility thesis in the last section.

<sup>69</sup> Here, I follow Douglas’ reading of the bioconservative thesis: “[e]ven if it were technically possible and legally permissible for people to engage in biomedical enhancement, it would not be morally permissible for them to do so” (2008: 228).

<sup>70</sup> One could also endorse a weaker version of the intrinsic impermissibility thesis stating that intrinsic badness of a practice *p* provides a strong reason against *p*’s permissibility. Even if this is the case, my conclusions apply.

<sup>71</sup> Negative intrinsic value has the same meaning as intrinsic disvalue or intrinsic badness.

used to perform the intervention, the intervention itself, and the result of the intervention. But here, I will focus on the *intervention* or *activity* itself, while assuming that the biomedical means/technologies are an indispensable part of the intervention.<sup>72</sup> To determine whether enhancement as an *act* can be intrinsically bad, we need to take a closer look at what it means for a thing to have intrinsic value.

Two senses of intrinsic value seem relevant for our purposes: the value a thing has *in itself* and the value it has *as an end*. The first sense is the standard, predominant interpretation of intrinsic value as value a thing has ‘for its own sake’, ‘as such’, or ‘in its own right’ (Zimmerman, 2001; Ronnow-Ramussen et al. 2005: xiii).<sup>73</sup> Since this kind of value does not depend upon anything else, we say it is nonderivative (as opposed to derivative value that a thing derives from something else). In other words, it is the value a thing has solely in virtue of its intrinsic (nonrelational) properties. In the second sense, intrinsic value need not entirely depend upon the object’s intrinsic properties – it can depend in part upon the object’s nonintrinsic relational properties (Kagan, 1998: 280).<sup>74</sup> This occurs when something that is extrinsically good is valued as an end because of the “interest that someone took in it, or the desire that someone had for it, for its own sake” (Korsgaard, 1983: 172). Take *uniqueness*, a nonintrinsic relational property contributing to the intrinsic value of an object such as a work of art. Similarly, intrinsic value could be ascribed based on *causal* properties. For instance, a car’s capacity to perform at a particular speed can be found valuable in itself without the car ever

---

<sup>72</sup> As I mentioned, it does not seem reasonable to think of biomedical means/technologies like medical machines or pills as bad in themselves. Also, equating human enhancement with the end result of the process imposes the risk of normative ladenness. Namely, this entails that enhancement makes us better off by definition – if an intervention does not make us better off, it cannot be considered enhancement. As far as our goal is the evaluation of the intervention as such, we should refrain from equating human enhancement with its end result.

<sup>73</sup> Some scholars warn that to ascribe intrinsic goodness to something is not to say that it is valued for its own sake, but that it has goodness in itself – it refers to the source of goodness rather than the way we value it in (Korsgaard, 1983).

<sup>74</sup> As Kagan explains, “if something does have value as an end, then there is reason to ‘promote’ it, to try to produce the valuable object, or perhaps to preserve and maintain it; we sometimes say that the world is better off ‘as such’ for the existence of the valuable object” (1998: 279). We can assume that the opposite applies to objects with negative intrinsic value: we should not try to produce, preserve and maintain them, and the world seems worse off ‘as such’ for their existence.

being driven or there being an intention to drive it.<sup>75</sup> Although this second view of intrinsic value is not without its difficulties, it has its appeal. I will approach both senses of intrinsic value as relevant for my analysis, but will take that the assessment of the first sense has greater analytical weight because of its prevalence.

In addition to the two senses of intrinsic value, we should note that intrinsic value can have at least two valences, positive and negative (good and bad).<sup>76</sup> In ordinary language, it is standard to understand ‘value’ and ‘valuable’ as something good or favorable. However, in philosophical parlance, it is not uncommon to distinguish between positive and negative value: “the claim that something has value may be predicated not on the judgment that it is good but, for example, on the judgment that it is bad, that is, that its value is the negative one” (Zimmerman, 2001: 3). Also, intrinsic good and bad come in degrees of intensity, which makes intrinsic value computable.<sup>77</sup> But when is it that things have intrinsic value? One traditional method for testing whether a particular thing can be a bearer of intrinsic value is Moore’s *method of isolation*. This test asks whether a thing is such that, if it existed by itself in absolute isolation, we would judge its existence to be good or bad (Moore, 1993).<sup>78</sup> Another method runs in reverse to Moore’s isolation test. Namely, the thing is a bearer of intrinsic value if we can imagine it to have value in any or all circumstances – that it carries its value with it, so to say (Korsgaard, 1983: 171).<sup>79</sup> Although there is more to both of these proposals, simplified versions will suffice for our current purposes. I will take it that a thing has intrinsic value if we found it valuable in all circumstances, or if nothing else existed in the world. What, then, are the preliminary implications of the intrinsic value debate for the discussion on the intrinsic value of HET?

---

<sup>75</sup> Also, certain skills that are instrumentally valuable could be valued in themselves – an ability such as superb cooking could be valued even if one never gets to exercise it – but its intrinsic value depends in part upon its usefulness. For a list of such relational properties and detailed examples, see Kagan (1998).

<sup>76</sup> Some things have no value – they are neither good nor bad, positive nor negative (Timmons, 2012: 8).

<sup>77</sup> For a detailed discussion on degrees of intrinsic value and its computability, see, for example, Zimmerman (2001) Chapter 5.

<sup>78</sup> For a detailed analysis of Moore’s isolation test, see, for example, Zimmerman (2001), Chapter 5.

<sup>79</sup> By contrast, if a thing is not good in any and all circumstances, its goodness is extrinsic – derived from or dependent upon circumstances (Korsgaard, 1983: 171).



## 2.2. Preliminary Argument

Keeping all of the above in mind, it seems that human enhancements cannot be bearers of intrinsic value, and, therefore, cannot be intrinsically bad. If, however, they could have intrinsic value, this value is more likely to be positive than negative. My argument rests on four premises.

First, it seems implausible to say that biomedical interventions, in general, have value ‘in themselves’ or ‘as such’, regardless of anything else. This is because their value seems to depend entirely upon external factors, such as purpose or efficacy. Although HE aims to improve specific capacities, this tells us nothing about the intervention’s intrinsic (or extrinsic) value.

Second, human enhancement interventions are unlikely to have value as ends because they are not desired as ends – they are merely a means to other valuable ends. For example, interventions that aim to improve the immune system or cognition are best understood as a means to some other valuable end, such as health or virtue. Even if we equated enhancement with the end result of the intervention, such as improved memory, hearing or empathy, it is still reasonable to say that we want these goods as means to some other end, like a good life or happiness. More precisely, enhancements are merely tools for acquiring all-purpose goods – things that are *necessarily* good, but should not be conflated with *intrinsically* good things – i.e., their value is always consequentially justified.<sup>80</sup> Thus, it seems that enhancement interventions cannot be bearers of intrinsic value on either of the aforementioned interpretations. This creates a considerable burden for the bio-conservative case.

Third, in order to claim that enhancements are bearers of intrinsic value, we would have to show that they have value in any or all circumstances or in absolute isolation. First, it is not at all obvious that biomedical interventions carry their value with them – in all conceivable circumstances – regardless of their interaction with other things. It is essential to know whether an intervention makes a person better or worse off, in order to judge it good or bad. This suggests the value we

---

<sup>80</sup> All-purpose goods such as memory, intelligence, patience, empathy, and morality, “are traits that are valuable regardless of which kind of life a person chooses to live – valuable on all plausible conceptions of well-being” (Savulescu et al. 2011: 11). Also, things that have *necessary* value (such as all-purpose goods) may be a relevant target of this discussion, but I engage with this later.

ascribe to interventions is *entirely* contingent (not intrinsic).<sup>81</sup> We can easily imagine scenarios in which an intervention to improve one's hearing or memory would make a person better off, as well as worse off. Looking at the intervention (or its intrinsic properties) in isolation, without taking into account any external factors, reveals nothing about its value. Unless we can determine the value of the intervention regardless of its consequences, it will not be a likely bearer of intrinsic value.<sup>82</sup> We draw a similar conclusion when we apply the second rationale. If an improvement in cognition was the only thing in existence (i.e., there are no viable targets of implementation), and we still found it valuable, only then would the value count as intrinsic. But ascribing intrinsic value to such interventions is not intuitively appealing, and even if we were to allow it, we would face the following implications.

Fourth, if HET can be bearers of intrinsic value, it follows (according to the two-valences rationale) that this value can be *positive*, as well as negative. In fact, bioconservative claims about the intrinsic badness of HET set the ground for a counterpoint about their possible intrinsic goodness. Conceptually, at least, we have equally good reasons to believe HET are intrinsically good or intrinsically bad. One might even suggest that we, in fact, have conceptually sounder reasons to believe HET are intrinsically good. Namely, if we were to judge HET a priori, it seems more reasonable to ascribe positive value to interventions that are designed to deliver good things such as improving the quality of life.<sup>83</sup> In principle, we are compelled to accept at least one of the following: either HET *cannot* be bearers of intrinsic value and are, therefore, intrinsically neither good nor bad; or they *can* be bearers of intrinsic value, in which case they can bear both intrinsic goodness and badness.

The debate on intrinsic value places serious constraints on bioconservative views about the intrinsic badness of HET. To determine whether their arguments can overcome such constraints, we need to examine them in more detail,

---

<sup>81</sup> This point is in line with the instrumentalist approach – a predominant view in philosophy of technology which sees technologies as neutral means for achieving human goals (as opposed to substantivism which gives technologies a deterministic role in society) (Borgmann, 1984).

<sup>82</sup> Thus, unless the improvement of cognition is valuable despite scenarios where it makes a person worse off, this action is not intrinsically valuable. Conversely, we would have to show that intrinsic disvalue is present even when an intervention is perfectly safe and makes someone better off in a relevant way.

<sup>83</sup> Perhaps the most reasonable thing to assume from a conceptual point of view is that HET are neither good nor bad, but neutral.

especially in terms of relevant intrinsic and nonintrinsic properties. In what follows, I will inspect three potential sources of intrinsic value, as found in the enhancement debate. The first two concern exclusively the intrinsic and relational properties that could ground negative intrinsic value, while the third relates to the *necessary* consequences that affect intrinsically valuable things.

### 3. Concerns about the Intrinsic Badness of HET

Bioconservative arguments relating to the intrinsic badness of HET can be expressed in the following three ways. First, HET are thought to be bad purely in virtue of the way they are – owing to their unnaturalness; or because their unnaturalness will disrupt the natural. Second, the pursuit of HET is thought to necessarily indicate (or generate) an intrinsically bad disposition to act – often referred to as the “desire for mastery”. Third, HET may necessarily have bad consequences that will undermine intrinsically valuable things.<sup>84</sup> In this section, I challenge each of these variants by taking into account previously established understandings of intrinsic value. I place special emphasis on properties that could conceivably ground the ascription of (negative) intrinsic value to HET. Although some aspects of these bioconservative concerns are relevant in their own right, I argue that they are, for the most part, inconsistent, misconceived, and overly speculative to convincingly establish that HET are intrinsically problematic.

#### 3.1. Concern 1: The Unnaturalness of HET

The ‘unnaturalness concern’ comes in a stronger and a weaker version: enhancements are bad owing to their unnaturalness alone, or they will, due to their unnaturalness, disrupt the preservation of the natural. Although these two versions raise fundamentally different concerns (intrinsic and contingent), they are closely and causally related, and can be jointly addressed. The goal here is to examine both claims with respect to the two senses of intrinsic value. In other words, I examine whether HET are bad in virtue of their intrinsic properties (such as

---

<sup>84</sup> This categorization approximately corresponds to the object, intention, and circumstances relevant for evaluating morality of a human act. In its strict sense, intrinsic badness is only applicable to the object (such as the one in the first concern). However, the other two variants aim to identify *potential* objects of intrinsic value within bioconservative concerns and inspect how that corresponds with standard interpretations of intrinsic value.

unnaturalness), and whether they are bad in virtue of their relational properties (such as the capacity to disrupt the natural).

The stronger version views unnaturalness as an intrinsic property of HET because enhancements are deliberate interventions (they do not occur naturally) brought about by artificial means. The ‘unnaturalness concern’ rests on the assumption that the natural is good, sacred, and should be honored, while the unnatural is bad and should be avoided (Sandel, 2004, 2007; Kass, 2003).<sup>85</sup> It follows that enhancements – as far as they are unnatural – are bad in themselves. This approach, however, fails to distinguish between the natural and the good – the natural is not always good (e.g., natural disasters), and the unnatural is not always bad (e.g., art) (see e.g., Kamm, 2005; Buchanan, 2011b). Not only is the strong version of the unnaturalness concern conceptually flawed, but it is also inconsistent with common practice. For instance, we rarely object to the use of artificial means in medicine merely because they are unnatural. Bioconservatives have themselves recognized the inconsistency of objecting to the means of enhancement due to their artificiality: “[since] the use of artificial means is absolutely welcome in the activity of healing, it cannot be their unnaturalness alone that upsets us when they are used to make people ‘better than well’” (Kass, 2003: 21). Although there is a sense that the naturalness of means matters, as Kass notes, the problem of means “lies not in the fact that the assisting drugs and devices are artifacts, but in the danger of violating or deforming the deep structure of natural human activity” (2003: 22). Thus, it seems that unnaturalness alone is not an intrinsic property of HET that can ground negative intrinsic value. This brings us to the second part of the unnaturalness concern and the other sense of intrinsic value.

The weaker version is concerned with HET’s capacity to disrupt the natural; this capacity could be a nonintrinsic relational (most likely causal) property that affects HET’s intrinsic value. As I already acknowledged, bioconservatives are not concerned with unnaturalness simpliciter, but rather with the *preservation of the natural* (status quo). Naturally given processes such as natural procreation, the human life cycle and flourishing are inherently precious and should be preserved (President’s Council on Bioethics, 2003: 288). Enhancements therefore represent

---

<sup>85</sup> The concept of human nature used here equates nature with physicality or biology. It is sometimes possible to distinguish between physical and metaphysical aspects of human nature. However, it is not in the scope of my paper to engage with those views.

a threat to the natural – they can interfere with or override it.<sup>86</sup> This view suggests that there are necessary consequences (in this case, negative ones) caused by intrinsic properties of enhancement.<sup>87</sup> Thus, if an object can have value derived from relational properties such as causal properties (e.g., Kagan 1998), this could affect the value HET have as ends.

However, even if all of these claims are true, they do not decisively determine the intrinsic badness of HET. First, we should not conflate necessity with intrinsicity. Necessary consequences do not show that HET are intrinsically bad. Even though necessary consequences can strongly affect moral judgment, their actual value is always consequentially (derivatively or relationally) justified. For instance, pollution is a necessary feature of air travel, but not its intrinsic property, i.e., it does not make flying intrinsically bad. Even if HET were to have necessarily bad consequences, this would not decisively determine their intrinsic value.<sup>88</sup> Second, even if enhancement's capacity to disrupt the natural is indeed a casual property that can impact its intrinsic value, this tells us nothing about the *valence* and the *degree* of that value. If value as an end need not be based on intrinsic properties alone, since the object can have value as an end in virtue of some subset of its properties (Kagan, 1998: 291), this would equally apply to all sorts of nonintrinsic properties. In order to plausibly claim HET are intrinsically bad, we have to show that intrinsic value based upon relational properties (such as the capacity to disrupt the natural) is not only *negative*, but *so* negative that no amount of positive value could justify the use of HET. Conceptually (as argued in the previous section), we have no particular reason to assume HET's intrinsic value is negative. Empirically, evidence is not yet available to support either positive or negative intrinsic value in HET.

To summarize, regardless of whether our focus is unnaturalness alone or the disruption of the natural, claims about HET's intrinsic badness are not justified. It does not necessarily follow that HET have negative intrinsic value in virtue of

---

<sup>86</sup> The idea of preserving the natural and human nature is paradoxical because these things are constantly changing and enhancements can become essential for preserving the status quo (Harris, 2007; Buchanan, 2011b). Also, it is erroneous to think that most enhancements would change human nature since this would require a modification of a central trait on a population level (Daniels, 2009). I will therefore focus on challenges prompted by the debate on intrinsic value.

<sup>87</sup> I will address specific aspects of necessary consequences caused by intrinsic properties of HET, as well as their relational properties, later in the article.

<sup>88</sup> This issue will be addressed in more detail under the third concern.

their intrinsic properties, such as unnaturalness. This is because it is conceptually mistaken to equate the unnatural and the bad, as well as inconsistent with common practice to object to biomedical means based solely upon their artificiality. Even if value as an end can be affected by nonintrinsic properties, we have no particular reason to assume this value would be negative in sum. Perhaps the source of intrinsic badness lies elsewhere, e.g., in the very *desire* to pursue enhancements or disrupt the natural.

### 3.2. Concern 2: Pursuing HET Is an Intrinsically Bad Disposition

On this version of the bioconservative view, pursuing HET indicates bad character, i.e., it reveals the possession of an intrinsically bad disposition to act. For instance, Sandel explains we should not be so concerned about enhancements undermining valuable things such as effort or human agency, but instead about the attitude and dispositions that prompt the drive to enhancement. This concerns the problematic aspiration to “remake nature, including human nature, to serve our purposes and satisfy our desires. The problem is not the drift to mechanism but the drive to mastery” (2004: 54). This desire is not only detrimental (or instrumentally bad) to our sense of giftedness and humility, but it also indicates (intrinsically) bad character. However, it is far from clear how we should interpret the drive to mastery: does it motivate enhancement, does it constitute it, or is it perhaps identical with it? Still, we can try running these different possibilities against our two main senses of intrinsic value. In this section, I examine whether HET are bad in virtue of their intrinsic properties, such as (indicating) bad character; and whether they are bad in virtue of their relational properties such as desiring mastery.

First, let us consider whether HET are bad in themselves in virtue of indicating bad character. Pursuing HET indicates bad character, i.e., it reveals the possession of an intrinsically bad disposition to act. Thus, if we take bad character as a property intrinsic to practicing enhancement, we could perhaps claim that this grounds the negative intrinsic value of HET. A similar interpretation comes from Buchanan, who explains the “concern that the pursuit of enhancements, independently of its consequences, itself *indicates* bad character” as the expressivist or nonconsequentialist type of character concern (2011b: 69). By contrast, consequentialist concerns are “predictions that the pursuit of enhancements will *result* in a worsening of our characters” (Buchanan, 2011b: 69). The expressivist concern can be further understood as the claim that a stable desire to enhance is itself a

manifestation of vice or, at least, predominantly the expression of a vice (Buchanan, 2011b: 69). But can the agent's character plausibly ground the value some activity has in itself?

The suggestion that HET are bad in themselves because they are motivated by bad character is flawed in several respects. It is generally mistaken to think about desires, motivations, and character traits as intrinsic properties because they are typically subjective/relational. These properties must be *intrinsic* to an enhancement intervention in order to ground its intrinsic value. However, they are not constitutive of its description or definition, or in any other way part of its intrinsic nature. An agent's character traits, motives, and desires might count as relevant *nonintrinsic* properties, but surely they do not determine the value that an enhancement intervention has *in itself*. Enhancement critics may *identify* enhancement with the desire for mastery or assume that they are intertwined/closely related. But even then, they would still have to show why desiring mastery is bad in itself – what is intrinsically wrong about taking control over (human) nature (assuming that taking absolute control is even possible)?<sup>89</sup> The wrongness of such interventions is not self-evident because the 'hesitation' and 'unease' we may feel about enhancements are not decisive for determining their intrinsic badness.

The concern about mastering our nature may collapse into the previously discussed concern about the value of the natural. Although it is flawed to assume that natural is always good, as I previously showed, bioconservatives seem to be making an implicit claim that deliberate changes to human nature are illicit or immoral: "[to] successfully claim that a change in a person's nature is intrinsically immoral, we need a premise that there is an obligation to limit ourselves to the capacities provided by evolution" (Lindsay, 2012: 19). Hence, if there is even the slightest chance of HET being intrinsically bad, because they presumably indicate a bad desire, bioconservatives need to show why mastering our nature is bad in itself. This is yet to be proven, but such a claim seems difficult to sustain.

On the second account of intrinsic value, even if the indication of bad character cannot ground the value of an enhancement intervention in itself, some non-intrinsic relational properties could contribute to its value as an end. Nonintrinsic relational properties such as subjective experience, a manifestation of excellence, causal properties, etc. can contribute to a thing's intrinsic value (e.g., Kagan,

---

<sup>89</sup> Actual mastery and the desire for mastery are two distinct concepts (Kamm, 2005). It is improbable that actual mastery (absolute control over nature, including human nature) is attainable because while most enhancements involve changes in an individual's nature, they do not have the potential to overthrow human nature entirely (Daniels, 2009).

1998). Desiring enhancement (or mastery – assuming they amount to the same thing) as an end would count as relevant subjective experience. This assumption seems *prima facie* justified since desires are typically relational and desiring a thing as an end is one of the valid ways to ascribe intrinsic value. However, I argued in the first section that it seems most reasonable we desire enhancement as a means to some other end, such as health, virtue, or beauty – not as an end in itself.<sup>90</sup> Thus, as far as we desire mastery instrumentally, the bioconservative assumption is wrong, and even if we desired mastery as an end, mastering our nature would not be proven necessarily bad.<sup>91</sup>

Furthermore, it is erroneous to equate enhancement with complete mastery. Even if we take mastery to represent a manifestation of excellence, which counts as a relevant nonintrinsic relational property and could contribute to the value a thing has as an end, it does not follow that enhancement is that thing. Enhancement should not be equated with mastery because mastery stands for improving a skill to the point of perfection, whereas enhancement is typically understood as any (and not necessarily the highest) degree of improvement above the norm. A distinction between greater and complete mastery would allow us to show, at best, that enhancement indicates a desire for greater mastery broadly considered, but not necessarily complete mastery. Bioconservative claims seem most plausible when we talk about complete mastery, less so for less than complete mastery. Most advocates of enhancement (except perhaps radical transhumanists) would say that enhancement does not aim for complete mastery (perfection) at all, but merely an improvement on the current state of affairs.<sup>92</sup> If so, the entire argument from ‘perfection’ might be missing its target.

One might object that enhancement could lead to mastery (or other intrinsically bad things) on similar grounds as it may lead to the disruption of the natural – in virtue of its *causal* properties. Causal properties of an object are relevant for intrinsic value when the object produces or is a means to another valuable object (Kagan, 1998: 283).<sup>93</sup> Thus, if enhancement is a means to (or produces) mastery,

---

<sup>90</sup> Some scholars also suggested we could master nature as a side effect, without desiring it as an end (Kamm, 2005: 6).

<sup>91</sup> I address this further later in the paper.

<sup>92</sup> If the exact target is radical transhumanism, then the argument is inapplicable to *all* enhancements, since most entail minor to moderate changes.

<sup>93</sup> Kagan’s example of such a causal property is the capacity of a racing car to perform at a particular speed. We might think the car is intrinsically valuable in virtue of its relational properties, including its causal properties (1998: 284).



and mastery is intrinsically bad, then enhancement may produce negative intrinsic value in virtue of its relational/causal properties. However, as I already argued, not only is enhancement not necessarily a means to complete mastery, but complete mastery is not decisively intrinsically bad. If we are, in turn, discussing only greater levels of mastery, in broader terms, the bioconservative argument applies with even less strength.

Even if we grant that the desire for mastery necessarily motivates enhancement and that mastery is intrinsically bad, which in turn affects the intrinsic value of enhancement (in virtue of its causal properties), it does not follow that the value of enhancement as an end is necessarily negative on balance. We would need to show it to be so overwhelmingly bad that it grounds the negative intrinsic value of HET. Since nonintrinsic properties only *contribute* to intrinsic value, their effect is not decisive. Contributive value is commonly understood as the value of a part in an intrinsically valuable whole (Korsgaard, 1983). Thus, the bioconservative assumption that the desire for mastery is so bad that it outweighs all other contributing factors is not obviously true and calls for additional support. I argued earlier why such a claim is not conceptually stronger than the claim about the intrinsic goodness of HET, but I offer additional reasons in the next section.

### 3.3. Concern 3: HET's Necessary Consequences

So far, we have examined the possibility of different intrinsic (unnaturalness; bad character) and nonintrinsic (disrupting the natural; desire for mastery) properties grounding the negative intrinsic value of HET. We saw that the debate on intrinsic value poses various challenges and, absent further argument, offers no reason to assume HET are intrinsically bad. However, there is another variant of the bioconservative argument: HET may vitiate intrinsically valuable things; as far as they do so *necessarily*, a relevant concern arises regarding the intrinsic value of HET.<sup>94</sup> This argument raises two concerns: i) do HET *in fact* represent a threat to intrinsically valuable things (do they necessarily generate bad consequences), and ii) even if they do, can necessarily bad consequences ground intrinsic value? Here, I mainly focus on the second concern, showing that HET are not necessarily a threat

---

<sup>94</sup> It is important to distinguish between necessarily having/generating bad consequences and contingently generating consequences that are themselves necessarily bad. What I mean here is the former.

to intrinsically valuable things (not all HET will generate bad consequences), and even if they are, this does not determine their intrinsic value.

I mentioned earlier that bioconservatives believe enhancements represent an aspiration to remake human nature and take absolute control over our lives. The negative side of this aspiration, according to Sandel, lies in the possibility of destroying the appreciation for the gifted character of human powers and achievements; in other words, we would be missing the sense of life as a gift (2009: 53-54).<sup>95</sup> Genetic enhancements will “undermine our humanity by threatening our capacity to act freely, to succeed on our own and to consider ourselves responsible – worthy of praise or blame – for the things we do and for the way we are” (2009: 78). Similarly, Fukuyama (2002) fears that biotechnologies threaten to undermine our human essence and dignity, and are likely to create a genetic underclass. His argument about human dignity states that enhancement will undermine the grounds for a nonarbitrary claim to equal respect: “What the demand for equality of recognition implies is that when we strip all of a person's contingent and accidental characteristics away, there remains some essential human quality underneath that is worthy of a certain minimal level of respect--call it Factor X” (2002: 149).<sup>96</sup> What we want to protect from future advances in biotechnology is “the full range of our complex, evolved natures against attempts at self-modification. We do not want to disrupt either the unity or the continuity of human nature, and thereby the human rights that are based on it” (2002: 172). The question then is whether HET will necessarily undermine values such as giftedness and human dignity and how this relates to their intrinsic value.

Enhancements are not necessarily incompatible with, nor will they necessarily undermine intrinsically valuable capacities such as giftedness and dignity. Several scholars challenged Sandel's argument stating not only that the deterministic approach to enhancement is false, but that enhancements might in fact improve some capacities we find intrinsically valuable (e.g., Kamm, 2005; Savulescu,

---

<sup>95</sup> Other examples of intrinsically valuable things threatened by enhancements include humility, spontaneity, the mystery of life, openness to the unbidden, human flourishing, etc. Hauskeller explains this concern in greater detail by claiming that attempts to make better people seem to be detrimental to humility, a virtue that Sandel highly values. Hence, the main concern is the harm we might inflict on ourselves as enhancers, by allowing the reign of our inherent drive to mastery (Hauskeller, 2013: 167). Harm, however, is not a question of intrinsic value.

<sup>96</sup> Factor X is a combination of qualities such as “the possession of moral choice, or reason, or language, or sociability, or sentience, or emotions, or consciousness, or any other quality that has been put forth as a ground for human dignity” coming together in a whole (Fukuyama 2002: 171).

2009; Buchanan, 2011b; Lindsay, 2012; Hauskeller, 2013).<sup>97</sup> Even if some extreme versions of enhancement might represent threats to intrinsically valuable capacities (perhaps radical transhumanism), it is not a *necessary* feature of enhancements that they do so. It is more important to focus here on the concerns that considerations of intrinsic value bring about: can necessary consequences ground intrinsic value?

Following the standard interpretation of value that a thing might have in itself, it is conceptually implausible to ground intrinsic value on consequences, even if they are necessary. I already mentioned, while addressing the first concern, that necessity and intrinsicity are two distinct concepts. The former concerns the consequences of an act and, *ipso facto*, its value is always derivative. The latter concerns the source of value, which resides in the thing itself. I previously mentioned the example of pollution being a necessary consequence of air travel, without making air travel intrinsically bad. Translated into enhancement terms, even if undermining giftedness/dignity were a necessary consequence of enhancement (imagine their occurrence is intertwined/closely related), this does not make it an intrinsic feature of enhancement – especially not the sort that can ground intrinsic value. If anything, the necessary property could impact the contingent value, but this does not warrant a conclusion about HET's necessary intrinsic badness.

Let us look more closely at Fukuyama's argument about the undermining of human dignity. Fukuyama holds that the genetic lottery is inherently unfair, but also profoundly egalitarian, while replacing it with choice threatens to increase the disparity (2002: 157). In other words, the natural lottery is not bad despite its necessarily bad properties, while self-modification such as HET is bad for being brought about deliberately. Fukuyama suggests that unfair but egalitarian circumstances (such as being equally subject to nature) are better than unfair and inequalitarian (such as deliberate self-modification). While this is a relevant concern in its own right, it does little work for determining the negative intrinsic value of HET. If the natural lottery is good merely because it is natural, then this argument collapses into the (un)naturalness concern. If the natural lottery is good because it is not as bad as deliberate change, then the concern is not about HET's intrinsic but rather its contingent properties.

---

<sup>97</sup> For example, Kamm (2005) argued that valuable things promoted by enhancements (e.g., health, virtue, beauty) are not incompatible with other things valued in themselves (e.g., spontaneity, giftedness, human flourishing). The deterministic effect of enhancements was challenged by Savulescu (2009; 2010), Buchanan (2011b), and Lindsay (2012). Some weak points of the argument from giftedness were analyzed by Hauskeller (2013).

Alternatively, we may take that necessary consequences are relevant for the value a thing has as an end, assuming they count as nonintrinsic relational properties. It is disputable whether consequences can count as nonintrinsic relational properties because consequences are usually distinct from something's properties. But even if this were possible (for reasons of necessity or causality), challenges similar to those discussed previously would emerge. We need to show that the value as an end would be, in sum, negative – the badness of HET's nonintrinsic properties (necessary consequences) would have to outweigh all other considerations. I have already argued that this claim is not particularly convincing from a conceptual standpoint; considering the points I make against the necessary badness of HET in this section, it seems even weaker. At any rate, a more substantive claim is needed if we want to argue that HET are intrinsically bad in virtue of their consequences. Critics of enhancement have not offered plausible arguments on this matter so far, and my analysis suggests this can hardly be expected.

In summary, bioconservative arguments grounded in unnaturalness, the desire for mastery, and necessary consequences do not warrant any conclusions about HET's intrinsic value. They are, as they currently stand, highly speculative, incoherent, and empirically unfounded. Now that we have drawn tentative conclusions about the intrinsic status of HET, we need to explore how this reflects upon the intrinsic impermissibility thesis, presented in the first section.

#### **4. The Gap between Intrinsic Badness and Moral Permissibility?**

The examination of HET's intrinsic status was primarily motivated by its relation to the intrinsic impermissibility thesis, i.e., the claim that HET are morally impermissible regardless of their effects. What do our tentative conclusions about the intrinsic status of enhancement say about its moral permissibility? My analysis showed that considerations of intrinsic value place significant constraints on the intrinsic impermissibility thesis. Ascribing negative intrinsic value to HET, absent further argument, seems unwarranted. Let us, nevertheless, assume that more cogent arguments for HET's intrinsic badness become available. Would this resolve that they are morally impermissible? Probably not, and here is why.

The intrinsic impermissibility thesis expresses a very general bioconservative stance towards enhancements, but it also assumes a strong connection between a thing's intrinsic value and its moral permissibility. This assumption, however, is not entirely justified. A thing's intrinsic value may carry significant normative weight (it can affect moral judgments and the moral status of an object), but,

strictly speaking, it does not necessarily determine that thing's moral permissibility. Things can still be morally bad and permissible (e.g., war). Moral theories organize and relate accounts of the right and the good differently: e.g., duty-based theories place lesser weight on value concepts, while theories of value explain right and wrong action in terms of how they bear on intrinsic value (Timmons, 2012: 10-11). Therefore, goodness and badness do not necessarily affect the (im)permissibility of an act.

Consider once more the bioconservative concern about HET being intrinsically bad because they indicate bad character (the desire for mastery). One could argue that the agent's character and motivation can be relevant for the moral status of an act, but does not affect its moral permissibility. Some scholars have put forward arguments along these lines. For example, Kamm (following Scanlon and Thomson) suggests that the permissibility of an act can be seen as independent from our intentions or dispositions to carry the act out – intentions and attitudes of an agent reflect on the agent's character but not on the permissibility of their act (Kamm, 2005: 7). For example, a scientist motivated by the desire for mastery in her quest to find a cure for some nasty disease may not be a good person on every moral account, but we would not find her conduct impermissible. Similarly, Buchanan states that motivation is not the only relevant factor for determining permissibility, given that even bad motivation can result in the right act: “even if it were true that the pursuit of enhancement is always driven solely by bad character, it would not follow from this that enhancement is morally impermissible” because “one can perform the right act as a result of bad motivation” (2011b: 72). Hence, even if there were reasons to think HET are intrinsically bad, this would not necessarily imply their moral impermissibility.

This is not to say HET are generally permissible. Even if the intrinsic impermissibility thesis is false, and we cannot plausibly claim HET are morally impermissible for intrinsicality-based reasons, they can be impermissible for other reasons. Their permissibility may depend upon potentially bad consequences, i.e., the plausibility of what I have previously called the *contingent impermissibility thesis*. The contingent impermissibility thesis states that HET are morally impermissible insofar as they are reasonably expected to be contingently bad. Unlike the intrinsic impermissibility thesis, the contingent impermissibility thesis is not an absolute claim. It allows for the conditional assessment of HET: *If* HET are bad in their interactions with other things, then they should be deemed morally impermissible. This implies that HET could be morally *permissible* if we could reasonably expect them not to have bad effects. If we endorse this line of reasoning, future

research should consider contingent instead of intrinsic factors.<sup>98</sup> My examination has hopefully cleared the way for a more fruitful discussion focused on the evaluation of the contingent rather than the intrinsic value of HET.

## 5. Conclusion

In this article, my primary aim was to show that notions of intrinsic value place serious constraints on claims about the intrinsic badness of HET. I argued that enhancement interventions are not typical bearers of intrinsic value on any of its plausible understandings. Even if we granted such a possibility, I argued that there were no compelling reasons to accept that the intrinsic value of HET was negative. Additionally, even if HET had negative intrinsic value, this would not warrant an unfavorable verdict about their moral permissibility. The examination of HET's intrinsic and relevant nonintrinsic properties, such as unnaturalness, the agent's desire for mastery, and the necessary production of bad consequences, was not decisive in establishing their intrinsic badness. Under scrutiny, most of these concerns collapse into one another, and ultimately into questions about contingent factors that could, in fact, play a prominent role in accounting for the moral permissibility of HET.

---

<sup>98</sup> A similar idea comes from Kamm, as she suggests that in accounting for the permissibility of an act: “[w]e will have to decide whether particular changes are permissible independently of the aims, attitudes, and dispositions of agents who act” (2005: 7).







## Chapter 4

# The Role of Emotion Modulation in the Moral Bioenhancement Debate

**Abstract:** This paper aims to analyze technical and internal aspects of one particular type of human moral enhancement, i.e. enhancement of moral motivation via direct emotion modulation. More precisely, it challenges the assumption that modifying certain emotions will have the results desired by the advocates of this proposal. It is argued that neuropsychological understanding of the role and function of emotions, as well as of underlying cognitive mechanisms, might be relevant for the discussion about biomedical enhancement of moral capacities. Moreover, typical claims about direct emotion modulation seem to be contradicted, or at least seriously challenged, by available neuroscientific data. Particular attention is paid to the theory that emotions are evolved and functionally specialized programs whose task is to coordinate other adaptive mechanisms of human psychology in order to promote one's fitness. If this view of emotions is plausible, it can be argued that several difficulties for moral bioenhancement ensue. Neuroscientific and evolutionary-psychological perspectives seem to indicate that emotions do not fulfill necessary requirements to serve as the vehicles of moral enhancement and it should, therefore, take into account the role and function of entire cognitive modules associated with moral decision-making.

This chapter is a linguistically improved version of the article: **Kudlek K. 2019. The role of emotion modulation in the moral bioenhancement debate, *Topoi* 38, 113 -23. DOI: <https://doi.org/10.1007/s11245-017-9481-9>**



# The Role of Emotion Modulation in the Moral Bioenhancement Debate

## 1. Introduction

The idea that human moral character might be in serious need of improvement became widely discussed in contemporary bioethics. Ambition to reshape our moral nature is not new in human history, but introducing biomedical science and technology into this equation caused a fair amount of scholarly disagreement. Hence, the suggestion that biomedical means could and should be used to directly improve our moral capacities became one of the hottest bioethical topics. Whereas the controversy and the concerns about ethical permissibility of biomedical moral enhancement preoccupy the majority of theorists, this paper aims to analyze more technical and internal aspects of one particular type of human moral enhancement, i.e. enhancement of moral motivation via direct emotion modulation. More precisely, it challenges the assumption that modifying certain emotions will have the results desired by the advocates of this proposal.<sup>99</sup>

Enhancement can be generally defined as any deliberate intervention that aims to improve a capacity that most or all human beings typically have or to create a new capacity in a human being (Buchanan, 2011b: 23). It can also be defined as any deliberate intervention that aims to (i) improve existing capacity, (ii) select for desired capacity or (iii) create a new capacity in a human being (DeGrazia, 2014: 361). Accordingly, moral enhancement aims to improve human moral capacities and traits – regardless of whether the nature of moral enhancement is traditional or biomedical. Biomedical human enhancement differs from traditional enhancements only in so far as it applies biomedical means directly to one's body or brain in order to improve some existing capacities or create new ones. There are several types of capacities usually associated with improving moral judgment and

---

<sup>99</sup> It is worth bearing in mind, as one of the reviewers of this journal correctly observed, that MBE is in principle problematic idea which entails disturbing claims about maladapted human morality. Nevertheless, it seems worth asking whether modifying emotions would actually have the result desired by proponents of MBE. Since the latter is the exclusive goal of this paper, general ethical permissibility or justification of MBE will not be discussed.

behavior. We can target *affect*, *emotion* and *motivation* (Buchanan, 2011b) or try to improve certain *behaviors* and *dispositions* (Jebari, 2014).

Contemporary discussions about moral bioenhancement (MBE) are typically focused on the direct modulation of certain emotions with the goal of enhancing moral motivation. Moral enhancement might, therefore, comprise of the *attenuation* of counter-moral emotions<sup>100</sup> (as suggested by Douglas 2008; 2013) or in *strengthening* our altruistic emotions and a sense of justice or fairness (as suggested by Persson and Savulescu 2008; 2011; 2012). According to the mentioned authors, human morality is, allegedly, based in certain biological dispositions whose elementary function is reciprocity. Namely, central moral dispositions like altruism and a sense of justice were selected for in the course of evolution because they increased one's fitness and inclusive fitness – giving a survival advantage to the most successful participants in various tit-for-tat interactions. Since the world we currently live in considerably differs from the environment in which our emotions evolved, MBE proponents see the need for “boosting” our altruism and properly “attuning” related emotions in order to make them morally appropriate. Vividly put, MBE via direct emotion modulation would consist in “updating” our “moral software”. The feasibility of such an intervention appears to be backed up by scientific findings which show that manipulation of our biological make-up can have morally desirable effects. Proponents of MBE find it reasonable, therefore, to assume that biomedical means can and should be used to boost moral motivation and improve moral decision-making even in healthy individuals.<sup>101</sup>

The available neuroscientific evidence seems to suggest that targeting moral motivation may be the most promising way for promoting moral behavior (Crockett, 2014). It also suggests that our motivational processes have neurobiological underpinnings and, consequently, that there is a link between modulation of emotions and modulation of neurobiological mechanisms. Although many details of this complex relationship are still unknown, MBE proponents do not seem to take seriously the important differentiation between mind's various programs and the way they operate. The basic idea of this paper is to show how neuropsychological

---

<sup>100</sup> Possible examples of counter-moral emotions are a strong aversion to certain racial groups and an impulse towards violent aggression (Douglas, 2008: 231).

<sup>101</sup> It is generally assumed that enhancement procedures go beyond restoring health and are supposed to be applied to “normal” individuals, boosting their capacities above some species-typical standard. MBE discussed here, therefore, is about the application of biomedical means to healthy individuals (although findings relevant for this debate are mostly from medical cases of abnormalities which predispose for immoral behavior).

understanding of the role and function of emotions, as well as of underlying cognitive mechanisms, might be relevant for the discussion about biomedical enhancement of moral capacities. It can be argued, moreover, that typical claims about direct emotion modulation seem to be contradicted, or at least seriously challenged, by available neuroscientific data.

This shall be demonstrated through four main steps. First, the main goals of MBE (as presented by Persson and Savulescu) will be outlined. It will be shown that certain difficulties arise because of insufficient conceptual clarity of this proposal. Secondly, basic ideas of MBE via emotion modulation will be expanded through comparison with evolutionary-psychological and neuroscientific perspectives. Particular attention will be paid to the theory that emotions are evolved and functionally specialized programs whose task is to coordinate other adaptive mechanisms of human psychology in order to promote one's fitness. If this view of emotions is plausible, it can be argued that several difficulties for MBE theory ensue. Potentially problematic aspects of emotion modulation will be discussed in the third part of the paper, whereas its possible negative effects will be presented in the final section.

## 2. The Moral Bioenhancement Proposal

The intention behind the MBE proposal is to bridge the gap between (a) scientific and technological progress achieved by the human species during the past few centuries and (b) the absence of comparable progress in our species-typical moral psychology.<sup>102</sup> Its proponents argue that human beings are not naturally equipped with a moral psychology that would empower them to adequately cope with the moral problems that these new conditions of life create (Persson and Savulescu, 2012: 1). These drawbacks of human moral psychology are manifested primarily as the lack of moral motivation – as the propensity to actually do what one already believes it ought to be done – and they might ultimately cause great harm to the entire life on Earth. Since this mismatch presents a serious threat to the human survival as well, proponents of MBE offer a sort of “evolutionary short-cut” or “artificial update” of our moral psychology in order to make us better “morally

---

<sup>102</sup> Moral psychology can be simply understood as psychology of a moral agent. Persson and Savulescu do not provide its explicit definition, but they clearly understand it as a combination of common-sense morality and some related psychological dispositions (explained in more detail later in this article).

adapted” to our current environment and conditions of life. As Persson and Savulescu explain:

We hypothesized, based on evidence from evolutionary biology and psychology, that *the moral psychology of humans is adapted to the former conditions*, which have obtained for most of the time the human species has existed. This mismatch is a serious matter because humans now have at their disposal technology so powerful that it could bring about the destruction of the whole planet if misused. (Persson and Savulescu, 2011:1, italics added)

Persson and Savulescu assume that occurring problems in human moral behavior might be caused by the limitations of our moral psychology which was originally adapted to importantly different conditions. Namely, human beings lived in small, close-knit societies with primitive technology for almost all of their history, whereas now, most people live in large communities with access to advanced technologies, enabling them to exercise an influence that extends all over the world and far into the future (2012; 2015a). This mismatch is reflected in several psychological tendencies or biases that don't seem adequate for the moral requirements of the modern world. For example, human beings appear to be temporally and spatially “myopic” which means that we are “disposed to care more about what happens in the near future to ourselves and some individuals who are near and dear to us” (2015a: 338).

Consequently, we are incapable of responding adequately to the suffering of distant, unknown individuals and larger collectives and we are often inclined to morally undesirable behavior such as selfishness, nepotism, xenophobia or groupishness. Psychological myopia also causes inadequate responses in the context of care for future people, the environment and non-human animals. Persson and Savulescu conclude: “[s]uch a limited moral psychology is an ineffective brake on misuse of technology when modern weapon technology enables us to create weapons to kill large numbers at long distances” (2015a: 338). As already announced, a possible solution to our problems is a biotechnological update of our moral software.

The most fundamental question that every morality-related theory encounters is the meaning of concepts such as “moral” and “morality”. Proponents of MBE are aware of countless accounts of what “morality” and “moral” might mean, but they prefer the view which, in their opinion, is relatively neutral and shared by many. According to this view, “the core of our moral dispositions comprises, first, a disposition to altruism, to sympathize with other beings, to want their lives to go

well rather than badly for their own sakes” (2008: 168).<sup>103</sup> There is also a set of dispositions from which the *sense of justice* or *fairness* originates and their elementary form is tit-for-tat.<sup>104</sup> It is worth noting that the general understanding of morality that accompanies this view is the so-called “folk” or “common-sense” morality. By common-sense morality, Persson and Savulescu mean “a set of moral attitudes that is a common denominator of the diversely specified moralities of human societies over the world” (2012: 12). Thus, they find it reasonable to assume that the dispositions for altruism and justice constitute the centerpiece of this morality because they motivate us to act in accordance with plausible basic moral principles (2015b: 349).

The concept of MBE encompasses the enhancement of human motivation to act morally through the application of biomedical and pharmacological means, like drug treatment or genetic engineering. Such interventions seem possible because, as already explained, our moral dispositions are biologically based and shared with many non-human animals. If these core moral dispositions (altruism and sense of justice) are biologically grounded, then they could be malleable by biomedical and not only by traditional means. Biomedical means could be used for strengthening our altruism and making us more just or fair, i.e., for “attuning” certain tit-for-tat emotions in order to make us properly grateful, angry, forgiving, etc. For example, “more altruism is likely to initiate more tit-for-tat exchanges, though too much altruism may be an obstacle by making us turn the other cheek when tit-for-tat requires retaliation” (Persson and Savulescu, 2008: 168-9).

---

<sup>103</sup> This notion of altruism includes benevolence and Persson and Savulescu believe that altruism conceived in this way must be central to morality since “morality requires the setting aside of our own interests for the sake of others, though to what precise extent is a matter of controversy” (2015b: 349). It is important to note here that altruism can be seen as an insufficient foundation of morality, but this point will be further elaborated in the final section.

<sup>104</sup> This means that collectives in which the pattern of reciprocal reactions (driven by emotions such as anger, remorse, guilt, pride, admiration, contempt or forgiveness) is widespread are most successful in terms of survival and reproduction (2008: 169). For example, recognizing needs in kin increases inclusive fitness, whereas cooperation and reciprocating can increase chances of individual survival, which is why altruistic emotions were selected in the course of evolution. The sense of justice and fairness is based on the feeling of anger or contempt at those who do not reciprocate (free-riders), gratitude or admiration to those who do, guilt for not doing right to someone, shame for not being able to properly reciprocate or retaliate, etc.

Regardless of whether the focus is set on *strengthening* pro-moral or *diminishing* counter-moral emotions,<sup>105</sup> the bottom line is that modulation of emotions seems like a promising ground for enhancing human moral motivation. Manipulating emotions is not new in human lives, after all, it is one of the oldest educational and upbringing methods. Emotions are already successfully manipulated through the administration of various mood-altering pharmaceuticals, and there are also certain neuroscientific findings indicating that eliminating dispositions for immoral behavior may be within the reach of biomedical or genetic treatment. However, since the scope of this paper is limited to examining the role of emotions in the enhancement debate, it is necessary to examine whether modifying emotions would, in fact, have the result desired by proponents of MBE. This is a simple resume of the argument to be challenged:

1. Human moral psychology is outdated (it has been adapted to the past environment);
2. This problem manifests itself as a lack of moral motivation;
3. Moral motivation can be boosted by biomedical emotion modulation;
4. Therefore, emotion modulation can fix the problem of outdated moral psychology.

Although the conclusion seems to follow from the premises, it could be argued that it contains a flaw. The second premise can be expanded with the following clarification: our moral psychology consists of species-typical neurobiological mechanisms (or modules) designed by natural selection for the living conditions of distant evolutionary past with the primary goal of promoting fitness – i.e., a lack of moral motivation is caused by maladapted psychological mechanisms. An example of such maladapted mechanisms is the aforementioned temporal and spatial moral myopia which causes a lack of care or compassion for the suffering of

---

<sup>105</sup> Whereas Persson and Savulescu discuss the possibility of increasing altruism, Thomas Douglas's proposal is focused on biomedical attenuation of counter-moral emotions that tend to interfere with sound moral reasoning, sympathy and other plausible candidates for "morally good motives" (Douglas, 2011: 2). The implications of emotion attenuation are not to be discussed *per se* in this paper, but they do fall within the "emotion modulation" scope and, accordingly, the same conclusions apply to them too.



distant strangers.<sup>106</sup> It seems more appropriate, therefore, to talk about modulating these mechanisms rather than about modulating emotions, because: If human moral shortcomings are caused by maladapted mechanisms of moral psychology (not by pure lack of motivation), how can moral motivation be boosted by emotion modulation? Or in other words: If human moral shortcomings are caused by maladapted mechanisms of moral psychology, how can emotion modulation fix the problem of outdated moral psychology?

Although Persson and Savulescu begin by discussing the problem of maladapted mechanisms of human moral psychology, they proceed to emotion modulation suggestion without sufficient explanation of the role of emotions and mechanisms in the mental economy. Some strong connection between the two is presumed, implying that tampering with emotions will supposedly solve the problem of insufficient moral motivation, but that is hardly self-explanatory.<sup>107</sup> Given that they subscribe to many principles of evolutionary moral psychology, evolutionary biology and neuroscience, it seems logical to expand this discussion with relevant perspectives.

### 3. The Evolutionary and Moral Psychology of Emotions

Since the proponents of MBE subscribe to a significant number of ideas of evolutionary psychology (especially the idea of our maladapted moral psychology), it seems logical to evaluate their proposal of emotion modulation by comparing it with the view of emotions proposed by prominent evolutionary psychologists.<sup>108</sup>

---

<sup>106</sup> Other examples of psychological mechanisms causing lack of moral motivation in the current environment are: "(1) a bias towards the near future, according to which we heavily discount the importance of events in the more remote future, (2) an altruism which is restricted to kin and a small circle of acquaintances, (3) an incapacity to sympathize with larger numbers of people, (4) an act-omission doctrine, according to which it is harder to justify causing harm than letting harm occur which functions as a bar against the greater easiness of causing harm and (5) a conception of responsibility as causally based, according to which we are responsible for an effect in proportion to our causal contribution to it, so that our responsibility is proportionally diluted when we cause things together with other agents" (Persson and Savulescu, 2015a: 338).

<sup>107</sup> To be fair to the authors it must be said that they eventually express the doubt that emotion modulation will do the job necessary for moral enhancement. However, they don't elaborate the difference between emotions and mechanisms or the other problematic aspects of emotions which will be discussed further on.

<sup>108</sup> Although there are many relevant sources within this field, the focus of this paper will mostly be directed to a couple of capital works by J. Tooby, L. Cosmides and R. Nesse who enjoy a special

Tooby and Cosmides, for example, view emotions as “information-processing relations – that is, programs – with naturally selected functions” (2008: 116). They maintain that emotions evolved as a special class of biological adaptations (or programs) with the task to coordinate other adaptive psychological mechanisms (better known as *modules*) in order to promote fitness. Their general position is that the brain is a computer - a physical system that came into existence to carry out computations that were needed to solve the adaptive problem of regulating behavior. More specifically, emotions can be understood as the software of the mind – providing for the mind what software provides for the computer – adjusting its various parameters to the needs of a particular task (Nesse, 1990: 269).<sup>109</sup>

There is a consensus among psychologists, biologists and neuroscientists regarding the view that the human mental architecture is crowded with evolved, functionally specialized programs. Although these diverse programs were tailored by natural selection to solve different adaptive problems during our evolutionary past, their existence created adaptive problems of their own. The crucial problem was that they could, if activated simultaneously, conflict with each other and interfere with each other’s functioning. It is for this reason, according to Tooby and Cosmides (2008: 114-116), that our mind came to be equipped with another set of special, so-called, “superordinate” programs that performed the task of overriding, activating and deactivating other (subordinate) programs in specific environmental circumstances. Emotions perfectly fit the profile of such coordinating programs as they manage, harmonize and align other programs into the proper configuration at the right time.<sup>110</sup>

---

reputation in the field and their work perfectly fits the needs of this discussion. However, for more topic related overviews check e.g. Lewis, M. et al. (2008) or Al-Shawaf, L. et al. (2015).

<sup>109</sup> As one anonymous reviewer of this journal observed, the vexed relationship between the mind and the brain might be invoked. Since deeper analysis of this complex philosophical problem is not within the scope of this discussion, the reader is asked for a maximally neutral and naturalistic understanding of the terms in question: “The brain is a machine designed to process information [...] one can define the ‘mind’ as a set of information-processing procedures (cognitive programs) that are physically embodied in the neural circuitry of the brain. For cognitive scientists, ‘brain’ and ‘mind’ are terms that refer to the same system (Cosmides and Tooby, 2000: 97).

<sup>110</sup> Tooby and Cosmides explain that “sleep and flight from a predator require mutually inconsistent actions. When such or similar condition or situation of an evolutionarily recognizable kind is detected, a signal is sent out from the emotion program that activates the specific constellation of subprograms appropriate to solving the type of adaptive problems that were regularly embedded in that situation, and deactivates programs whose operation might interfere with solving those types of adaptive problems” (2008: 118).

We argue that such coordination is accomplished by a special class of programs: the emotions that evolved to solve these superordinate demands. In this view, the best way to understand what the emotions are, what they do, and how they operate is to recognize that mechanism orchestration is the function that defines the emotions, and explains in detail their design features. (Tooby and Cosmides, 2008: 117)

Although Tooby and Cosmides discuss emotions in general, very similar logic seems to apply to *moral emotions*. Evolutionary psychologist David M. Buss claims that moral emotions can be understood as “commitment devices” – serving to promote pro-social deeds, reparation of harm and punishment of cheaters (free riders), but also signaling to others that one is a good coalitional ally (2007: 402). According to Buss, each moral emotion was tailored to a specific kind of conduct (especially for solving certain groups of adaptive problems) and “although morality is sometimes viewed as being within the province of cognitive psychology, it clearly cannot be divorced from the social adaptive problems it evolved to solve” (2007: 402).

When it comes to moral emotions, the perspective of moral psychology is just as relevant.<sup>111</sup> Philosophers generally do recognize that moral emotions play a role in moral judgment and moral motivation, but such views are often presented with a high level of abstraction and without systematic empirical support (Prinz and Nichols, 2010: 111). As Prinz and Nichols point out, although there is a considerable disagreement about the exact role of emotions in moral cognition, there is a corresponding agreement that they are important for morality. The most important role of emotions, among others, is the one associated with moral motivation and promoting pro-social behavior.<sup>112</sup> In a nutshell, experimental results in the domain of cognitive psychology show that moral judgments can be influenced by manipulating emotions. For example, the emotion of *disgust* can be used to manipulate making negative moral judgments in the most unexpected ways. A study showed that subjects were inclined to give higher wrongness rates to various actions if they were placed at the filthy desk while filling out questionnaires. The same effect was obtained by using disgusting films and “fart spray” (Schnall et al.

---

<sup>111</sup> A profitable read on this topic is *The Moral Psychology Handbook* by J. Doris et al. (2010), especially chapters number two and four.

<sup>112</sup> However, it remains an open question whether emotions motivate us to act morally in the absence of a moral judgment or as a consequence of a moral judgment – i.e. moral emotions are emotions that are either constitutive of moral judgments or causally related to moral judgments in a special way (Prinz and Nichols, 2010: 119).

2008, as cit. by Prinz and Nichols, 2010: 114). Also, empirical research about emotions and moral judgments suggests that emotions can serve as “moral intuitions” or “heuristics”: “The basic idea is that, under some circumstances, we do not come to believe that something is, say, morally wrong by reasoning about it. Rather, we have an emotion, and the emotion leads us to judge that it is wrong” (Prinz and Nichols, 2010: 114).

Emotion theorists differentiate between several categories of emotions: pro-social emotions which promote morally good behavior (empathy, sympathy, concern, compassion), self-blame emotions (such as guilt and shame) and other-blame emotions (such as contempt, anger and disgust).<sup>113</sup> Prinz and Nichols explain the existence of so many different emotions along the same line as Tooby, Cosmides and Buss: each emotion has a different functional role (Prinz and Nichols, 2010: 122). More precisely, emotions have different roles corresponding to different moral norms.<sup>114</sup> Views from moral psychology suggest that emotions are essential for the preservation and practice of morality, but also that they drive our moral conduct in accordance with our evolutionary history.

Another group of authors claims that “moral judgment is the product of both intuitive and rational psychological processes, and it is the product of what is conventionally thought of as ‘affective’ and ‘cognitive’ mechanisms” (Cushman et al. 2010: 48). They argue that

the cognitive system operates by “controlled” psychological processes whereby explicit principles are consciously applied, while affective responses are generated by “automatic” psychological processes that are not available to conscious reflection. Thus, we suggest, the cognitive/affective and conscious/intuitive divisions that have been made in the literature in fact pick out the same underlying structure within the moral mind. (2010: 49)

This understanding of moral cognition resembles the evolutionary psychological views on cooperation between the mind’s various programs (such as emotions and

---

<sup>113</sup> Prinz and Nichols note that it is possible that empathy and sympathy are not emotions, but rather capacities that enable experiencing other people’s emotions. They, also, emphasize that *anger* and *guilt* count as the most important emotions in Western morality.

<sup>114</sup> For example: “Contempt arises when people violate *community* norms, such as norms pertaining to public goods or social hierarchies. Anger arises when people violate *autonomy* norms, which are norms prohibiting harms against persons. [...] [E]motions such as self-righteousness, gratitude, admiration, and elevation, [which] may serve as rewards for good behavior” (Prinz and Nichols 2010: 122).

mechanisms). There are similar neuroscientific explanations. For example, Crockett argues that although targeting moral motivation seems like a promising way for promoting moral behavior, “[f]uture research is needed to identify the specific types of motivational processes that contribute to moral behaviour, and to uncover their neurobiological mechanisms” (2014: 371).

It should be obvious that the role of emotions in human biology and moral cognition is highly complex, not to mention the indistinct relation between emotions and other mental machinery. Nevertheless, certain compatibility and prevailing agreement are evident. Biological, psychological and neuroscientific perspectives warrant the following points: (i) emotions are functionally specialized programs that evolved to promote fitness (evolutionary-biological perspective); (ii) emotions play an important role in moral cognition, driving moral decision-making in accordance with the basic evolutionary principles (perspective of moral psychology); and (iii) emotions have an interdependent relationship with underlying neurobiological mechanisms which play a substantial role in regulating human behavior (neuroscientific perspective). Bearing in mind these three points, what follows is the critical examination of the emotion modulation claim.

#### 4. Emotion Modulation and Its Problems

Contemporary debate on MBE has been swarmed with various aspects of what might be morally wrong with such a proposal, as well as what might go technically wrong in the case of its implementation. Most of these concerns express general skepticism regarding MBE, but some are also questioning the role of emotions in the MBE debate. J. Harris thus argued that “to believe that emotions can deliver answers to moral dilemmas or generate moral judgments is like believing that the gut is an organ of thought or one that can answer complex, combined theoretical and empirical, questions” (2013: 288). F. Jotterand commented: “neural moral enhancement focuses mostly on the manipulation of moral emotions. But moral judgments require more than the control of emotions, [...] while the manipulation of moral emotions might change the behavior of an individual, it does not provide any content, for example, norms or values to guide one’s behavioral response” (2011: 6).

Emotions are a questionable basis for morality, especially for improving moral judgment or moral decision-making. Bearing in mind our discussion so far, it could be argued that mental machinery which evolved to increase fitness is completely irrelevant in the present moral context. One could say that even if human

morality has a biological and emotional basis, this is not something worthy of promoting or enhancing. Moreover, maybe we should strive to eliminate such elements from moral reasoning as morally irrelevant or even detrimental.

It was also mentioned, however, that the majority of experts on emotions agree about their importance in moral reasoning and moral motivation. So, if emotions are involved in and relevant for morality, but at the same time biased and unreliable, how should we treat them? MBE proponents might be on the right track by suggesting to take better control of them, but at the same time, they are missing the bigger picture. It can be claimed that emotions are relevant, yet “empty” or “blind” for sophisticated moral reasoning. Therefore, MBE via emotion modulation would be in the same way “blind” and “blunt” intervention, downsizing rational decision-making.<sup>115</sup> To cope with this problem, a deeper understanding of emotions and their underlying cognitive mechanisms might be crucial. In the light of previously presented views, emotion modulation should be further critically examined.

#### 4.1. Emotion-Mechanism Correlation

It is generally assumed that it would be impossible to manipulate human mental machinery safely and reliably. According to evolutionary psychologists, emotions are highly complex, intricate, functionally organized and sensitively related to the structure of respective ancestral (evolutionary) problems. They are probably far more sophisticated engineering achievements than it may seem at first sight and many decades of work will be required before they are comprehensively mapped (Tooby and Cosmides, 2008: 134). It follows that, to implement MBE successfully, we should be familiar with specific functions of and interactions between capacities involved. Unfortunately, many relevant questions remain unanswered: To what extent do emotions control psychological mechanisms and what is the nature of that control? Does it consist only in activation-deactivation or does it go deeper than that? Would it be possible at all, through emotion modulation, to affect particular mechanisms of human psychology relevant for regulating moral behavior? Finally, would it be possible to affect them in the desired manner and without any negative side effects?

---

<sup>115</sup> This point can be associated with prevailing disagreement about what constitutes morality and which mechanisms comprise our moral psychology. Some aspects of this discussion can be found in Shook (2012) and Specker et al. (2014).

Even though the exact nature of the background interaction between various mental programs remains mostly unknown, proponents of MBE seem to assume the following hierarchy:

1. Mechanisms control behavior.
2. Emotions control mechanisms.
3. Therefore, emotions control behavior.
4. *Therefore, if we can control emotions, we can control behavior.*

At the most general level, this inference may be valid. Its premises, unfortunately, are far too imprecise to be of any practical use. The relevant point in this context, emphasized by Tooby and Cosmides (2008: 117), is that “emotions evolved to deal with a particular, evolutionarily recurrent situation type” and that “the design features of the emotion program, when the emotion is activated, presume the presence of an ancestrally structured situation type (regardless of the actual structure of the modern world)”. It follows that emotions reliably affect (control) mechanisms only in a limited number of specific situations (so-called *evolutionarily recurrent situations*).<sup>116</sup> Therefore, the second premise needs to be supplemented with the relevant data from the evolutionary psychology of emotions:

2. Emotions control mechanisms *only in a number of specific situations*.
3. Therefore, emotions control behavior *only in a number of specific situations*.

Another important perspective is that these specific situations, explained by evolutionary psychology, are not nearly the same as the situations described by advocates of MBE (regarding their complexity, predictability, reliability, etc.). For example, in a great number of ancestral cases of mate competition, sexual rivals could be driven off by violence. This is the reason why sexual jealousy was selected for: to prepare the body for possible combat and to motivate violent behavior. However, as Tooby and Cosmides emphasize (2008: 117), “in modern situations of potential or actual infidelity, police and prisons create additional consequences,

---

<sup>116</sup> What Tooby and Cosmides mean by *evolutionarily recurrent situation* in this context is “a cluster of repeated probabilistic relationships among events, conditions, actions, and choice consequences that endured over a sufficient stretch of evolutionary time to have favored some variant designs over others. Many of these relationships were probabilistically associated with cues detectable by humans, allowing psychophysical triggers to activate the task-appropriate program” (2008: 117).

and so violence against a sexual rival is likely to lead to maladaptive outcomes now.”

#### 4.2. Moral Emotions-Moral Reasoning Correlation

Moral emotions, as we have seen, drive our reasoning in accordance with certain archetypical scenarios faced by our ancestors. Therefore, if our “moral software” was programmed for radically different circumstances, betting on it might not always be a good idea. Morally relevant emotions such as *compassion* or *guilt* are triggered in a limited number of situations that resemble the evolutionary recurrent ones. This is why we are currently experiencing problems of insufficient care for the suffering of distant strangers and larger collectives. We are generally incapable to sympathize with distant strangers and we don’t feel particularly guilty about that, precisely because cognitive mechanisms of our moral psychology have been adapted to the environment where such problems did not exist. Living in small close-knit societies with limited impact on other people’s lives didn’t equip us for dealing with problems of the modern world. What is important to emphasize here is that moral emotions (just like ordinary emotions) operate on top of underlying psychological mechanisms – i.e. psychological mechanisms provide the instructions for desirable behavior and emotions serve as “short-cuts” for enabling faster reactions.<sup>117</sup>

The above conclusions combined suggest that maladaptive behavior in the current environment emerges from the outdated design of our emotion-driven psychological mechanisms. This implies that the enhancement of moral motivation via emotion modulation would not affect human moral behavior in a significant number of cases. In the best-case scenario, emotion modulation might positively affect moral motivation in a limited number of specific situations, such as the ones described by the evolutionary psychological notion of “ancestrally structured situations”. It seems plausible to conclude that proponents of MBE are right when they say that humans should biomedically speed up the process of their moral

---

<sup>117</sup> This understanding of moral emotions and underlying cognitive mechanisms corresponds well with the views that emotions serve as “moral intuitions” or “heuristics”. Whereas heuristics are generally understood as cognitive shortcuts enabling faster decision-making. Although they are very useful in most cases, they often lead to systematic errors in cognitive reasoning, i.e. to cognitively biased choices. Respectively, emotions govern automatic, intuitive judgments which can be unreliable rules in moral decision-making.



adaptation. However, the idea of tampering with emotions to boost the moral motivation seems problematic for at least two following reasons.

When it comes to the general claim that emotions can influence human behavior, that influence is limited to a specific number of situations resembling those from our evolutionary past. When it comes to the more specific claim that emotions are relevant for moral decision-making, then we face the problem of downsizing moral reasoning. The bottom line seems to be the unbreakable bond between emotions and underlying psychological mechanisms, suggesting that there are multiple factors involved in the process of moral cognition. Accordingly, emotions seem to be operating on an upper level of this computing system (second-order programs), whereas psychological mechanisms are basic instruction bearers (first-order programs). If my analysis of this problem is valid, the lack of moral motivation which is crucial in the MBE proposition is merely a manifestation of what appears to be a substantial problem of maladjusted mechanisms of human psychology. This conclusion implies that MBE should necessarily take into account the entire cognitive modules involved in moral reasoning. However, it is interesting to show what could be the possible outcome of implementing MBE via emotion modulation as it was initially proposed. There are at least two directions of such action: (i) solving the problem partially or even (ii) aggravating the initial situation. They will be examined in more detail in the following section.

## 5. The Outcomes of Emotion Modulation

Assuming that the problem of our suboptimal morality in the current environment lies in the outdated mechanisms of our moral psychology, it is likely that emotion modulation will not solve this issue comprehensively. At best, it might solve it partially for a limited number of cases and circumstances. As already argued, emotions can reliably affect mechanisms of human psychology only in a limited number of specific situations, e.g. those sufficiently resembling situations from our evolutionary past. Emotion modulation, therefore, is likely to be effective within such constraints. The problem with emotions as coordinating programs is that they do not make substantial changes to their subordinate mechanisms: they only manage, harmonize and align them into the proper configuration at the right time (Tooby and Cosmides, 2008: 116). In other words, if a substantial change in our morality is needed, emotion modulation will not do the job.

Since psychological mechanisms were adapted to living in specific conditions of our evolutionary past (small societies based on kinship and direct reciprocity),

they are designed to promote behavior that was profitable in such conditions. For example, if the emotion of compassion were to be enhanced, it would make a moral agent more compassionate in accordance with his/her disposition for evolutionary altruism.<sup>118</sup> Since disposition for evolutionary altruism comprises psychological mechanisms designed for behavior that was profitable in the distant past (such as temporal and spatial myopia), an enhanced agent would be motivated to act within such constraints. That means, being more compassionate or altruistic towards individuals that are near and dear to us or caring more about events in close future that are important to ourselves. This type of emotion modulation could be effective when the goal is to boost the moral motivation for the limited behavioral patterns. However, if the goal is to expand moral horizons beyond in-group reciprocal altruism, emotion modulation might not do the job.

On the other hand, it seems possible that emotion modulation could aggravate the current state of affairs. If we assume that emotion modulation would enhance motivation for particular moral actions (that is, actions that are already “natural” to us), then we would enhance motivation for exactly those behavioral patterns which were programmed into our moral psychology in the course of evolution. A moral agent with enhanced altruism will, therefore, become even more caring towards members of his/her in-group (such as kith and kin), but simultaneously he/she will become even less caring – or hostile – towards out-group members (such as strangers and distant individuals). If emotion modulation does not revise human psychological mechanisms from scratch and in a morally appropriate manner, but only enhances our motivation to act in accordance with our existing dispositions, this might actually boost human motivation for undesirable behaviors such as selfishness, nepotism, xenophobia or groupishness.

One could say that this view is opposed by recent pharmacological evidence and neuroscientific findings which suggest that emotional MBE might work. For example, hormone and neurotransmitter oxytocin has been shown to promote trust and cooperation; selective serotonin reuptake inhibitors (SSRIs) – like citalopram based anti-depressants - have been shown to increase cooperation and

---

<sup>118</sup> According to Stich et al. (2010: 154-5): “a behavior (or a behavioral disposition) is *evolutionarily altruistic* if and only if it reduces the inclusive fitness of the organism exhibiting the behavior and increases the inclusive fitness of some other organism. [...] It is logically possible for an organism to be evolutionarily altruistic even if it is entirely devoid of mental states and thus can’t have any ultimate desires”. It could be argued that MBE via emotion modulation would affect precisely this type of altruism (“devoid of mental states”) which does not seem as a sufficient foundation of morals.

reduce aggression; methylphenidate, mostly known as Ritalin, appears to reduce violent aggression when given to persons with Attention Deficit Disorder; some personality disorders which predispose for immoral behavior, have been linked to certain biological basis suggesting that if these conditions are better understood, interventions might be developed to improve behavior (Persson and Savulescu, 2008; Douglas, 2008).<sup>119</sup> It seems, therefore, that certain cases of manipulating human emotions are successful when it comes to making us behave more morally. What is then the ground for challenging emotion modulation as a means of biomedical moral enhancement?

Although experimental results indeed indicate that manipulation of biology can have moral effects, these effects could turn out to be undesirable in the bigger picture. Research on previously mentioned pharmacological evidence such as trust-promoting neurotransmitter oxytocin is twofold.<sup>120</sup> Some of the earlier experimental results showed that oxytocin can indeed increase trust and pro-social behavior (Zak et al. 2004; Kosfeld et al. 2005), but more recent experiments showed that it appears to be sensitive to group membership (De Dreu et al. 2010; De Dreu, 2011). This means that pro-social effects of administering oxytocin can be limited to in-group members, promoting in-group trust and cooperation, while simultaneously excluding out-groups and triggering defensiveness and territoriality towards out-group members and intruders (De Dreu et al. 2010).<sup>121</sup> De Dreu (2011) also showed that higher levels of oxytocin promote human ethnocentrism, i.e. oxytocin amplifies the intensity of trust and reciprocity within an already favored group (this can trigger discrimination by race or class). Nevertheless, some studies of oxytocin effects on human decision-making show strong context-

---

<sup>119</sup> An additional example of effective administration of pharmaceuticals in dealing with disorders which predispose for immoral behavior is that, e.g., anti-libinal drugs reduce sex drive in compulsive sex offenders. These drugs make it easier for persons to do the right thing, and resist doing the wrong thing.

<sup>120</sup> Similar results were found in studies on neurotransmitter serotonin. For example, Crockett et al. (2013) showed that serotonin regulates retaliatory motives in costly punishment. Depletion of serotonin reduces fairness and increases punishment for behavior directed towards oneself, but it can promote fair behavior in the group. This study also showed context-dependent effects of serotonin on social behavior. These findings have similar implications for the discussion on MBE as in the case of neurotransmitter oxytocin.

<sup>121</sup> This study also showed that in competing situations humans display parochial altruism because it had a strong survival function (De Dreu et al. 2010). Such example of limited or constrained reaction strongly supports the thesis about evolutionary biased tendencies which could become aggravated through emotion manipulation.

dependency (Bartz et al. 2011; Radke and de Bruijn 2012). This means that pro-social implications of oxytocin strongly depend on characteristics of the situation, as well as on stable characteristics of the individuals to whom it is administered. Finally, both of these studies suggest that context- and person-dependency requires investigation into the more basic psychological or biological mechanisms underlying the social effects of oxytocin.

It seems that pharmacological evidence does not refute the claim that MBE via emotion modulation could deepen the problem of undesirable behavior instead of expanding moral horizons beyond in-group reciprocal altruism. Moreover, it seems to support it. It is important to emphasize, however, that Persson and Savulescu do acknowledge that administration of hormones such as oxytocin would not by itself be effective, but that it has to go hand in hand with reasoning that “undercuts groundless moral differentiation” (2012: 119-120). This might entail the recognition that MBE via direct emotion modulation is not a comprehensive solution to moral problems of the modern world. A complete solution could come in certain forms of human enhancement such as genetic engineering (an often discussed example is fixing the MAO mutation of the X chromosome that has been linked to anti-social behavior). In this case, Persson and Savulescu recognize the need to fix both cognitive mechanisms and corresponding emotions – which could be a strategy compatible with the view presented in this paper. However, genetic enhancements by definition involve radical modifications which go beyond emotion manipulation and, for that reason, it is plausible to expect them to be more effective. This does not change the fact that emotion modulation as such might be problematic or insufficient.

## 6. Conclusion

The main intention of this paper was to expand the discussion about the basic ideas of MBE via emotion modulation through comparison with evolutionary-psychological and neuroscientific perspectives on emotions. One could claim, of course, that evolutionary psychology is a controversial theory and that being consistent or inconsistent with it is insufficient to count for or against any philosophical view.<sup>122</sup> However, the comparison with evolutionary psychology was made because proponents of MBE theory themselves subscribe to a significant number

---

<sup>122</sup> For some interesting critical observations of evolutionary psychology, see: Confer, J. C. et al. (2010) and Panksepp, J. (2000) or for more fundamental overview check Buss, D. (ed.) 2005.

of its ideas and theories. As it was shown, emotion modulation does not take into account the hierarchy between different mental programs and the way they operate. It was also argued that the emotional basis of morality plays an important role in human moral cognition, but it requires a more careful examination. The neuroscientific and evolutionary-psychological perspectives seem to indicate that emotions don't fulfill necessary requirements to serve as the vehicles of moral enhancement. If key features of our moral cognitive machinery appear to be insufficient or unreliable in the current environment, then their thorough, not just partial, revision is required. Therefore, if my analysis of this problem is valid, MBE should necessarily take into account the role and function of entire cognitive modules associated with moral decision-making.



A. Karametas, *Subservient* (2015)

## Chapter 5

# On the Uneasy Alliance between Moral Bioenhancement and Utilitarianism

**Abstract:** Moral bioenhancement (MBE) is often associated with a consequentialist, especially utilitarian, framework, due to its capacity to prevent great harm and motivate acts in accordance with basic moral principles such as universal impartial altruism or benevolence. However, it remains unclear whether we could de facto justify MBE on utilitarian grounds. This article examines whether there is a plausible utilitarian case for MBE and what the obstacles for justifying MBE on utilitarian grounds could be. More specifically, it explores the relationship between MBE and basic utilitarian principles, as well as its effects on utilitarian moral judgment. It seems that MBE could modify moral agents in ways that would accord with main utilitarian demands and facilitate the adoption and realization of utilitarian prescriptions. Although MBE would, in principle, create preconditions for achieving utilitarian ends, I argue there are certain limits to this claim. I identify and elaborate several ways in which MBE could undermine utilitarian moral judgment.

This chapter is published as: Kudlek, K. 2022. On the uneasy alliance between moral bioenhancement and utilitarianism, *Bioethics* 36, 210– 217. DOI: <https://doi.org/10.1111/bioe.12974>





# On the Uneasy Alliance between Moral Bioenhancement and Utilitarianism

## 1. Introduction

Enhancements are often associated with a consequentialist framework. A tacit assumption in the debate seems to be that they would be overall optimistic – that their benefits would outweigh the costs. For example, some authors suggest that the debate largely moves along standard consequentialist lines and often employs a utilitarian model of aggregation, usually some version of the rule-utilitarian maximization principle (Heinrichs and Stake, 2018). Others suggest that particular enhancements, such as moral bioenhancement (MBE), may even supplement utilitarian morality by motivating us to act in accordance with basic moral principles such as universal impartial altruism or benevolence (Persson and Savulescu, 2015b: 349). Namely, the aim of MBE to improve dispositions to altruism and a sense of justice overlaps with the basic demands of utilitarian moral theory (Persson and Savulescu, 2015b). Thus, not only is MBE discussed as a means of preventing great harm that threatens all the planet’s inhabitants, but also of turning us into better utilitarian agents. However, in order to determine whether MBE is in fact justified on utilitarian grounds, we need to thoroughly examine its relationship with basic utilitarian principles, as well as its effects on moral agents and utilitarian moral judgment. Therefore, my goal here is to make a ruling on whether there is a plausible utilitarian case for MBE and what the potential obstacles are to justifying MBE on utilitarian grounds.

MBE seems not merely permissible, but desirable on utilitarian grounds because it improves psychological dispositions of people to *act* like utilitarians. Utilitarianism is famous for demanding radical impartiality and strong benevolence in moral agents. It urges us to prevent bad things from happening whenever this is in our power (Singer, 1972). However, most people fail at adopting and following these principles due to a lack of moral motivation, and “[a]s many of us are incapable of complying even with commonsense morality, and the proposed consequentialist extension or revision of it is more demanding, it is obvious that compliance with this extended morality will require an enhancement of the moral motivation of many of us” (Persson and Savulescu, 2015b: 349). MBE promises to

boost our moral motivation, making us more compliant with this extended morality, and helping us to deal with the pressing issues of the modern world (Persson and Savulescu, 2012; 2015b). It is expected to modify our commonsense morality to the extent of counteracting the features of our moral psychology that are commonly associated with decisions resulting in bad outcomes, such as “the limitation of our altruism to those who are near and dear” and “the numbness to larger numbers of sufferers” (Persson and Savulescu, 2012: 123). To put this in utilitarian terms, it seems that MBE will promote good consequences by reinforcing the duties recognized by commonsense morality (Persson and Savulescu, 2015b).

In my attempt to explore the possible justifications of, and obstacles to MBE on utilitarian grounds, I mainly focus on Persson and Savulescu’s proposal for motivational MBE, which suggests improving particular features of commonsense morality.<sup>123</sup> Nevertheless, I consider the likely need for a more sophisticated kind of MBE – one that would encompass both motivational and deliberative capacities – and how such a possibility would resonate with utilitarian standards. My analysis consists of two main parts. First, I make a *prima facie* utilitarian case for MBE by looking at how MBE fits with basic utilitarian principles – such as impartiality and utility maximization. I examine whether MBE can make us better utilitarian agents and whether it is indeed *optimific* – such that it would yield greater benefits over costs. I show that MBE roughly meets basic utilitarian demands – enhancement of altruism and a sense of justice could contribute to the overall good. Second, I argue that MBE might not *always* be *optimific*, due to limitations we need to take into account. I argue that MBE’s effects on utilitarian moral judgment would not be reliable and systematic but a matter of chance, and could, therefore, undermine the best utilitarian outcomes.

I look into three particular concerns: a) enhancing commonsense morality may create tension between utilitarianism and MBE through fostering intuitive instead of rational reasoning; b) enhancing altruism may promote partial instead of universal and impartial concern due to the phenomenon of parochial altruism; c) enhancing a sense of justice may lead to prioritizing just over utility-maximizing actions. These concerns show that although there is, in principle, a legitimate case for MBE, it comes with certain limitations. As I will note, there may be ways to tackle these concerns, but the strength of these concerns will still depend on different accounts of utilitarianism and the kind of MBE applied (such as

---

<sup>123</sup> I also take the feasibility of this proposal at face value. i.e., I do not discuss whether such enhancement is possible, but how we should morally evaluate it if it were possible.

direct/indirect utilitarianism and motivational/deliberative MBE). My take is that direct utilitarians are unlikely to support MBE, whereas indirect utilitarians may have good reason to do so.

## 2. The Utilitarian Case for MBE

According to its proponents, MBE is expected to increase the likelihood that we correctly estimate the right thing to do and act upon it. However, the estimation of what constitutes the correct action will depend on personal beliefs and preferences:

To be morally enhanced is to have those dispositions which make it more likely that you will arrive at the correct judgement of what it is right to do and more likely to act on that judgement. It is disputed what the right thing to do is and how we would arrive at the right course of action. What constitutes moral enhancement will depend on the account one accepts of right action. (Savulescu and Persson, 2012: 406)

In order to understand what this entails for utilitarian morality, we can start by examining whether the ends and means of MBE are right/permissible on utilitarian grounds. Thus, in this section, I will examine (i) how MBE affects moral agents and their actions (whether it promotes utilitarian ends), and (ii) whether the act of enhancement itself is right or permissible on utilitarian grounds (whether the means of MBE are acceptable). First, I look into MBE's correspondence with basic utilitarian principles and show that it could modify moral agents in ways that would *indirectly* facilitate utilitarian ends. Second, I explore the conditions that MBE would need to satisfy to be optimific, and I argue that there are good reasons to believe it would meet these requirements.

### 2.1. Making Better Utilitarian Agents?

Advocates of MBE envision this type of moral betterment as an extension of duties recognized by commonsense morality because such an approach may have the best overall consequences. "Folk" or "commonsense" morality is a globally shared set of moral attitudes that are "a common denominator of the diversely specified moralities of human societies over the world" (Persson and Savulescu, 2012: 12). It amounts to "a set of psychological dispositions to react in particular ways in certain types of situations" (Gligorov, 2018: 370). MBE is supposed to modify these dispositions. To fix some of the reoccurring flaws of moral psychology,

Persson and Savulescu propose “a rather modest extension of commonsense morality, an extension which puts greater emphasis upon duties that commonsense morality already recognizes” (Persson and Savulescu 2012: 123). MBE is supposed to strengthen pro-moral emotions (sympathy, cooperation, etc.) or, alternatively, diminish counter-moral emotions (racial aversion, violent aggression, etc.) (Douglas, 2008).<sup>124</sup>

Although commonsense morality typically exhibits deontological features, proponents of MBE are confident that it will produce the best overall consequences:

For it may be that the structure of commonsense morality is so deeply embedded in our nature that it will have best consequences in terms of our underlying consequentialist theory if we try to live by something akin to commonsense morality, somewhat revised to be better aligned with the underlying consequentialist theory. (Persson and Savulescu, 2015b: 349)

They continue:

Many consequentialist theories nowadays have this sort of two-level architecture: a ground level of true consequentialist principles—in our case about universal beneficence and justice—and a superstructure of literally false commonsensical principles, the endorsement of which is pragmatically justified by its good consequences, as determined by the ground-level consequentialism. (2015b: 349)

Let us grant this *arguendo*. It follows that MBE will, through the reinforcement of duties recognized by commonsense morality, likely promote an underlying consequentialism or the maximization of utility. Such an outcome seems to be in line with some kind of indirect utilitarianism (such as rule or motive utilitarianism), which instructs us to follow particular rules, motives, virtues, etc., because they have been shown to maximize utility in the long run (Brandt, 1992; Sverdlik, 2011). Unlike direct utilitarianism, indirect utilitarianism does not place so much normative weight on the outcomes of actions – rightness depends on whether the action follows from rules and dispositions of character that are themselves utility-maximizing (Alexander, 1985). For example, rule and motive utilitarianism determine an action’s rightness by its conformity to the best set of rules or by the motivations with which it is performed (Woodard, 2019). Since MBE amounts to

---

<sup>124</sup> Although diminishing counter-moral emotions is as relevant as boosting pro-moral emotions, I will limit my analysis to the latter. One reason is the strong utilitarian focus on active contribution to overall well-being.

improvements of the agent's character, it is more likely to be justified on indirect utilitarian grounds, whereas the rightness on direct (act) utilitarian grounds should be determined directly from the improvement's consequences.

Moreover, strengthening altruism and a sense of justice could facilitate the adoption of the standard utilitarian principles of impartiality and utility. MBE could broaden the scope of our moral concern, make us more likely to maximize utility, and less likely to harm others. By always urging us to maximize the welfare of everyone affected by our actions (Bentham, 1948; Mill, 1864), utilitarianism presupposes strong altruistic and benevolent character traits in a moral agent: “[it] requires him to be as strictly impartial as a disinterested and benevolent spectator” (Mill, 1864: 24). Mill also held that impartiality is an obligation of justice that can be influenced only by proper considerations (while resisting other motives). Impartiality is closely related to equality and, therefore, “one person's happiness (...) is counted for exactly as much as another's” (Mill 1864: 93). Hence, if MBE makes us more altruistic, impartial and just, it will create necessary preconditions for facilitating utilitarian ends.

One of the MBE project's main agendas is to broaden our moral concern beyond the limits of kith and kin, to include those in need (Persson and Savulescu, 2012). A successful MBE is expected to override various biases that seem to be hard-wired into our moral psychology (such as selfishness, nepotism, xenophobia, groupishness, etc.), which often cause us to behave in morally undesirable ways. The utilitarian principle of impartiality dictates precisely that “we broaden our concerns so that we are not focused just on ourselves, or on our friends, family or fellow citizens” but on “everyone whose well-being may be affected by our actions” (Shafer-Landau, 2012: 125). Hence, not only does MBE accord with the classic utilitarian principles of agent-neutrality and impartiality, but it creates the grounds for promoting the kind of behavior desired/favored/advocated by utilitarians.

Utilitarian moral theory also gives weight to reducing pain and suffering of all sentient beings. One crucial aspect of MBE is preventing harm in terms of reducing crime and violent behavior, which contributes to minimizing pain and suffering in the utilitarian sense.<sup>125</sup> Preventing harm is also nothing but avoiding bad consequences, and MBE proponents deem that goal (including the means to achieve it) justified similarly as utilitarians justify certain actions in terms of the goodness of outcomes. Thus, MBE can be understood as a means of preventing

---

<sup>125</sup> Preventing harm can be brought about by strengthening altruism, but also by diminishing counter-moral emotions like racial aversion and violent aggression (Douglas, 2008).

harm. In broad terms, utilitarianism requires us to adopt *any* means that will maximize the sum total of welfare in the world.<sup>126</sup> If we can reasonably expect that the means of biomedical enhancement would maximize the sum total of welfare, there would be no reason for utilitarians to oppose enhancement.<sup>127</sup> I will now explore whether MBE as an intervention could be justified on utilitarian grounds.

## 2.2. Is MBE Optimific?

Aside from promoting desirable dispositions of character, utilitarians might be more interested in how MBEs score on the utilitarian calculus. Generally, if an action is to be morally justified on utilitarian grounds, it must be optimific – such as to yield the greatest balance of benefits over costs (Rachels, 2003). In other words, the rightness of actions is explained in terms of the goodness of outcomes.<sup>128</sup> For example, act consequentialists hold that an action is right if it results in an outcome which is “at least as good as that of any relevant alternative” (Woodard, 2019: 4).<sup>129</sup> Utilitarians hold that one outcome is better than another if it contains more well-being. Thus, MBE would be right on utilitarian grounds when it creates more (or no less) well-being than any available alternative. In terms of an agent’s well-being, MBE is good insofar as it promotes value within an

---

<sup>126</sup> Standard consequentialist rationales instruct us that motives and intentions are not directly relevant for an action’s rightness; the only thing that matters are the consequences (Shafer-Landau, 2012). According to utilitarians, we can only know that our intentions are right, but the rightness of the action is evaluated only after it has been performed and its consequences became evident. Utilitarians differentiate between the morality of actions and intentions – intentions are irrelevant to the morality of actions (see Shafer-Landau, 2012:124). There are, of course, some exceptions among consequentialist accounts (e.g., motive consequentialists give special weight to agent motives).

<sup>127</sup> By contrast, deontologists may be more reluctant about the permissibility of biomedical means because they might be concerned about their intrinsic properties.

<sup>128</sup> This is one of the three defining features of utilitarianism. Following Sen (1979), Woodard (2019) states that consequentialism explains the rightness of actions in terms of the goodness of outcomes. Welfarism posits that all well-being is intrinsically valuable, whereas everything else is valuable in terms of its contribution to well-being. Sum-ranking explains the value of an outcome in terms of the sum of the goods (and bads) within that outcome. Utilitarians are committed to these claims and they hold that one outcome is better than another if it contains more well-being.

<sup>129</sup> Alternatively, rule consequentialists claim that the right action is permitted by the best set of rules, and motive consequentialists believe an action is right if it is performed by someone with the best motives (Woodard 2019: 4).

agent's life and bad insofar as it inhibits it. But we want to determine here whether MBE itself or *on the whole* is the best possible action compared to its alternatives.

Commonly discussed alternatives to MBE are, for example, the preservation of the status quo, traditional moral enhancement, and cognitive enhancement.<sup>130</sup> And although MBE faces fierce criticism, one could argue that the currently conceivable alternatives are not likely to score any better on the utilitarian calculus. Following basic utilitarian rationales, MBE would be morally right if we could reasonably expect its outcomes to be optimistic, i.e., that none of the available alternatives result in a larger sum total of welfare in the world. The ultimate motive for implementing MBE is not the quest for perfection, but the urgent need to deal with the pressing issues of the modern world, such as environmental degradation and threats of terrorist attacks (Persson and Savulescu, 2012). Hence, the right action has to produce the greatest balance of benefits over costs to urgently prevent catastrophic harm, in view of the lack of moral motivation and immoral behavior.

The preservation of the status quo is often advocated by bioconservative authors who believe that maintaining things as they are is better and safer than pursuing new and unfamiliar projects such as MBE (e.g., Sandel, 2007; Kass, 2003). New projects are often risky, controversial, and may not be worth our while – or in this context, the risks might outweigh the benefits. This view also implies that the quality of our lives is satisfactory as it is. However, biomedical enhancements might become necessary for *sustaining* the status quo rather than improving our situation – to prevent a decline in our average quality of life (Buchanan, 2011b: 76). The threat of decline (such as by environmental degradation) cannot be prevented by mere inaction. In other words, sticking to the status quo is likely to make matters worse, and will surely inhibit our well-being.<sup>131</sup>

---

<sup>130</sup> This is not an exhaustive list of alternatives to pressing issues in the enhancement debate. However, the listed possibilities are among those most commonly discussed. For a more detailed discussion on these topics, see Sandel (2007), Buchanan (2011b, Ch3), and Sandberg (2011).

<sup>131</sup> One could argue that bioconservative arguments apply regardless of whether they produce good consequences, since they supposedly indicate an intrinsic wrongness in enhancements. The only thing intrinsically valuable for utilitarians is well-being, while everything else is valuable in terms of contributing to well-being (for a more in-depth discussion, see Woodard [2019 pp.16-7, Ch. 4]). Hence, MBE would have to undermine well-being (or contributions to it) to the extent that all other considerations are outweighed. I believe this is highly unlikely. We do not have plausible reasons to believe enhancements are intrinsically bad, regardless of whether we evaluate their intrinsic properties or necessary consequences. I offer a detailed account of this in Kudlek (2021), here Chapter 3.

Traditional moral enhancement resembles the preservation of the status quo, since it amounts to already established, noninvasive means of moral improvement. Although considerable moral progress has been achieved throughout history, traditional moral enhancement works too slowly and is, therefore, insufficient in the face of the need for urgent solutions (Persson and Savulescu, 2012). For traditional moral enhancement methods to be sufficient, we would already need to be morally motivated to a significant extent (Persson and Savulescu, 2008: 168). Even if they were significantly more advanced than they currently are, the traditional means (like moral education, upbringing, and socialization) would hardly be successful in boosting our motivation for dealing with pressing issues, or in preventing harm on a larger scale. Thus, traditional moral enhancement is unlikely to result in outcomes better than those expected by MBE.

Cognitive enhancement might have a more immediate effect on gaining the relevant knowledge and skills to deal with pressing issues. However, “we do not necessarily do what is right and good as soon as we gain knowledge of what this is” (Persson and Savulescu, 2008: 168). Some aspects of cognitive enhancement might be necessary as one ingredient of effective moral enhancement, but it is not likely that cognitive enhancement would be sufficient on its own for reaching morally desirable improvement. Most importantly, the risk of abuse appears much greater with cognitive enhancement than with MBE because the former directly benefits the enhanced, but may disadvantage the unenhanced.<sup>132</sup> MBE, on the other hand, is expected to benefit others (Douglas, 2008: 230).

Insofar as there is no better alternative for meeting pressing issues, MBE could be optimific on the utilitarian calculus. This analysis takes the feasibility of MBE at face value and presupposes it would achieve its stated goals. But even if it would have negative side-effects, such as harming the moral agent by undermining their freedom or autonomy, MBE may still be optimific. This is because utilitarianism places no constraints on the maximization of utility – it permits us to cause harm for the sake of maximizing utility. Moreover, if MBE is optimific, and thus right from the utilitarian point of view, its implementation is required; not conducting the right action counts as moral wrongdoing on utilitarian grounds (Shafer-Landau, 2012).

---

<sup>132</sup> This primarily applies to competitive environments like sports and work, in which cognitive enhancements are associated with positional goods – goods whose value is dependent on others not having them (Sandberg, 2011: 83).



Thus far, I have identified some conditions under which MBE would be justified on utilitarian grounds. It seems that not only is MBE in line with basic utilitarian principles, but it could indirectly maximize utility by modifying certain dispositions of character. Also, as far as MBE is our best bet (optimific) for solving the pressing issues of the modern world, utilitarians should fully support its implementation. However, there might be some limits to these claims. In the following section, I raise several concerns and identify conditions under which MBE could undermine utilitarian judgment.

### **3. The Utilitarian Case against MBE**

The motivational model of MBE that I have described may have effects on moral agents and their judgment that could be problematic from the utilitarian point of view. I will address three such concerns in this section. Some of these concerns have been acknowledged in the broader debate, but their implications for utilitarian morality are yet to be discussed. I start from some conceptual difficulties regarding utilitarianism and commonsense morality, which can create tension between utilitarianism and MBE. I then examine how specifically the enhancement of altruism and a sense of justice could conflict with utilitarian demands for impartiality and the maximization of utility.<sup>133</sup> I also take notice of how these concerns could be mitigated.

#### **3.1. The Tension between Utilitarianism and Commonsense Morality**

One possible source of tension between MBE and utilitarianism is their arguably different relation to commonsense morality. Namely, MBE tends to extend parts of commonsense morality (as explained previously), whereas utilitarianism departs considerably from this moral framework. If so, it follows that MBE departs from utilitarianism, and should therefore be alarming to utilitarians. I noted at the beginning that MBE is conceived as a modest extension upon the globally shared set of moral attitudes that are based in biology and often referred to as folk or commonsense morality (Persson and Savulescu, 2012). On this view, commonsense morality amounts to a set of psychological dispositions for reacting in

---

<sup>133</sup> These concerns are not meant to be exhaustive of all problems utilitarians may have with MBE. However, they do cover a range of fundamental questions that MBE raises.

certain ways (Gligorov, 2018); MBE would boost either these dispositions, or the commonly shared intuitions about what is the morally right thing to do.

Conversely, utilitarianism is typically regarded as a view rooted in rational reflection and conflicted with many commonsensical intuitions. The strongest opposition to utilitarianism is, in fact, due to its clash with common moral judgment and intuitions (Singer, 2005; Rachels, 1991). For example, Kahane et al. describe two ways in which utilitarianism radically departs from commonsense morality: i) utilitarianism places no constraints on the maximization of aggregate well-being and ii) requires us to maximize the well-being of all sentient beings (2018: 132). By placing no constraints, utilitarianism permits us to harm others for the sake of maximizing well-being. Such reasoning often conflicts with commonsense intuitions. Hand in hand with counter-intuitiveness goes the problem of demandingness. By requiring universal and impartial maximization of well-being, utilitarianism strikes many as too demanding (Kagan, 1984; Cullity, 2003). Thus, utilitarianism can permit too much and demand too much, and is, therefore, in conflict with our common moral judgments. Unlike utilitarian views, common moral views have their source in gut reactions and intuitions shaped by evolutionary pressures (Kahane et al. 2018).

Biologically based moral attitudes, which are altered via MBE, are often regarded as quick intuitive judgments in moral psychology. Such intuitive judgments (typical for commonsense moral reasoning) often lead to conclusions favored by deontological and rights-based doctrines, whereas careful rational reflection was shown to produce more utilitarian-friendly conclusions (Haidt, 2012; Greene, 2007; Paxton, Bruni and Greene, 2014). Assuming that utilitarianism radically departs from commonsense morality and that MBE is conceived as extending/building upon the latter, it follows that MBE radically departs from utilitarianism. If MBE fosters commonsensical reasoning, utilitarians, who tend to prefer a radically rational decision-making approach, could rightfully doubt whether such an intervention is overall optimistic.

This concern may be mitigated. For example, it has been argued that motivational MBE will have to be supplemented by the enhancement of deliberative capacities (cognitive enhancement) in order to achieve the desired aims (Persson and Savulescu 2012; Earp et al. 2018). If this were the case, we might expect a more balanced approach to enhancing moral decision-making that would not

necessarily undermine rational deliberation.<sup>134</sup> Hence, as far as MBE fosters intuitive thinking, utilitarians would find it problematic. But even under these circumstances, not all utilitarians need to think alike. Another possibility is to resort to the previously discussed view that commonsense morality is *de facto* consequentially justified on *indirect* utilitarian grounds because (as I show in Section 2) it would boost behavioral tendencies that maximize well-being in the long run.<sup>135</sup> If commonsense moral tendencies can be justified on utilitarian grounds, then there was no conflict between commonsense morality and utilitarianism to begin with. Instead, they may form some sort of friendly alliance (Sidgwick, 1907).<sup>136</sup> I, however, believe that the gap between commonsense morality and the utilitarian doctrine is further reflected in the practical aspects of enhancing altruism and a sense of justice.

### 3.2. The Problem of Partial Altruism

On a more practical level, enhancing the psychological disposition to altruism could undermine impartial maximization of utility due to the natural tendency towards partial or parochial altruism. Although this concern has been noted in the bioenhancement debate (e.g., Persson and Savulescu, 2012), it has not been previously raised in discussions about utilitarianism. My goal here is to examine the

---

<sup>134</sup> Although the discussion has recognized that motivational enhancement might not be sufficient to achieve desired moral betterment on its own, it is still discussed as a legitimate “standalone” type of MBE. One reason for this might be that a more sophisticated MBE may be overly speculative at this point. In addition, the distinction between direct and indirect MBE captures the difference between motivational and deliberative MBE to some extent (Schaefer 2015; Earp et al. 2018).

<sup>135</sup> Even if MBE can be justified on indirect utilitarian grounds, it could still be problematic from a direct utilitarian perspective. Act (outcome) utilitarians would evaluate MBE by its immediate consequences.

<sup>136</sup> In philosophical theory, the relationship between utilitarianism and commonsense morality can be seen as a friendly alliance because commonly accepted moral rules (often referred to as ‘rules of thumb’) usually tends to maximize happiness/utility (Sidgwick, 1907). In the context of MBE, however, a biological approach to commonsense morality (which implies their conflict) is more relevant and, therefore, given more attention in this discussion. Some have argued that commonsense morality can be consequentialized (see Portmore [2011], especially Chapter 4). Regardless of whether the relationship between utilitarianism and commonsense morality should be characterized as one of alliance or dissent, MBE (extending upon commonsense morality) can cause problems for utilitarians. For some views about how MBE could disrupt the valuable balance between commonsense morality and utilitarianism, see Agar (2015b).

implications of partial altruism for utilitarian morality. To have an attitude of altruism and sympathy towards other beings is to want things to go well for them for their own sake (Persson and Savulescu, 2018). To enhance the capacity for sympathy would then mean to enhance the probability that we will do what we believe we ought to do (in response to reasons for it). In this sense, MBE can be interpreted as a suggestion to make human beings, on average, more prone to altruism and sympathy as the most moral people among us already are.<sup>137</sup> In practice, this would mean that a person with enhanced altruism will have a stronger tendency to perform altruistic acts. For example, a person who already gives money to charity will now give even more. Insofar as we believe that giving to charities is optimistic, but that there are psychological constraints on our ability to sacrifice our well-being for the sake of others (Singer, 1972), MBE can bring us closer to the utilitarian ideal.

However, since the natural attitude of sympathy tends to be partial, it may conflict with utilitarian demands for impartiality:

It is well-known that the attitude of sympathy, as it occurs spontaneously, tends to be partial: we tend to sympathize in particular with members of our family, friends, and people before our eyes. (...) Utilitarianism, which takes sympathy or altruism to be the one and only fundamental moral attitude, opposes this partiality, by declaring in its most familiar form, roughly speaking, our moral goal to be to see to it that things go as well as possible for as many as possible. (Persson and Savulescu, 2019: 8)

It turns out that we naturally tend to sympathize *more* with members of our in-group like friends and family, while our willingness to cooperate or sympathize with out-group members tends to be reduced and hostility towards them increased.<sup>138</sup> Increasing altruistic emotions in human beings in order to increase the probability that one does what one believes to be right could, therefore, *further* strengthen in-group, instead of out-group altruism. Such side effects were reported by several studies about the administration of oxytocin and serotonin, which are typically expected to increase altruistic tendencies in humans (see, e.g., De Dreu et al. 2010; De Dreu 2011; Crockett et al., 2013). Hence, strengthening our psychological disposition to altruism could indeed promote moral concern for

---

<sup>137</sup> One example are higher average levels of sympathy in women as compared to men (Baron-Cohen, 2003).

<sup>138</sup> I offer a more detailed account of this concern in Kudlek (2019) [Chapter 4].

members of our close group, but the concern may not necessarily become more extended beyond that group.

By virtue of the phenomenon of partial (or parochial) altruism, the effects of MBE might be limited in scope, and utilitarians might find this troubling. Since utilitarianism places significant normative weight on agent neutrality, enhancement of altruism could, under some circumstances, undermine the principle of impartiality. For example, parents with enhanced altruism might prefer marginal increases in their children's well-being over massive increases in the well-being of a great number of other children (see, e.g., Drake, 2016). As long as the enhancement of altruism promotes only in-group cooperation and sympathy, its potential to prompt utilitarian judgments among agents will be limited.

However, promoting in-group altruism can be consistent with utilitarianism when it maximizes the overall good for all affected parties. Utilitarians will sometimes judge MBE to have morally desirable effects even when the consequences of actions are strictly limited to one's group. Imagine a negligent parent who undergoes enhancement to become more altruistic towards her own children. In this case, the enhancement of "partialist" altruism would be compatible with utilitarian demands because all affected parties belong to one's group. Also, if it turned out to be possible for MBE to go "hand in hand with reasoning which undercuts race, sex etc. as grounds for moral differentiation" (Persson and Savulescu, 2012: 120), then this concern would be alleviated. Another solution could be a distinction between "utilitarian" and "utile", where utile implies doing more good for others even if it is a restricted set of others. Apart from cases where utilitarianism allows certain forms of partiality, the enhancement of altruism would not necessarily facilitate utilitarian ends.

### 3.3. Prioritizing Just over Utility-Maximizing Actions

To solve the problem of partialist altruism and to supplement the principle of utility with some principle of justice, MBE advocates postulate another moral attitude – a sense of justice or fairness in moral agents (Persson and Savulescu, 2015b: 349). As a moral theory, utilitarianism is often criticized for being unable to take matters of justice into proper account (Persson and Savulescu, 2015b: 349), since maximizing overall well-being may sometimes strike us as unjust or unfair (Persson and Savulescu, 2019: 8). Therefore, "the moral goal is not just that things go as well as possible for as many as possible, but that how well things go for different beings be as much as possible in line with justice" (Persson and

Savulescu, 2019: 8). In short, MBE is expected to “motivate us to act in accordance with plausible basic moral principles” (Persson and Savulescu, 2015b: 349), like impartial altruism and benevolence, which happen to overlap with basic utilitarian principles.

But justice is an exceptionally complex and controversial notion, and there is little consensus on what it consists in. I will understand it here as a desire to do what one thinks is just or fair, which, in turn, makes the enhancement of one’s sense of justice simply the enhancement of one’s desire (or the motivation and probability) to do what one regards as just (Persson and Savulescu, 2019). Authors acknowledge that our common sense is “firmly wedded to deserts and rights,” and that our sense of justice originates in reciprocity and tit-for-tat strategies: justice requires getting what you deserve (Persson and Savulescu, 2019). So, instead of balancing out the partial altruism problem, the enhancement of a sense of justice can create additional problems for utilitarians.

As with altruism, the biological sense of justice does not necessarily align with the utilitarian sense. According to the utilitarian conception of justice, just actions are those that maximize utility. For example, the justice of institutions requires that institutions maximize utility (Woodard, 2019).<sup>139</sup> I mentioned previously that most people are not naturally inclined to utilitarian moral reasoning – we often find it counterintuitive and unjust. Thus, one could argue that a moral attitude that is supposed to ensure that things are in line with justice can, in some circumstances, undermine or conflict with the willingness to maximize utility. For example, we often have difficulty sacrificing an innocent person’s well-being for the sake of maximizing utility. If you have a strong intuition that it is not fair to sacrifice one for the sake of many, boosting your sense of justice could only deepen this inclination. Hence, boosting a biological sense of justice and fairness (i.e., enhancing the desire to do what we regard as just) might reduce the likelihood of arriving at the correct utilitarian judgment, as well as acting upon it.

Even if utilitarians conceive of the ‘good’ and the ‘just’ independently from one another, these values may conflict given that it is not always possible to maximize the good and act justly, all at once. Sometimes you have to lie to your friends to make them feel better. In such cases, act utilitarians are required to always maximize the good, even when that means committing unjust acts. Indirect

---

<sup>139</sup> The lack of a plausible distinction between the notions of justice and goodness is a common problem for utilitarian moral theory. For an in-depth discussion on this issue, see Woodard (2019), Chapter 7.

utilitarians, however, acknowledge that, although utility is the ultimate foundation for morality, the direct pursuit of utility is not always the best course of action (Alexander, 1985: 316). Therefore, they conform to the rule that lying does not pay off overall. Now, an enhanced sense of justice would entail a stronger inclination to do what we deem just, regardless of what directly maximizes utility. This means that enhanced agents will be more likely to follow indirect than direct utilitarian prescriptions. In other words, MBE is more likely to be justified on indirect than direct utilitarian grounds.

An enhanced sense of justice can accord with utilitarian demands in a limited capacity or when supposedly just acts have been shown to maximize overall utility. But this would be merely a matter of chance – there is no direct connection between MBE and utilitarian outcomes. In other cases, MBE can interfere with maximizing utility – it can make us favor just over utility-maximizing acts (assuming the notion of justice does not coincide with maximum goodness). The enhanced sense of justice can also interfere with the utilitarian demand for altruism – it can make us favor just over altruistic acts (when altruism would be preferable from a utilitarian perspective). Simultaneous enhancement of a sense of justice along with the capacity for altruism will not resolve any of these issues. On the contrary, it may only deepen them because moral agent will experience stronger preferences or a stronger conflict of preferences.

A remedy to these concerns could once again lie in the further development of technology that would fine-tune the enhancement of motivational and deliberative capacities. If the technology can become sophisticated enough, some of the aforementioned problems could be avoided. Whether this is indeed possible is, however, a separate question. From where we are standing now, tampering with parts of commonsense morality (specifically the capacities for altruism and justice) may, under the described circumstances, compromise reaching the best utilitarian judgment. How strict these limitations will be will greatly depend, as already mentioned, on the different notions of utilitarianism and different approaches to MBE. But it seems safe to conclude that direct utilitarians are less likely to support MBE, whereas indirect utilitarians may have good reasons to do so.

#### **4. Conclusion**

My main goal here was to explore whether there are legitimate reasons to think that MBE is justified on utilitarian grounds. My analysis showed that there is indeed a *prima facie* case that MBE would be appealing to utilitarians. It seems to

roughly accord with utilitarian demands and principles, especially in terms of expanding moral concern and adopting agent-neutrality. We could deem MBE optimistic as far as it is our best solution to urgently prevent harm caused by the lack of moral motivation and immoral behavior. However, there are important limits to this claim. I argued that one such constraint is the well-known tension between utilitarianism and commonsense morality. Since utilitarianism places great normative weight on radically rational mode of decision-making, it might not favor enhancing quick intuitive judgments typical for commonsensical reasoning. This problem is further reflected in the potential to promote partial altruism and prioritize just over utility-maximizing actions, which could compromise utilitarian moral judgment. Although these concerns may place tight constraints on justifying MBE on direct utilitarian grounds, they apply with lesser strength in the case of indirect utilitarianism and (as of yet hypothetical) sophisticated types of MBE, such as the fine-tuned enhancement of motivational and deliberative capacities.







A. Karametas, *Dilemma* (2014)

## Chapter 6

# The Kantian Promise and Peril of Moral Bioenhancement

**Abstract:** Moral bioenhancement (MBE) aims to fix our moral agency itself in order to prevent us from engaging in negligent or harmful behavior. Although such (self)-paternalistic practice might very well produce good outcomes, it can be argued – inspired by Kant – that it is intrinsically disrespectful towards our future agency. Hence, we are faced with the following ethical dilemma: the failure to engage in MBE seems reckless and negligent, which can be considered a serious moral wrong; but engaging in MBE presupposes that we treat our faculties (our future agency) in disrespectful self-paternalistic manner. In this paper, we want to resolve the described dilemma by suggesting a novel way of understanding Kantian objections to MBE. We argue that a careful engagement with Kantian moral psychology does provide a space for MBE but that it also describes a potential danger of MBE that has, at best, been only superficially described. That is, we offer a different Kantian understanding of MBE as a means to bring our empirical and noumenal selves together as a coherent whole to achieve what Kant describes as “genuine accountability to others.”

This chapter is published as: Kudlek, K., Smith, P.T. 2022. The Kantian promise and peril of moral bioenhancement. *Journal of Applied Philosophy*. DOI: <https://doi.org/10.1111/japp.12575>



# The Kantian Promise and Peril of Moral Bioenhancement

## 1. Introduction

Moral bioenhancement (MBE) has been discussed as a potential tool for fixing certain features of human moral psychology which seem to hinder our moral progress and impose risks on others in our current environment. Human moral psychology, the argument goes, seems to be maladapted to the moral requirements of a globalized world in terms of lack of moral concern for things that are personally and temporally distant to us (Persson and Savulescu, 2012). Those tendencies reflect a strong intuitive appeal to negative duties rather than positive ones – we feel more responsible for the harm we cause than benefits we fail to cause (Persson and Savulescu, 2012). These tendencies often interfere with our long-term interests and aims to be better people. As moral agents, we have legitimate reasons not to trust ourselves to make good choices. MBE is, hence, a proposal to remove some of those psychological drawbacks and help us expand our moral concern beyond the “near and dear” circle.<sup>140</sup> It is a novel way of “tricking” ourselves into doing the right thing. By diminishing certain emotions and tendencies, MBE can reduce and prevent harm and negligent behavior, and as such, should be in line with the demands of virtually all moral doctrines. However, the way in which we treat our moral agency when we want to prevent ourselves from engaging in negligent or harmful behavior (i.e., when engaging in MBE) can be a bit puzzling and this is something we want to examine more closely in this paper.

Consider the following examples. Odysseus asked to be tied to the masts of his ship so the Sirens’ song would not lead to his doom. Smokers hide their cigarettes to make it harder for them to fall off the wagon. A dieter makes sure there is no

---

<sup>140</sup> But MBE does not exclusively or necessarily amount to promoting the interests of others (positive rights) – it is also determined to prevent the damage to other’s interests (Savulescu and Persson, 2012). It is not only about increasing altruism, sympathy or empathy, it is also about decreasing levels of certain morally undesirable emotions or tendencies such as anger, violent aggression, racial aversion, narcissistic and psychopathic tendencies etc. (Douglas, 2008; Savulescu and Persson, 2012).

ice cream in the house. A person gives their keys to a friend at the beginning of a party so that they do not drive home drunk. Everyday life, history, and mythology are replete with examples of individuals responding to character flaws or a temporary loss of their faculties through a variety of mechanisms that make us more likely to comply with our own better judgment. As agents, we understand that we can be akratic – weak-willed – and our ability to make autonomous decisions can be weakened, undermined, or lost entirely. In response, we treat ourselves and our agency as a kind of object to be manipulated, directed, or coordinated. There is nothing obviously wrong about these everyday cases, but they can become problematic when evaluated from certain ethical perspectives – such as Kantian moral philosophy.

MBE looks like, and borrows some of its initial plausibility from, these kinds of everyday acts of self-paternalism. Just as the person who gives up their car keys because they know they might make bad decisions when drunk is being responsible, the person who engages in MBE knowing about the flaws in their character in order to comply with their own best moral judgment is also being responsible. And conversely, just as a person who refuses to take steps to ensure they do not drive home drunk is being negligent, a person who refrains from MBE might also be negligent. We believe, following Shiffrin, that negligence is often downplayed in philosophical literature and considered to be a slight wrong, while in fact, it can be a serious moral and political wrong, even when considered separately from its consequences (Shiffrin, 2017: 197). Although moral negligence does not necessarily involve bad outcomes, moral agents can be held culpable and morally responsible for their negligent actions.<sup>141</sup> Besides, negligence risks harm. This is not to claim that intentional and negligent harm are the same, but it is to say that they are both, at least *prima facie* morally wrong, while MBE can help prevent this risk.<sup>142</sup>

We are faced with a serious ethical dilemma: the failure to engage in MBE seems reckless and negligent, which can be considered a serious moral wrong,

---

<sup>141</sup> As Shiffrin explains, a negligent agent is an agent who “has transgressed against an actual, valid moral standard of due care in a negligent manner for which she is morally responsible” (2017: 202). In addition, we also agree with Shiffrin, that non-negligence can be considered a moral virtue.

<sup>142</sup> Some advocates of MBE even take a more rigorous stance – not only claiming that negligence and omissions are a serious moral wrong – but also that failing to benefit may be as morally wrong or bad as harming (Persson and Savulescu, 2012). Engaging in this discussion any deeper is not crucial for our goals here.

but engaging in MBE presupposes that we treat our faculties (our future agency) in a disrespectful self-paternalistic manner. Although paternalism might very well produce good outcomes, many have argued – inspired by Kant – that paternalistic actions are intrinsically disrespectful (Sandel, 2007, Hauskeller, 2013; Wilks, 2018; Vedder, 2019). In this paper, we want to resolve the above dilemma by suggesting a novel way of understanding Kantian objections to MBE. We understand MBE as a voluntary or consensual form of pre-commitment contract and question whether such contract can undermine our agency by treating it as an object to be manipulated. MBE is often meant to solve the problem of our excessive selfishness or insufficient altruism and it aims to fix our agency itself. However, this can be tremendously disrespectful on Kantian grounds, even if aimed at *our own* future agency. Nevertheless, we will show that a careful engagement with Kantian moral psychology does provide a space for MBE *but* that it also describes a potential danger of MBE that has, at best, been only superficially described. That is, we offer a different Kantian understanding of MBE as a means to bring our empirical and noumenal selves together as a coherent whole to achieve what Kant describes as “genuine accountability to others.”

The following section explains that arguably more acceptable forms of MBE, such as indirect MBE, represent a *generalized* skepticism about our agency and aim to fix our agency itself. Such attitude towards one’s agency applies even in the case of a voluntary, pre-commitment contract. In the third section, we introduce the analogy between MBE and self-paternalism and raise the concern about problematic ways in which self-paternalism relates to one’s agency. In the fourth section, we respond to this concern by looking at different stages of Kantian moral psychology and argue that there is a Kantian interpretation of MBE such that it is not generalized self-paternalism and therefore may be permissible. Finally, we show that while Kantians can endorse MBE under certain circumstances, there are important potential pitfalls that must be avoided. This interpretation sets certain constraints on how and why we might engage in MBE.

## 2. MBE as a Generalized Skepticism towards Our Moral Agency

Two approaches are typically discussed regarding MBE’s potential effects on our faculties: direct and indirect moral enhancement. Direct MBE is focused on correcting beliefs, motives and/or actions according to some external normative standard – e.g., inculcating the belief that murder is wrong (Schaefer, 2015: 262). It can also be said to target basic features of our moral psychology (our first-order

moral capacities), such as empathy and a sense of fairness (Earp et al. 2018). It simply amounts to feeling “more empathy” or “less aggression” (Earp et al. 2018; Douglas, 2013). By contrast, indirect MBE aims at making people more capable of reliably producing the morally correct ideas, motives and/or actions without specifying the content (Schaefer, 2015: 262). Unlike direct MBE, which relies on the enhancer’s moral beliefs, indirect MBE relies on the connections between certain processes and the correctness of moral beliefs (Schaefer, 2015: 262). It is focused on augmenting higher-order capacities which enable the flexibility of response to different situations and, therefore, can ensure a more reliable moral enhancement (Earp et al. 2018: 169-171). So, instead of simply feeling “more empathy”, one would have an improved ability to recognize when it is morally desirable to be empathetic and act upon it (Earp et al. 2018: 170). It is this latter (indirect) form of MBE that generally seems to be more acceptable.

Although the distinction between direct and indirect MBE comes close, it does not distinctly capture the context-specific and temporal dimensions of different types of enhancements. Therefore, we want to introduce a slightly modified distinction between what we will call *localized* and *generalized* MBE. Localized MBE is context-specific, applied to everyday practice, and has a temporary effect. It could, in fact, be said that localized MBE has the effect of a moral *therapy* rather than enhancement of some complex ability to understand or reflect upon our moral choices. Take, for example, an introvert who wants to leave a good impression on her social group, so she takes a pill to fit in without anxiety. The introvert need not view her introversion as generally problematic or burdensome but only wishes to manage a specific arena where it is sub-optimal.<sup>143</sup> By contrast, generalized MBE aims to fix a more general idea that we reason poorly about our moral obligations. Unlike localized MBE, which supplements or enables our broader agency, generalized MBE aims to fix our agency itself. So, our introvert would be generally enhanced if introversion was viewed as moral failing that needed to be repaired, either through chronic treatment or a single, permanent intervention. We believe that general MBE is both more controversial and attractive, much like indirect MBE. Yet, it is important to see that indirectness and generality need not go hand in hand: I could think that my direct capacities need general improvement or that my indirect capacities need localized improvement.

---

<sup>143</sup> This is analogous to Kant’s example of occasionally using wine to temporarily assist in one’s socializing if one is shy (MM VI 428). We thank an anonymous reviewer for the suggestion.



To illustrate more vividly how this differentiation maps onto the existing discussion, let us look at the famous God Machine example, developed by Savulescu and Persson (2012). Imagine a technology that monitors every human being's thoughts, desires, and intentions and can modify them in nanoseconds. This machine is designed to prevent great harm, injustice or grossly immoral behavior from happening. However, it only intervenes when intentions for such immoral behavior are formed – otherwise; human beings are entirely free in their moral (mis)conduct. Subjects are aware that God Machine exists but unaware of its interventions – in case they form an immoral intention, they simply feel “the change of mind” (Savulescu and Persson, 2012: 413). The God Machine resembles, in a sense, what we have called localized MBE because subjects are temporarily “treated” in specific situations without any reflection on their moral misconduct. It is important to note that although prevention of harm is among the main goals of MBE, Savulescu and Persson do not consider God Machine to be an instance of MBE because “it does not enhance your motivation to do what is morally right. Rather, it deprives you of your freedom – and even ability – to decide to do and do what is wrong” (2016: 265).

Even though God Machine, or moral enhancement in general, may produce good outcomes, many have argued that the ability to commit immoral acts – the so-called “freedom to fall” is tremendously (intrinsically) valuable and represents a constitutive part of personal and moral autonomy (Harris, 2011; Hauskeller, 2017; Sparrow, 2014a; Wilks, 2018). So, as far as MBE would deprive us of freedom to fall, it could be deemed impermissible – especially from a deontological moral framework.<sup>144</sup> In response to these criticisms, MBE advocates emphasize the importance of *voluntary* commitment or *consent* to engage in MBE. Namely, if enhancements were to be implemented voluntarily, the concerns about undermining autonomy, agency and freedom could be alleviated. If we take that “autonomy is the power to make well-grounded, rational decisions and to act in accordance with them”, “voluntarily connecting to the God Machine would then be an example of pre-commitment contract” and, therefore, it would not compromise autonomy (Savulescu and Persson, 2012: 414). In other words, our choice to

---

<sup>144</sup> Although the concern about depriving a moral agent of their freedom and autonomy most likely applies to all types of moral enhancements (direct/indirect, localized/generalized), its strength should be lesser in the case of indirect/generalized MBE. Generalized MBE should not radically change what we want but simply expand our moral concern/horizon and boost our motivation to do what we believe is right.

engage in MBE can itself be an expression of our autonomy/agency (DeGrazia, 2014).<sup>145</sup>

This issue has also been discussed from the Kantian perspective and, similarly, some scholars argue that MBE would not represent a threat to the Kantian conception of autonomy, as the capacity to deliberate and legislate a moral law to oneself.<sup>146</sup> For example, Hickey argues that not developing our capacities (employing human enhancement) would be disrespectful towards our rational will – because the will is authoritative and autonomous over capacities (2018: 178). He also argues that “[p]assively developing talents through a pill can seem to still recognize one’s status as a rational agent and end setter. It still represents an expression of our will. It clearly expresses the notion that one’s capacities are worth developing to be put towards one’s autonomously chosen projects” (Hickey, 2018: 183). On similar grounds, Wilks argues that as far as enhancement is achieved through natural causes, it cannot affect the autonomy of the will because “the will is free, i.e., is able to act autonomously” (2018: 139). She further concludes that autonomy is immune to (it is not affected by) both effects of direct and indirect MBE, insofar as those means are naturalistic. But this is something we will discuss further in Section 3.

We, however, believe that some aspects of Kantian ethics present a challenge even for voluntary (pre-commitment) MBE. In particular, using MBE as a pre-commitment strategy requires that we treat our agency as an object to be manipulated to generate positive actions – we use ourselves as mere means. MBE could present a kind of disrespectful self-paternalism and, as such, undermine our autonomy. In the following section, we develop this analogy between MBE and self-

---

<sup>145</sup> Not only has it been argued that MBE is not a threat to freedom and autonomy of a moral agent because it will not bypass moral deliberation and undermine moral freedom (e.g., DeGrazia, 2014; Danaher, 2018), it has also been argued that MBE will *improve* moral agency by removing biological barriers to careful and reflective moral reasoning and decision-making (e.g., reducing impulsiveness or alleviating certain biases can have beneficial effects for moral reasoning and moral conduct) (e.g., Savulescu, 2010; Douglas, 2013). For the discussion about enhancement’s potentially beneficial effects on autonomy/agency from Kantian perspective see, e.g., Carter (2017); Bauer (2018); Clewis (2017); Hickey (2017).

<sup>146</sup> As Wilks explains, according to Kant, autonomy is the capacity of persons to legislate to themselves the moral law of reason – to have their will guided in its choice and motivation by this law (Wilks, 2018: 124).

paternalism and identify some of the overlooked challenges that Kantian ethics seems to impose on MBE.<sup>147</sup>

### 3. Self-Paternalism and the Kantian Objection to MBE

Consider again poor Odysseus, who asked to be tied to a mast of his ship so he would not be lured to his death. This example has been used to justify a voluntary pre-commitment contract and to illustrate how a frustration of someone's desires can, under certain circumstances, represent respect for their autonomy (Savulescu and Persson 2012). But, as Rumbold argues, even if tying oneself to a mast is overall justified, such action comes at a moral cost which is incurred by enhancing oneself morally (2017: 550). We want to develop this concern further, especially with regards to the cost of one's own moral agency. Namely, if I tell you to tie me to a tree or a mast and you do it, usually I am permitted to change my mind. If you refuse to untie me after I have asked you to do so, you have wronged me. But under these sorts of Odysseus-type contracts, you are supposed to keep me tied up even after I have withdrawn my consent because I have, in some sense, lost my capacity to withdraw. It is only when I have regained my capacity that I can withdraw my consent and then be released. Following Andreou (2018), let us call this set of strategies for dealing with potential problems with our own agency, "self-paternalism."

To begin, it is important to see what is wrong with paternalism in general from the Kantian perspective and to show that such concerns apply, *mutatis mutandis*, to at least some cases of self-paternalism. Paternalism is the use of coercion or manipulation in order to force or push the targeted agent to act in their own best interests. Paternalistic actions can range from subtle interpersonal manipulation to nudges to legal prohibition with criminal sanction but they all share, from the Kantian perspective, a worrisome attitude of disrespect towards the agency of the target. Seana Shiffrin writes:

Even when paternalist behavior does not violate a distinct, independent autonomy right, it still manifests an attitude of disrespect toward highly salient qualities of the autonomous agent. The essential motive behind a

---

<sup>147</sup> For Kant, freedom and autonomy are related to the appropriate relationship and combination of the empirical and noumenal perspectives on our personhood. In this paper, we will be rejecting overly metaphysical accounts of the Kant that would more directly intervene on libertarian vs. compatibilist discussions.

paternalist act evinces a failure to respect either the capacity of the agent to judge, the capacity of the agent to act, or the propriety of the agent's exerting control over a sphere that is legitimately her domain. (Shiffrin, 2000: 220)

There are some areas of our life where we have especially strong interests in deciding for ourselves. However, even in those cases where such interests do not apply, we have a *general* claim to be treated as capable of running our own lives. And unlike other ethical views, one cannot violate or disrespect a person's humanity – their rational agency – in order to improve their well-being or to make them happy. Fahmy (2018) has identified three different Kantian grounds for the rejection of paternalism: it fails to respect our humanity, it undermines the development of our moral personality, and it violates our right to freedom insofar as it is coercive. On a deeper level, the problem with paternalism is that it inherently treats the target as an inferior; the paternalistic agent *knows better* than the target and should decide for the target because the target is incapable of deciding for themselves. For the Kantian, taking oneself as morally special, as superior, is not just wrongful, it is the fundamental source of all wrongdoing: the temptation of self-love to treat one's own interests or judgment as more important than others.<sup>148</sup>

Yet, it seems like self-paternalism avoids these problems. After all, in cases of self-paternalism, we are either constraining *ourselves* or we are consenting to be constrained by someone else. In other words, the constraint is an expression of our agency, of our judgment that we are going to be acting in ways contrary to our own understanding of our interests. We do not want to continue smoking but are weak-willed or addicted, we think driving drunk is terrible but do not trust our inebriated selves to act rightly. We know that following the Sirens will lead to our death but also know that we will be too brainwashed to resist. As Andreou (2018) points out, self-paternalism involves a kind of bifurcation of the self: we have a view of the right choice and a view about the choice that our flawed selves will make and create a world to compensate for our flaws or incapacities. This involves treating *part* of our selves – our addiction, our vulnerability to brainwashing, our diminished capacities from intoxicants or disease – as an object to be manipulated, an obstacle to get around. Yet, there is still another part of ourselves that we are respecting to direct our lives, the rational part, that is nonetheless coordinating our actions. When that part is in control, we are autonomous, and our decisions must be respected. The problem is that this part is not always in full control.

---

<sup>148</sup> See Kant's *Religion within the Limits of Reason Alone* (31-32).

MBE advocates rely upon an analogy between standard, everyday self-paternalistic interventions and the more general idea of making our capacities for moral deliberation and motivation better. If I can anticipate that I will be weak willed when faced with the ice cream in the refrigerator, why is it wrong to anticipate that I will be flawed in my basic reasoning capacities and take steps to mitigate the risk I will behave wrongly? The analogy fails, however, and that failure grounds a Kantian objection to MBE as self-paternalism. Self-paternalistic actions are typically understood to be temporary or context specific, supplementing our broader agency in order to get to act in accordance with our prior judgments. In other words, self-paternalism can be understood as expressing a respect for our agency by ensuring that the judgments that reflect our endorsed values are operative.

Yet, MBE is not like that as it represents a *generalized* skepticism about our moral agency. For example, MBE is often meant to “solve” the problem of our excessive selfishness and insufficient altruism (Persson and Savulescu, 2008; 2012). It is clearly true that I could take temporary or context specific steps – such as doing a monthly subscription for charitable giving that I make difficult to cancel – in order to increase my altruistic behavior. But MBE is designed to fix the more general idea that we are poor at reasoning about or acting on altruistic obligations. But it is important to see that this is quite different from the specific case: sometimes altruism is all things considered justified but other times being selfish is perfectly permissible. MBE is not designed – or not solely designed – to resolve the issue of “I have resolved to be altruistic but worry about following through” but rather “I am bad at reasoning about when I should be altruistic.” So, at some time in the future I will be presented with a choice between altruism and selfishness and I will *wrongly* judge that it is permissible to be selfish. And this will *not* be because of some temporary or other kind of problematic reduction of my agency – such as depression, intoxication, or distraction – but rather this is because *my agency itself* is flawed and needs to be fixed. This attitude would be tremendously disrespectful, at least on Kantian grounds, if aimed at another person. In fact, one might suggest that this attitude of superiority – that I know better and can use my agency to force you to follow my judgment rather than yours – replicates the fundamental Kantian grounds for all wrong-doing: the failure to treat rational agency with appropriate respect.

It is important to see that the distinction between *generalized* and *localized* self-paternalism cross-cuts the direct/indirect and also the revocable/irrevocable distinctions in MBE (as well as, we shall see, consensual/non consensual MBE). That

is, an MBE intervention can be reversible or not, direct or not, and still be an instance of generalized self-paternalism. Consider the following case:

James is a district court judge in a multi-ethnic area. He was brought up in a racist environment and is aware that emotional responses introduced during his childhood still have a biasing influence on his moral and legal thinking. For example, they make him more inclined to counsel jurors in a way that suggests a guilty verdict, or to recommend harsher sentencing, when the defendant is African-American. A drug that is administered daily is available that would help to mitigate this bias. (Douglas, 2013: 161)

As it stands, this is a direct and revocable intervention. Yet, what matters to the claim of generality is the *justificatory logic* of the intervention. That is, generalized self-paternalism is justified by the idea that one's agent cannot compensate or make correct judgments generally. James cannot trust his own judgment concerning how racist he is acting, so he takes the pill every day. It is not that his agency has been undermined by some other factor but rather that his agency is intrinsically or fundamentally in need of correction or improvement. So, even though James is taking a pill with temporary effects – like being tied to a mast temporarily – the justification of his action is not that his agency has been undermined but that his own agency is not up to the task of ensuring non-racist or anti-racist behavior. Adopting some language from Gary Watson (1975), localized changes are justified by a *mismatch* between the evaluative system (what the agent values) and the motivational system (what the agent wants or does) while generalized actions are justified by a claim that the evaluative system is itself flawed.

Of course, there seems to be an obvious difference between paternalism and self-paternalism. In the interpersonal case, I am using another person's agency – their humanity – as an object to be manipulated, an obstacle to be avoided. This looks like a disrespectful way of relating to them, an expression of superiority. In the generalized self-paternalism case, I am relating to myself. It is not immediately obvious if it is wrong to adopt an attitude of superiority towards myself. Yet, it is important to see that MBE is a way of relating to my *future* self, the person I am going to be. And the distinction between relating to *other* people and relating to my future self is not as sharp as one might think. Take Christine Korsgaard's discussion of self-paternalism and our future selves:

The 19th century Russian is now, in his youth, a socialist, and he plans to distribute large portions of his inheritance, later, when it comes to him, to the peasants. But he also anticipates that his attitudes will become more conservative...So he makes a contract now...to distribute the land when he

gets it, which can only be revoked at the consent of his wife, and he asks his wife to promise not to revoke it then, even if he tells her then that he has changed his mind.... Parfit says, "It might seem as if she has obligations to two different people." (Korsgaard, 2009, 185)

The point is that once we are choosing for future versions of ourselves, then the notion of generalized self-paternalism becomes more problematic. We might think this akin to constraining future generations. Of course, there is *some* sense in which we share a political community with our ancestors and our descendants, but they are sufficiently different from us such that making decisions on their behalf and in their absence requires special legitimation (Barry, 1977). The same applies in the case of the Russian nobleman. In a case of limited specific self-paternalism, we could argue that the person constrained is the same person agreeing to the constraint, especially since the idea is that the case where constraint is necessary represents a deviation from their rational agency. But in the MBE version of the Russian nobleman, we are constraining a version of ourselves in the relevantly distant future because we do not think they will make the right decision even if they are freely using their own rational capacities. In that case, future versions of ourselves have a legitimate complaint that they should have been entrusted to run their life. This need not involve any kind of radical metaphysics; the idea that we can resent the constraints imposed on us by our younger selves is commonplace. Furthermore, this kind of constraint can put third parties in the difficult position of deciding which claims to respect. This is not to say that we can never make decisions for our future selves or that we do not bear some special – and especially intimate – relationship with our future selves. But the way in which we make those decisions – and the reasons behind them – can be disrespectful towards our future selves and to our agency by expressing the idea that we would be better off if our choices were constrained because we cannot be trusted.

#### 4. MBE and Kantian Moral Psychology

We now have what looks to be a significant ethical dilemma and a serious potential problem for the application of Kantian moral philosophy to MBE. On the one hand, the failure to engage in MBE seems reckless, a deliberate abrogation of one's responsibility (or even duty<sup>149</sup>) to develop one's moral faculties and to avoid placing

---

<sup>149</sup> For an interesting discussion about how human enhancement could align with the Kantian duty to oneself, see Bauer (2018).

needless risks on others. On the other hand, the attempt to assimilate MBE as an instance of unproblematic self-paternalism fails and, as a result, MBE seems disrespectful of our future agency. In this section, we will argue that a careful engagement with Kantian moral psychology does provide a space for MBE *but* that it also describes a potential danger of MBE that has not been adequately developed in the literature. That is, we offer a different Kantian understanding of MBE as a contribution to appropriate relation or harmonization of our phenomenal and noumenal selves in order to achieve what Kant describes as “genuine accountability to others.”

At first glance, the Kantian resolution to this dilemma seems both easy and broadly sympathetic to MBE. On the standard view, Kant’s moral psychology is strongly dualistic. Our noumenal selves – the locus and grounding of our rational agency and our free will – is entirely separate from our phenomenal, empirical selves. Since our noumenal selves have radical freedom, no manipulation of the empirical self can affect the status of our noumenal self. Wilks writes

[I]t is difficult to see how the bioenhancement of an individual’s cognitive or other capacities may be *transferred* to the individual’s *will*, thereby enhancing its autonomy. On Kant’s view, this transference cannot happen, not even by the most indirect and remote means. The reason is that the will is *free*, i.e., is able to act autonomously, precisely in so far as it is *not* affected by the efficient causality operative in nature; it operates only in accordance with the law of practical reason, i.e., the moral law, which the will legislates to itself. The autonomy of the will cannot be accessed via natural means either to augment or diminish it. (Wilks, 2018: 139)

So, MBE is simply irrelevant to the autonomy of the will even if an admirable, empirical self may – in some way – make moral action easier. Thus, there is no dilemma. We can manipulate our phenomenal selves however we wish in order to generate more reliable compliance with the moral law, and it cannot – at least not fundamentally – affect our status as autonomous beings.

In what follows, we wish to focus on a set of issues that is best represented by the following question, “How does one *become* autonomous?” On this dualistic, metaphysical interpretation of the noumenal and phenomenal, the capacity for rational action simply emerges at some point in our empirical development. Yet, this question – what is the relationship between the development of our empirical psychology and our status as rational agents – is a key focus of Kant’s later moral psychology, especially in *Religion within the Limits of Reason Alone*. In other words, if one views our rational agency – our humanity, in Kant’s terminology – as a



separate capacity from our empirical selves, then it is unclear how education and moral development are supposed to work. However, if rational agency is something that we *achieve*, in part, through our empirical selves, then the relationship between our empirical and our noumenal selves is much closer.<sup>150</sup> That is, our humanity is a process.<sup>151</sup>

This understanding of Kant's moral psychology is interpretatively and philosophically plausible while having complex and productive consequences for the normative analysis of MBE. First, unlike the metaphysical conception, the process conception is consistent with MBE being highly problematic insofar as it undermines the process of becoming fully rational (or virtuous) in the Kantian sense. Second, the process conception creates space for MBE to be highly beneficial insofar as it encourages or helps us in the process of becoming virtuous. Thus, we need not understand MBE as self-paternalism. We suggest nuanced judgment concerning MBE: it is either permissible or impermissible based upon its role in the process of our moral development.

In the *Religion*, Kant is trying to explicate the sense – consistent with his underlying moral theory – in which human beings might be understood as “radically evil” or suffering from original sin. On Kant's view, this is really a question of how we move from being creatures with what seem to be quite ordinary psychologies to those who are capable of free, autonomous action. For Kant, the key element that divides good from evil is the “subjective principle of volition” or the principle by which one decides which sorts of reasons – at the most fundamental level – justify action. Evil beings are those whose subjective principle of volition is “self-love” and thus accept reasons are only binding, authoritative, or obligatory insofar as they serve our interests, or at least, our understanding of them. Good agents, by contrast, adopt the moral law as subjective principle of volition, subjecting the claims that their empirical selves make to the categorical imperative and thus become accountable to other people. It is important to realize that being “evil” in this sense is both unavoidable and, as we shall see, consistent with quite complicated and reliable behavior in *compliance* with the moral law even if it is not, yet,

---

<sup>150</sup> See David Sussman's *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* (2001) and Christine Korsgaard's “Morality of Freedom” in her *Creating the Kingdom of Ends* (1996).

<sup>151</sup> This is related but not equivalent to recent work to show that Kant's view allows more space for emotions in his later work (Carter, 2017). Carter does not focus on the process of development but rather focuses on the role emotions can play in the mature moral agent. Thus, she tends to focus on works other than *Religion within the Limits of Reason Alone*.

motivated out of respect for that law.<sup>152</sup> Thus, Kant's claim is that we are *necessarily*, due to the way in which we relate to our empirical psychology, default committed to self-love as a subjective principle of volition and we must learn and develop and come to understand the authority of the moral law. This requires a "revolution" in our subjective principle of volition that is made possible and enabled by a more gradual process of development – an "evolution" – of our empirical selves.<sup>153</sup>

This process is based on a gradual development by which our psychologies become more receptive and committed to the moral law. There are three primary stages:

1. The predisposition to animality: mechanical self-love in the pursuit of self-preservation, propagation of the species, and community with others
2. The predisposition to humanity: the inclination to acquire worth in the opinion of others
3. The predisposition to personality: the capacity for respect for the moral law as in itself a sufficient incentive of the will. (R, 21-23)

So, animality represents the claim that our empirical psychologies make up for basic needs – food, water, shelter and so forth – while humanity is something like the reasons generated by our social standing with others, including reputation and recognition. It is only in the final stage – personality – where we become genuinely accountable to other people. It is important to see that the first two dispositions are not *intrinsically* bad – Kant reviews to them as dispositions towards the good – but they can also be the source of vicious action. For Kant, we always begin with the first two dispositions – which are evil insofar as they rely upon the principle of self-love – and we must convert or transition ourselves to the third, final, and truly moral stage of our development.

In understanding the transition from animality and humanity to personality, we need to understand some key features of the process. First, the relationship between the subjective principle of volition and action is complicated; weakness

---

<sup>152</sup> Thus, Kant says these dispositions are all, in a sense, good as they can lead us to the moral law even if they can also incentivize vicious action (Religion, 23).

<sup>153</sup> See Marijana Vujošević "Kant's Account of Moral Strength" (2020) and "Kant's Account of Moral Weakness" (2019) for the idea that one can run into issues on both the "revolutionary" level of adopting the correct maxims and subjective principle of volition and the level of executing those maxims via one's psychology.

of will results when a person with a sincere commitment to moral law nonetheless fails to act because of a problematic set of animal and human dispositions.

However, that kind of weakness may not be the most important or insidious problem a developing agent faces. Rather, a significant threat to proper moral development is the way that certain desires – especially those associated with our humanity and thus our honor, reputation, and status – appear to our agency like genuine moral reasons but in fact reflect a kind of moral corruption or rationalization. We can imagine the person so obsessed with honor and reputation that they lose sight of acting in ways that are worthy of public honor in the first place. However, this is also the power of animality and humanity: if appropriately structured these psychologies can be strong inducements to comply with the moral law even if the agent is not appropriately motivated by respect for the moral law itself. In fact, it may be the case that it is impossible to tell whether one is truly motivated by the moral law in any particular case so indistinguishable can actions with different principles of volition appear externally and, perhaps, even to the agent herself. But all self-love based principles will ultimately founder in the sense that the compliance with the moral law is *accidental* and not necessary.

If the metaphysical, dualistic understanding of the relationship between the noumenal and the empirical selves was true, then we would expect Kant to say that we become good when our *rational capacity* as a distinct entity is committed to the moral law and is strong enough to overcome the deficiencies in our psychological character. And yet, this is not what Kant says. Rather, he suggests, rather mysteriously, that the commitment to the moral law (and to personality) is a “timeless” choice:

[M]an is under the necessity of, and is therefore capable of, a revolution in his cast of mind, but only of a gradual reform in his sensuous nature... That is, if a man reverses, by a single unchangeable decision, that highest ground of his maxims whereby he was an evil man (and thus puts on the new man), he is, so far as his principle and cast of mind are concerned, a subject susceptible of goodness, but only in continuous labor and growth is he a good man. That is, he can hope in the light of that purity of the principle which he has adopted as the supreme maxim of his will, and of its stability, to find himself upon the good (though strait) path of continual *progress* from bad to better...But in the judgment of men, who can appraise themselves and the strength of their maxims only by the ascendancy which they win over their sensuous nature in time, this change must be regarded as nothing but an ever-during struggle toward the better, hence as a gradual reformation of the propensity to evil...(R, 43)

This is a famously difficult passage to make sense of and we will not offer anything like a novel interpretation. What we will suggest is that what he says is consistent with the process interpretation and that the process interpretation is coherent and plausible. The idea here is that there is no discrete action where we can say definitely that a particular capacity “takes over” but rather than an extended series of actions – taken together – allow us to understand ourselves as committed to the moral law and to understand the fragility of that commitment. That is, we have developed to the point where we can understand our agency as committed to the “space of reasons” where we understand ourselves to be accountable to others. This commitment is not timeless in the sense that it occurs outside of time by a rational faculty with libertarian free will but rather in the sense the “true” or “most reasonable” interpretation of our commitments is always subject to revision, tentative, and can change with the next choice that we make. We can only understand our revolutionary commitment *in terms of* an unceasing struggle to reform our empirical nature and to act in accordance with the moral law. From our point of view, the unceasing struggle – combined with the hope that this reflects a revolutionary commitment to the moral law<sup>154</sup> *just is* the revolutionary commitment of our noumenal selves to the moral law; they are not distinct theoretical capacities but rather different ways of understanding ourselves practically. As Korsgaard says, “Kant’s theory of the freedom of the will requires neither extravagant ontological claims nor the unyielding theory of personal responsibility which seems to follow from those claims” (1996: 183).

## 5. The Kantian Promise and Peril of MBE

On this understanding of Kant, MBE will represent both opportunities and dangers because it can relate to this interpretation and the process of development in complex and varied ways. First, if MBE can be understood as a non-rational means of expressing a deeper truth, then it can help us “see” things differently and strengthen the move towards acting in ways that reflect movement towards the

---

<sup>154</sup> Kant argues that we can only hope that our continual efforts to improve our character generate the preconditions for the self-understanding as a moral agent: “Man cannot attain naturally to assurance concerning such a revolution...for the depths of his heart are inscrutable to him. Yet he must be able to *hope* through his *own* efforts to reach the road which leads thither and which is pointed out to by a fundamentally improved disposition, because he ought to become a good man and is to be adjudged *morally* good only by virtue of that which be imputed to him as performed by himself.” (Religion, 46)

ideal of understanding ourselves as committed to the moral law. To put it another way, nothing in Kant's view requires that each step in our development occur through rational argumentation. We can come to understand ourselves as free and autonomous – and thereby understand ourselves as committed to the moral law – in other ways. In other words, it is perfectly consistent with the process view that some steps in the process be guided, motivated, or made possible by non-rational means such as MBE *as long as* we can understand that non-rational contribution is moving our judgment towards deeper appreciation of the moral law and its demands.<sup>155</sup> For Kant, there is *no* step-by-step rational path from animality to humanity to personality; in some sense, the process needs to incorporate other paths of moral development. Cora Diamond argues, for example, that literary works can develop our ethical imagination in a critical fashion but not necessarily through argument:

I have spoken of literary and non-literary works which invite us to respond emotionally or to take up some moral attitude or view of life; what I need to add is that such works may include in the 'invitation' an invitation to just the kind of awareness and critical reflection I have described. We are familiar enough with the kind of critical attention invited by philosophical argument, the kind of work demanded by it of the reader; but critical attention to the character and quality of thought in a work may be asked of a reader in many other ways as well. Further, a work may invite a reader to *elaborate and develop a way of looking and to respond critically to it* then as a possibility, perhaps leaving open in various ways how it is to be elaborated, perhaps incorporating any number of suggestions. (Diamond, 1982: 37, emphasis added)

MBE, or other pharmaceutical interventions such as anti-depressants, can – under the right circumstances – be seen as an invitation to direct our ethical imagination in a new, better way. This “invitation” need not be argumentative for us to understand this new sensibility as superior, as moving towards personality in the Kantian sense. To understand how this might work, let's look at a pharmacological intervention that is commonplace and *comparatively* well-understood: drug treatments for depression. One way to understand treatment for depression is that the drugs *help* the depressed individual to see the world as valuable and other activities

---

<sup>155</sup> Kyla Ebels-Duggan (2019) suggests that we can come to have reasons through “direct appreciation” where we cannot fully articulate why we have reasons to act and value as we do. This is a non-argumentative process by which we come to understand something we did not understand before. She argues that this is fundamentally compatible with the Kantian picture.

as worth doing; these interventions can lead us to see the world differently as if a fog is lifted from our consciousness. They do not dictate that we find any particular thing or activity valuable, but offer us a way of seeing the world that is different and perhaps superior to how we saw it before. We can then endorse that change from our new perspective. Just as a novel or a bit of a political rhetoric can cause our attention to shift and to see the world differently, so too can pharmaceutical treatments both by removing impediments to taking up the new perspective and by inducing us to explore it. *Insofar* as MBE can be understood in this way – and we are agnostic as to when and how it can – then the fact that MBE relies on a non-rational causal intervention into our empirical selves is not an objection any more than it is an objection to developing our moral capacities via reading a literary work or experiencing natural beauty. Furthermore, MBE can help us combat moral corruption and rationalization by undermining the attractiveness of passions that *appear* to be expressions of the moral law but are actually self-love. This can include, for example, in-group bias, racial stereotyping, or dispositions towards anger and hatred. Again, this is contingent as sometimes these moral emotions – especially anger – can be important and essential to claiming our *rightful* status as well, but if we think there are dispositions that move us further away from personality and MBE can prevent that kind of backsliding, then MBE can be useful.

Yet, MBE also represents a clear risk to this process that is related to the paternalism objection and has, perhaps, received insufficient attention. Rather than generating progress, MBE could lead to developmental “cul-de sacs” where the apparent ease of improvements in our animality and humanity may undermine our ability to develop towards personality. This could happen in a variety of ways. For example, it may be the case that viewing one’s agency as something to be fixed is inconsistent with appropriately interpreting oneself as consistently *struggling* towards personality. Of course, we now see that MBE need not *necessarily* represent that attitude but it may be the case that some MBE interventions – or MBE interventions in a certain social milieu – should or would be interpreted that way, undermining our moral progress. For example, in any social environment that emphasized comparison and competition, then we might see MBE as a way of “getting up” on the competition or as a way to engage in moral “entrepreneurship” to improve our standing within the community. This might lead us to act in greater *compliance* with the moral law but could potentially move us even further away from achieving genuine personality even if our performance improves. In other words, a bad social milieu may push us to interpret our use of MBE (or

motivate its use) in ways that are ultimately self-defeating in Kantian terms. Additionally, we might trade increased compliance with the moral law in “normal” situations through MBE with lower performance in less likely situations. We might, through complacency, use MBE that generates superior performance on a narrower set of concerns, trading off moral robustness for reliability. But this modal robustness – will we act rightly across a wider set of possible worlds where it does not serve our animality and humanity – is a key test of moral progress. So, rearranging our empirical self to generate more reliable *compliance* with the moral law may actually move us farther away from being *committed* to the moral law.

As an example, let’s return to our judge who uses an MBE intervention to make themselves less racist. Whether this intervention is permissible, on Kantian grounds, will depend on whether James can reasonably understand the intervention as contributing to the combined evolution/revolution of his character, moving from animality and humanity towards personality. That is, if the intervention allows James to “see” more clearly – to rationally grasp – the moral demands of racial equality by, for example, dispelling distractions or quieting the demands of self-love (understood in terms of a psychological disposition towards favoring his race however subconsciously), then the intervention may very well be justified. This could happen in a few ways. First, the MBE could be “active” or “positive” in directing our attention towards an aspect of the cases that tended to be occluded from James’s view: perhaps the MBE leads him to concentrate on the common humanity of the people before his court. Second, the intervention could be “negative” in the sense of de-emphasizing something that tended to take an out-sized role in our motivations, perception, or reasoning, such as the unintentional but racist disgust James may feel upon seeing certain people.<sup>156</sup> Our psychological

---

<sup>156</sup> For example, in the *Doctrine of the Method of Pure Practical Reason*, Kant argues we will need non-rational means to get individuals with the wrong empirical make-up to achieve understanding of the moral law such that they can be moved by the dictates of morality alone (CPR, 152): “To be sure, it cannot be denied that in order to bring about a still unmolded or a brutified mind into track of the morally good in the place, some preparatory guidance is needed to entice with its own advantage or scare it with harm.” The interpretation offered here is different from Kant’s analysis in the *Critique of Practical Reason* in two ways. First, it is in some ways more pessimistic: achieving true morality is not a matter of simply setting up our agency the right way and then letting reason take over, achieving personality is a never-ending project of progress and improvement towards an end point that is, in a sense, unachievable. Second, our understanding how we can train our empirical selves is more capacious and optimistic: we can intervene in ways that improve our understanding and motivation in ways that do not rely on self-interest or fear.

dispositions can undermine the process towards accountability. Removing or decreasing these obstacles or getting us to see something we previously did not due to limitations in our empirical selves through non-moral means may be both effective and respectful. The idea is that we can harmonize our empirical and noumenal selves both in making it *easier* for us to follow the dictates of morality but also as interpreting our struggle to train our empirical selves as a movement *towards* an ideal, as getting closer to a genuine accountability for others.

However, it is also the case that an MBE intervention can harden a psychological equilibrium around self-love by blocking the kind of critical thought that the constant striving for virtue requires. If the intervention allows James to “forget” that he is racist or undermines a disposition to deal with the underlying causes of the problematic dispositions, then the intervention would be less justified on Kantian grounds. Similarly, if the MBE intervention is motivated by a desire to retain position, honor, or reputation in society – or makes it possible for James to retain his social status without thereby committing to the moral law – then MBE interventions can be positively self-defeating. Does the MBE lead James to “elaborate and develop a way of looking and responding critically” in an anti-racist fashion? On our view, there is no reason to think that MBE *cannot* do this but there is also no guarantee that it *will*.

Whether these interventions can be reasonably interpreted as “deepening” our understanding and commitment to the moral law or whether they are self-defeating cul-de-sacs will be a complicated judgment. For James, it will depend on how the intervention affects his deliberations, the phenomenology of being under the influence of the treatment, how he responds when no longer on the drug, and context-specific features of his job, family life, and society. It is important to see<sup>157</sup> that James’s maxim – the principle upon which he understands his reasons for taking MBE – is an important part of this story but it is not dispositive. Whether James can reasonably interpret his taking of MBE as part of progress or as a cul-de-sac will depend upon a variety of factors, *including* his self-understanding of his own motivations.

But other factors will also determine how reasonable this self-interpretation will ultimately be. This is unsurprising since our empirical selves are influenced

---

However, it should be noted that, even here, Kant uses a *story* to impress upon the young listener the purity of truly moral motivation (CPR, 157). So, in a way, Kant already acknowledges the importance of these other methods for training our empirical selves. We thank an anonymous reviewer for bringing this to our attention.

<sup>157</sup> We thank an anonymous reviewer for pushing us on this point.



by things outside of our control: if non-rational factors can improve or undermine our pursuit of personality, then certainly things besides MBE, such as a hyper-competitive milieu, will influence us as well. What's more, the availability of this sort of reasonable interpretation will be orthogonal to other worries about enhancement, such as whether MBE makes virtue "too easy" – in terms of bypassing personal effort to be morally good which we often find valuable in itself and a relevant factor for ascribing moral praiseworthiness. On the Kantian view, there is a real sense that – from the perspective of any particular moment in time – virtue is impossible, so issues of "ease" or "effort" are not what's morally relevant. An MBE intervention need not be effortful for it to be justified on Kantian grounds. But this is unsurprisingly accidental, as aesthetic, or mystical conversion experiences may be positive in terms of moral development and yet not require much effort. The potential problem with James taking the pill to no longer be racist is *not* that the pill requires no effort but rather that the lack of effort *may* lead James to think that he has solved his racism problem when he has not.

Whether MBE can or should be understood as non-rationally helping us to develop our ethical imagination and commit to a set of Kantian values as opposed to blocking or undermining our moral development will be context specific and contestable. However, the same thing can be said about when good political rhetoric transmutes into propaganda or when a psychological treatment becomes an ideological imposition. There may be no bright line, but that is no reason to deny *either* the emancipatory or dangerous potentialities of the MBE interventions.

## 6. Conclusion

In this article, we offered a novel way of understanding Kantian objections to MBE, suggesting that generalized self-paternalism is a unique and problematic way to relate to one's agency. Although engaging in MBE essentially represents a voluntary pre-commitment contract aimed at avoiding bad outcomes, which is a morally responsible thing to do, it also presupposes that we treat our future agency as an object to be manipulated. We explained how such paternalistic approach to one's agency is objectionable on Kantian grounds. Nevertheless, we showed that a careful engagement with Kantian moral psychology opens up a space for MBE to assist in moral development. It is possible to understand MBE as an invitation to direct our ethical imagination in a new, better way. It can help us combat moral corruption and rationalization by undermining the attractiveness of self-love. However, we noted that MBE could also represent a risk insofar it eases the

improvement in our animality and humanity in ways that could undermine our ability to develop towards personality.



The Ethical Analysis of Moral Bioenhancement



A. Karametas, *The weight of life* (2014)

## Conclusion



# Conclusion

This dissertation examined how we can make progress in the MBE debate by addressing some of the main challenges through the lens of ethical theory. It raised pertinent questions about the debate's blind spots, the conceptual and theoretical sensibility and coherence within it, as well as its normative value. My analysis showed that many concerns could be mitigated, misconceptions clarified, and popular claims challenged through a systematic application of ethical theories. A general approach in this dissertation was based on the idea that a thorough conceptual, theoretical, and normative analyses can provide insights into the coherence, feasibility, and ultimately the permissibility of enhancement projects and, in that sense, contribute to the main research agenda. I proved that we can make progress in our analysis and develop conclusions that bring us closer to resolving some reoccurring questions in the debate with a higher degree of confidence.

In this conclusion, I will first summarize the contributions that previous chapters (individually) make to the debate and explain how they (jointly) serve the purposes of this thesis. This will be accomplished by circling back to the main research questions (set out in the introduction) and elaborating on what my analysis allows us to conclude. Aside from contributions, I will also mention some strengths and limitations of my approach together with suggestions for future research. Finally, adding to further implications of my work, I will give some special attention to the socio-political aspects of MBE.

## 1. Contributions to the Debate

The research in this dissertation has been guided by the overall question: ***How can we use ethical analysis (relying on normative ethics and metaethics) to improve the bio-ethical debate on MBE and our understanding of what is or is not morally permissible?***<sup>158</sup> Following this line of research led to several types of contributions. The dissertation provided an extensive review of the ethical debate, which enabled identifying the focal points of concern. This, in turn, gave the means to develop a methodological approach (or, at least, a proposal of methodological guidelines) for bringing

---

<sup>158</sup> Research questions are marked in bold.

us closer to resolving the issues at hand. Finally, the theoretical and normative analyses enabled the implementation of this methodological approach to demonstrate its applicability on concrete examples.

In what follows, I recapitulate the main results by answering the research questions laid down in the introduction. I tentatively divide the results in three broad categories that tie particular research questions to their corresponding chapters. For a summary, see the table at the end of this section.

### 1.1. Critical Review and Methodological Guidelines

In terms of contextualizing the story around MBE and setting the scene for my methodological analysis, Chapter 1 and Chapter 2 reviewed respectively *what the main challenges in the human and moral bioenhancement debate are and how we can move towards a systematic evaluation of enhancement projects*. These parts of the thesis reviewed some of the most pertinent issues in the human and moral enhancement debates and captured how the discourse has been carried out between opponents and proponents. This allowed me to identify the debate's most concerning aspects and develop a proposal for a systematic assessment of enhancements.

In Chapter 1, building on the identified concerns, I proposed three general categories most relevant for discussing the moral permissibility of HET: (i) enhancement's coherence and feasibility; (ii) its effects on fundamental moral values; and (iii) socio-political ramifications. A critical review of the enhancement discourse in Chapter 1 also clarified why we need a systematic theoretical and normative approach. I explained that the essential concern is whether enhancements are *morally* permissible, even if they were to become legally permissible, biotechnologically feasible, and safe. In addressing this concern, many authors provide an implicit normative evaluation that reflects some of the predominant normative standards (such as utilitarian, Kantian, Rawlsian, etc.), but this is rarely done explicitly and systematically. To minimize vagueness and arbitrariness, I proposed we should ascend to a more systematic, normative approach, given that moral permissibility is a subject of normative ethics, which can provide guidance for a better understanding of what is or is not morally permissible.

Chapter 2 narrowed the research subject down from human enhancement writ large to moral enhancement, in order to demonstrate the points made in Chapter 1 on a more concrete example. It captured the ethical debate on MBE, presenting the most prominent proposals for improving human morality by biomedical



means and reviewing the most common concerns. I pointed out that the same three types of concerns from the human enhancement debate are also typically raised in the moral enhancement discourse. Hence, I applied the same guidelines for discussing MBE's moral permissibility, as previously developed in Chapter 1, to help us move towards a more rigorous and systematic ethical analysis of enhancement projects.

In sum, the first research sub-question is answered by identifying the main points of concern and interest (listed above) – i.e., by identifying where there is polarization and lack of dialogue – and by proposing the utilization of instruments of ethical theory in pursuit of a more systematic assessment of moral permissibility of MBE.

## 1.2. Theoretical and Conceptual Analysis

Apart from reviewing the state of the debate and developing a methodological framework (tasks conducted in Chapters 1 and 2), other dissertation chapters implemented rigorous theoretical (conceptual) and normative analyses. Namely, Chapters 3 and 4 focused on *what a close theoretical examination of enhancement's intrinsic and extrinsic properties implies about its coherence and feasibility and how this is relevant for its moral permissibility*. This included two separate analyses. First, a broader metaethical analysis of HET's intrinsic properties determined whether such technologies are intrinsically bad and how this relates to their moral permissibility. Second, a more focused analysis of MBE investigated whether it is feasible in principle and would generate the anticipated effects.

More specifically, in Chapter 3, I challenged the (bioconservative) claim that HET, in broad terms, are intrinsically bad and should be considered morally impermissible. I investigated whether such an assertion is true when subjected to definitions and well-established views about intrinsic value. By bringing the theory of intrinsic value into this discussion, I showed that HET's intrinsic and relevant nonintrinsic properties (such as unnaturalness, the agent's desire for mastery, and the necessary production of bad consequences) do not justify assigning negative intrinsic value to HET. In short, I argued that (i) we cannot consistently claim across cases that unnaturalness is necessarily a bad property; (ii) that the desire for mastery does not have intrinsic or relational properties that can generate the negative intrinsic value of HET; and (iii) that necessary consequences that affect intrinsically valuable things cannot ground the (negative) intrinsic value of HET because 'necessary' and 'intrinsic' are distinct concepts, and we lack

theoretical as well as empirical support that HET will in fact undermine any valuable aspect of human life. Additionally, I proposed that even if HET had negative intrinsic value, this would not necessarily entail that they are morally impermissible, since moral permissibility need not entirely depend upon intrinsic value. Therefore, it is unwarranted to ground HET's moral impermissibility on the assumption about their intrinsic (dis)value. This chapter eliminated one of the essential concerns in the human enhancement debate and cleared the path for the following chapters to focus on other, predominantly contingent issues.

Chapter 4 was concerned with whether emotions could in fact serve as vehicles of moral enhancement, as proposed by some of its advocates (Persson and Savulescu, 2008; 2012). By relying on neuroscientific and evolutionary-psychological views, I showed that emotions do not fulfill necessary requirements and that their modulation could misfire. Namely, emotions are evolved and functionally specialized programs, the task of which is to coordinate other mental programs, with the ultimate task of promoting one's fitness. Basing our moral decision-making on these grounds can have potentially undesirable effects, such as sympathizing more with our close group instead of expanding our moral concern to others. Therefore, MBE by emotion modulation must be more nuanced and consider the role and function of entire cognitive modules associated with moral decision-making.

Discussions carried out in Chapters 3 and 4 tie to the second sub-question by reaching conclusions about (internal) coherence, feasibility, and effectiveness of particular enhancements. They mainly do so by looking into external but integrally linked theories. Employing theories of intrinsic value showed that there are no decisive reasons to think HET are intrinsically bad and morally impermissible on those grounds. This indicates that enhancement projects are legitimate subjects of ethical discussion, because, if they were intrinsically bad and consequently morally impermissible, further ethical discussion would be very limited. Chapter 4 also showed that, even if enhancements were not bad in themselves, some types could have undesirable or even harmful effects. My example was MBE by emotion modulation, which has a limited scope and may cause counter-productive effects. Although such outcomes reflect negatively on an intervention's coherence, feasibility, and effectiveness, this criticism does not apply to every conceivable type of MBE. By developing a more sophisticated type of MBE, e.g., one modifying relevant cognitive capacities on top of emotional ones, the above mentioned problems could be mitigated. A sophisticated moral enhancement is expected to make the agent more autonomous and improve cognitive and volitional capacities,

normative competence, and reasons-responsiveness. An intervention that reliably achieves intended effects can be considered coherent and, therefore, theoretically justified.

### 1.3. Normative Ethical Evaluation

Chapters 5 and 6 demonstrated how applying normative ethical theories can help us further illuminate moral enhancement's moral permissibility. They (respectively) challenged some common assumptions about MBE's relationship with consequentialist and deontological doctrines to address *how MBE aligns with prevailing moral norms and whether it conflicts with fundamental moral values*.

Chapter 5 analyzed whether MBE is justified on utilitarian grounds. There is indeed a prima facie case that MBE appeals to utilitarians due to its capacity to prevent harm and promote altruism and benevolence. I argued that MBE roughly accords with utilitarian demands and principles, especially in terms of expanding moral concern and promoting agent-neutrality. However, I also emphasized that there are potential pitfalls to consider, regarding the tension between commonsensical and classic utilitarian reasoning. I concluded that, although these concerns place tight constraints on justifying MBE on direct utilitarian grounds, they apply with lesser strength in the case of indirect utilitarianism and (as of yet hypothetical) sophisticated types of MBE, such as the fine-tuned enhancement of motivational and deliberative capacities.

Similarly, Chapter 6 focused on the implications of MBE in a deontological framework and looked into leeway and obstacles for justifying MBE on Kantian grounds. This part of my research offered a counter-balance to the discussion on utilitarianism, for it is often thought that consequentialism and deontology stand in opposition and channel conflicting or incompatible moral intuitions. Showing that there is at least some room to justify MBE from both perspectives makes a valuable contribution to the ethical debate.

Chapter 6 examined how MBE relates to one's moral agency and whether this is objectionable from a Kantian perspective. The chapter identified a novel objection to MBE based on problematic self-paternalistic tendencies that may constitute voluntary pre-commitment enhancement. It was argued that, although engaging in MBE essentially constitutes a praiseworthy, responsible action aimed at avoiding bad outcomes, it also presupposes that we treat our future agency as an object to be manipulated. On Kantian grounds, such a paternalistic approach to one's agency may be objectionable. However, a careful engagement with Kantian moral

psychology seems to leave enough space for MBE to assist in moral development and help us combat moral corruption and rationalization by undermining the attractiveness of self-love. On the other hand, MBE could also represent risk insofar it eases the improvement in our animality and humanity in ways that could undermine our ability to develop towards personality.

In response to the third research sub-question, MBE roughly aligns with basic utilitarian demands (especially in terms of indirectly reaching desirable outcomes), but constraints could be put in place with regard to its effects on an agent’s decision-making. Similarly, Kantians could object to the way we would treat our moral agency when engaging in MBE, but, on net, Kantian ethics should be sympathetic to MBE’s capacity to facilitate the moral development of an agent. That said, MBE does not explicitly undermine/contradict fundamental moral values (as described by utilitarian and Kantian ethical doctrines).

The table below summarizes contributions and how they tie to chapters and particular research questions:

| RESEARCH QUESTION   | CHAPTER   | RESULTS   |
|---|-----------|---|
| <i>What are the main challenges in the MBE debate and how can we move towards a systematic evaluation of enhancement projects?</i>  | Chapter 1 | - Systematic overview of the main challenges/concerns   |
|   | Chapter 2 | - Methodological guidelines for employing ethical theory  |
| <i>What does a close theoretical examination of enhancement’s intrinsic and extrinsic properties imply about its coherence and feasibility (and how is this relevant for its moral permissibility)?</i> | Chapter 3 | - HET’s intrinsic value is unlikely to be negative<br><br>- moral permissibility need not depend upon intrinsic value   |
|   | Chapter 4 | - MBE cannot be reliably implemented through emotion modulation only<br><br>- MBE needs to be more sophisticated        |
| <i>How does MBE align with predominant moral norms and does it conflict with basic moral values?</i>  | Chapter 5 | - MBE is roughly compatible with (indirect) utilitarian demands<br><br>- MBE may undermine (direct) utilitarian demands |
|   | Chapter 6 | - MBE can undermine one’s own future agency<br><br>- MBE can assist in moral development                                |

To summarize, the dissertation chapters make their specific (stand-alone) contributions to the debate, but they also complement each other in demonstrating the thesis' methodological contributions. On the whole, they serve to answer the main research question about *how we can use ethical analysis to improve the bioethical debate on MBE and our understanding of what is or is not morally permissible*. I showed, by answering the three research sub-questions, that we can use ethical analysis:

- i) to identify overviews and polarization in the debate and to develop a methodological framework for a systematic assessment of enhancement projects;
- ii) to challenge common assumptions about enhancement's intrinsic and extrinsic properties (e.g., although HET are not intrinsically bad, some can have undesirable effects);
- iii) to show that MBE does not explicitly undermine or contradict fundamental moral values (as described by utilitarian and Kantian ethical doctrines).

The conclusions from chapters serve as proof of concept that ethical theory can help us evaluate enhancement projects and resolve, or at least bring us closer to resolving, some pertinent issues in the enhancement debate.

Insofar as MBE's moral permissibility is concerned, my position is that MBE is, in principle, morally permissible. Surely, MBE's permissibility will depend on various contingent factors, but (I showed that) we do not have decisive reasons to treat MBE as utterly impermissible under realistic scenarios. Importantly, although moral permissibility is a positive moral status, this status is tentative and relatively weak – as some say, “morally permitted” is a morally indifferent form of behavior that we are ready to tolerate, even if we disagree with it (Babić, 2000). That MBE is morally permissible means, roughly, that it is not morally wrong under all (typical) circumstances and that we should consider it as a plausible alternative in ethical deliberations. Also, MBE is not morally obligatory since morally permissible actions are typically not required, but it could be supererogatory or praiseworthy (as suggested in Chapter 6). Any further discussion about whether MBE is morally right or obligatory will depend on various contextual considerations, such as the type of (enhancement) technology, the recipient, social and political circumstances, available practical alternatives, implementation safety, costs, risks, etc.

Conclusions reached in Chapters 3, 5, and 6 support this ultimate verdict. On the whole, I challenged the bioconservative thesis that enhancements should be always morally impermissible irrespective of their effects, safety, the fairness of distribution, etc. According to my arguments presented in Chapter 3, MBE is not intrinsically bad, and even if it was, its permissibility need not depend on its intrinsic value. Things can be intrinsically bad and still morally permissible. Additionally, Chapters 5 and 6 showed that MBE is roughly within the bounds of predominant ethical systems—utilitarianism and Kantianism—and, therefore, morally permissible.<sup>159</sup>

## 2. Strengths, Limitations, and Further Implications

Naturally, my PhD project took some turns that were not originally planned. This comes as no surprise in philosophical work, especially when dealing with a hotly debated, developing topic. Some scaling down was necessary for my project to be tenable and fruitful. I had to keep my scope narrow and my goals well-adjusted to the ongoing discussion. In this section, I will explain why I made certain choices, as opposed to some others, and indicate how the foregoing discussion may be lacking, but most importantly, how it can be further improved.

### 2.1. Methodological Aspects

Instead of conducting a full-scale ethical analysis of MBE's moral permissibility, as initially planned, this thesis focused on providing a template or a set of guidelines that could bring us closer to understanding what may or may not be permissible. Although such an approach may lack comprehensiveness, providing a reasonably flexible framework seemed more suitable for a developing topic such as MBE. This approach also allowed me to cover a selection of narrower topics

---

<sup>159</sup> Although we can use consequentialism and deontology as assessment tools, this does not imply that we have to accept their understanding of moral permissibility. For example, consequentialists find an option morally permissible if there is no alternative option with morally better consequences (Harman, 2015). The best possible option is then considered morally right and obligatory (as I explained in Chapter 5). I do not mean to imply here that my understanding of MBE's moral permissibility necessarily leads to endorsing such principles. We can accept the consequentialist evaluation of MBE without accepting their understanding of moral permissibility.

(instead of following a straight line), and publish work that is highly relevant and reinvigorating to the ongoing philosophical debate on biomedical enhancement.

Chapter 1 and Chapter 2 determined the scope by providing definitions and reviewing accounts that serve as points of departure of this thesis. For the most part, this meant understanding MBE as an attempt to boost moral motivation through emotion modulation, within which mildly invasive means, such as pharmaceuticals, are implemented (Persson & Savulescu 2008; 2012). However, other approaches to MBE would also make relevant research subjects. One such approach is Douglas' proposal to attenuate counter-moral emotions (described in Chapter 2). Understanding MBE as a means of *dialing down* certain morally problematic emotions such as violent aggression and racial aversion (Douglas 2008) sheds a different light on this technology. Instead of thinking of MBE as a means of extending our moral obligations to others (often at our own expense), we can think of MBE as a means of preventing harmful and negligent behavior. This aspect has been brushed upon in Chapter 6, but further developing this disparity would be a fruitful continuation of my discussion about justifying MBE in a consequentialist and deontological framework.<sup>160</sup>

Chapters 1 and 2 also reviewed a set of selected *pro-et-contra* arguments about human and moral enhancement to illustrate where there is tension and polarization in the debate. However, there are other moderate views, which were not thoroughly discussed.<sup>161</sup> Discussing such views in more detail may have painted a somewhat different picture of the state of the debate. My intention was to depict the discourse between strict proponents and opponents because their views are the most influential and dominant in the discussion. In my critical review, I found pro-enhancement arguments more convincing than their bioconservative counterparts. And although this is importantly different from promoting pro-enhancement views, some may find my presentation somewhat biased or lacking a more balanced and moderate criticism of HET. To mitigate this concern, I provide several references to other approaches to human enhancement, especially throughout Chapter 1, and I take a critical stance towards pro-MBE views, most explicitly in Chapter 4, as well as in Chapters 5-6.

Identifying the most pertinent issues in the debate led towards mapping out three broad points of discussion, which I found most relevant for evaluating the

---

<sup>160</sup> This approach could bring notions like positive and negative duties/rights into the discussion.

<sup>161</sup> Moderate views on enhancement often entail that evaluation should be made on a case-by-case basis. For an overview of moderate enhancement views see, e.g., Mukerji et al. (2014).

moral permissibility of HET. I suggested (in Chapter 1) that we should look more closely into: (i) coherence, feasibility, and effectiveness of enhancement projects; (ii) their relationship with fundamental moral values and norms; and (iii) their compatibility with or facilitation of socio-political goals of equality and justice. These categories have generally guided my research as I examined some issues that fall under the first (Chapter 3 and 4) and second category (Chapter 5 and 6), while the third category has been discussed only briefly (in Chapters 1 and 2). I will now say something about how the analysis in these categories can be further expanded with a special emphasis on socio-political aspects of MBE, since discussing these matters was beyond the scope of the thesis.

## 2.2. Theoretical and Conceptual Aspects

Although Chapter 3 resolved some misconceptions about the intrinsic value of HET, by allowing us to eliminate concerns about their intrinsic badness, it opened up some avenues for future research. Most importantly, it prompted questions about the *extrinsic* value of HET and how enhancements should be evaluated in terms of their *contingent* properties. I argued that even if we cannot plausibly claim that HET are morally impermissible for intrinsicality-based reasons, they can still be impermissible for extrinsic reasons, e.g., due to producing potentially bad consequences. Some of my chapters discuss the potential effects of MBE (primarily Chapter 4 and 5), but it would be interesting to conduct a more systematic analysis of this topic, as I suggested at the end of Chapter 3. Any substantial discussion about MBE's effects will have to be informed by empirical data which is currently unavailable (since the technology is still in its speculative phase). Many of the questions raised within the MBE debate will only be resolved once more empirical results are available.

Moreover, the discussion carried out in Chapter 4 (building on MBE accounts described in Chapter 2) about the suitability of emotion modulation as the (sole) “perpetrator” of MBE could benefit from introducing other (neuro)scientific views on emotions besides evolutionary psychology. One example is a recently published paper that builds upon the discussion about the role of emotions in the MBE debate by investigating the limits of emotion modulation through theoretical analysis of affective landscapes (Carman, 2020). In addition, others have attempted to link the discussion about emotional/motivational MBE to topics such as epidemiology, vaccination, and climate change (e.g., Gibson, 2020; Crutchfield, 2021; Rueda 2020).



### 2.3. Normative Aspects

The normative analysis in this thesis settled on two opposing ethical theories (consequentialism and deontology) and the examination of MBE's relationship with some of their most fundamental principles, such as utility maximization, agent neutrality, respect for one's agency, etc. My research could further benefit from connecting MBE to other ethical principles and doctrines, for instance, the third most prominent normative ethical theory – virtue ethics.<sup>162</sup> Since virtue ethics places great normative weight on improving desirable character traits and moral education as the primary method for inculcating these values, MBE seems suitable for facilitating such goals. It would be interesting to explore whether MBE could be a justified means for the inculcation of moral virtues on aretaic grounds, or it would undermine values such as *freedom of choice* and the *value of personal effort*. I would also be interested in exploring whether the potential compatibility of MBE and virtue theory reinforces my concerns about the incompatibility of common-sense morality and the utilitarian perspective. Examining such and similar aspects would undoubtedly complement and strengthen my approach.

Another perspective that could benefit my normative analysis is the role of MBE in harm prevention, which I touched upon in Chapter 6. For example, we might want to investigate further whether there is an *obligation* to enhance if MBE can prevent harmful and negligent behavior. Could deontological constraints on doing harm override other reasons deontologists might have for rejecting MBE interventions, such as preserving the freedom and autonomy of a moral agent? We might also wonder whether engaging in MBE can be seen as praiseworthy and virtuous, whereas refraining from it as blameworthy and negligent. These aspects can be investigated by applying contemporary theories of harm, as well as philosophical discussions about moral responsibility and agency.

Speaking of MBE in the context of harm brings us closer to discussing its practical implications, and in this regard, I want to say something more about MBE's socio-political effects.

---

<sup>162</sup> Luckily, this topic has been gaining attention from scholars, mostly at the human enhancement level (see e.g., Fröding, 2013; Jotterand, 2011), but also more recently at the moral enhancement level (Fabiano, 2021).

### 3. Socio-Political Ramifications

The third category of conditions relevant for the moral permissibility of HET relates to their compatibility with and facilitation of socio-political goals like equality and justice. It has been briefly discussed in Chapter 1 and Chapter 2 that, even if enhancements are coherent and sensible projects that do not contradict prevailing moral norms (as this dissertation attempted to show), their practical implications represent the next challenge. Addressing this challenge will largely depend upon data that can only be provided by empirical research in due time, but the methodological approach I developed in this thesis could be applied to investigate some theoretical aspects of enhancement's effects in the socio-political domain.

More specifically, I mentioned in Chapter 2 that for MBE, this would mean investigating whether it would exacerbate existing social equality and justice problems or instead serve as a tool for *reducing* the negative effects of the natural and social lottery. Related questions include: Would an enhancement be in the individual's and/or society's interest? Would it provide grounds for exploitation and discrimination, or would it reduce social inequality and injustice? This topic could further involve examining whether there are practical reasons for the development and implementation of MBE and how compelling they are. Namely, even if MBE turns out to be morally problematic, compelling pragmatic reasons for its implementation may still prevail under certain circumstances. This would mean examining whether, for example, limiting freedom for the sake of safety can be justified in particular circumstances, especially in a crisis.<sup>163</sup> These questions build upon my discussion from Chapter 5 about whether MBE is our best and quickest fix for pressing issues or there is a better alternative.

Some of these questions could be tackled with the help of well-established theories of social justice and equality. For example, to study how MBE would affect social justice, equality, or societal values in general,<sup>164</sup> we could apply models such as the Rawlsian maximin principle. The maximin principle is famous for its requirement that any redistribution of social goods can only be justified if it

---

<sup>163</sup> A good example here is the recent COVID-19 pandemic where most governments ruled that limiting certain freedoms was justified for the sake of general safety. Similarly, there could be a future crisis scenario where implementing MBE could be justified even if this meant undermining certain values.

<sup>164</sup> As some preliminary research has shown, the societal values that are and will be most threatened by HET include: autonomy, dignity, equality, fairness, health and safety, peace, privacy, respect for human life, and solidarity (Jensen et al. 2018: 6).

improves the wellbeing of those at the bottom of society (Rawls, 1971).<sup>165</sup> This approach harkens back to my discussion about MBE and utility maximization (from Chapter 5), especially in terms of the avoidance of unjust outcomes that are sometimes justified on utilitarian grounds. Also, the concerns of whether enhancements are unjust can be linked with my analysis of HET's intrinsic value. It would be interesting to see whether we could alleviate these concerns on the same basis as the concerns about intrinsic badness of HET were alleviated – i.e., by arguing that technologies cannot be unjust in themselves. Nevertheless, we could further ask whether they could be unjust in virtue of their contingent properties by, for instance, being unjustly distributed. This brings us, once again, to the point where we need to admit that conclusions about the moral permissibility of MBE will have to be cross-examined with practical justifications for its implementation.

To conclude, the potential to systematically analyze the moral permissibility of MBE lies beyond the theoretical tug-of-war – it lies in real-life applications that could improve the quality of human life. Although conceptual and normative analysis is an indispensable tool for helping us understand what is and is not permissible, it is only an initial part of a much greater process. As essential as they are, philosophical contributions should be coupled with various interdisciplinary views, the input of which will illuminate the prospects of new and emerging technologies such as MBE. That said, I would like to finish with two simple points to consider when studying MBE in the future. First, let us not prematurely dismiss and overlook the potential benefits of MBE, assuming it were to become feasible and safe. By improving moral motivation, responsibility, altruistic tendencies, and the sense of justice and fairness, MBE could reduce inequality and injustice, prevent harm in terms of reducing crime and violent behavior, raise awareness and prompt action against climate change and environmental degradation, attend to future generations, etc. Second, let us not underestimate the role of ethical theory and its valuable tools as we are developing this intriguing debate (and potentially the technology) further.

---

<sup>165</sup> Some scholars have already applied a Rawlsian approach in dealing with human enhancement. For example, Fritz Allhoff argued that “genetic enhancements are morally permissible if and only if they augment primary goods or create abilities that would lead to their augmentation” (Allhoff 2005, 54). More about employing “equality-conscious models” in the human enhancement debate, such as the Rawlsian maximin approach, can be found in Heinrichs and Stake (2018: 334).



## References

- Agar, N. 2015a. Moral bioenhancement is dangerous. *Journal of Medical Ethics* 41: 343–345.
- Agar, N. 2015b. Moral bioenhancement and the utilitarian catastrophe. *Cambridge Quarterly of Healthcare Ethics* 24: 37–47.
- Alexander, L. 1985. Pursuing the Good-Indirectly. *Ethics* 95(2): 315–332.
- Allhoff, F. 2005. Germ-line genetic enhancement and Rawlsian primary goods. *Kennedy Institute of Ethics Journal* 15(1): 39–56.
- Allhoff, F., Lin, P., Moor, J., and J. Weckert. 2010. Ethics of Human Enhancement: 25 Questions & Answers. *Studies in Ethics, Law, and Technology* 4 (1): 1–39.
- Al-Shawaf, L. et al. 2015. Human Emotions: An Evolutionary Psychological Perspective. *Emotion Review* 1–14.
- Anderson, R. T. and C. Tollefsen. 2008. Biotech Enhancement and Natural Law. *The New Atlantis* 20: 79–103.
- Andreou, C. 2018. Self-paternalism. In Grill, K. and Hanna, J. (Eds.) *The Routledge Handbook of the Philosophy of Paternalism* (1st ed.). Routledge: 59–65.
- Annas, G. 2002. Cell division. *Boston Globe*, 21 April, D1. Available online at [http://www.genetics-and-society.org/resources/items/20020421\\_globe\\_annas.html](http://www.genetics-and-society.org/resources/items/20020421_globe_annas.html).
- Archer, A. 2016. Moral Enhancement and Those Left Behind. *Bioethics* 30: 500–510.
- Babić, J. 2000. Justifying Forgiveness. *Peace Review* 12(1): 87–93.
- Baron-Cohen, S. 2003. *The essential difference*. Basic Books, New York.
- Barry, B. 1977. Justice between Generations, in Law, Morality and Society. *Essays in Honor of H. L. A. Hart*, P.M.S. Hacker and Joseph Raz (eds.), Clarendon Press, Oxford: 268–28.
- Bartz, J.A., Zaki, J., Bolger, N., Ochsner, K.N. 2011. Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Science* 15: 301–309.

- Bauer, K. 2018. Cognitive Self-Enhancement as a Duty to Oneself: A Kantian Perspective. *The Southern Journal of Philosophy* 56: 36-58.
- Beauchamp, T. L. 2004. Does Ethical Theory Have a Future in Bioethics? *Journal of Law, Medicine and Ethics* 32 (2): 209-217.
- Beauchamp, T. L. 2015. Are we unfit for the future?. *Journal of Medical Ethics* 41 (4): 346-348.
- Beauchamp, T. L., Childress, J.F. 1994. *Principles of biomedical ethics* [4th ed]. New York: Oxford University Press.
- Beck, B. 2014. Conceptual and practical problems of moral enhancement. *Bioethics* 29(4): 233-240.
- Bentham, J. 1984. *Introduction to the principles of morals and legislation*. Edited by W. Harrison. Oxford: Oxford University Press.
- Berghmans, R., ter Meulen, R., Malizia, A. and Vos, R. 2011. Scientific, Ethical, and Social Issues in Mood Enhancement. In *Enhancing Human Capacities*, eds. J. Savulescu, R. t. Meulen and G. Kahane, Blackwell Publishing Ltd: 151-165.
- Boer, T., Fischer, R. (eds.). 2013. *Human Enhancement: Scientific, Ethical and Theological Aspects from a European Perspective*. Strasbourg: Church and Society Commission of the Conference of European Churches.
- Borgmann, A. 1984. *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University of Chicago Press.
- Bostrom N., Roache, R. 2008. Ethical issues in human enhancement. In *New Waves in Applied Ethics*, eds. J. Ryberg, T. Petersen and C. Wolf, Pelgrave Macmillan: 120-152.
- Bostrom, N., Sandberg, A. 2009. Cognitive Enhancement: Methods, Ethics, Regulatory Challenges. *Science and Engineering Ethics* 15: 311-341.
- Bostrom, N., Savulescu, J. 2009. Human enhancement ethics: the state of the debate. In *Human Enhancement*, eds. N. Bostrom and J. Savulescu, Oxford University Press, Oxford: 1-21.
- Brandt, R. 1992. *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.
- Brooks, T. 2012. Moral Frankensteins. *AJOB Neuroscience* 3(4): 28-30.
- Brülde, B. 2011. Is mood enhancement a legitimate goal of medicine?. In *Enhancing Human Capacities*, eds. R. ter Meulen, J. Savulescu and G. Kahane, Blackwell Publishing Ltd: 218-229.

- Buchanan, A. 2011a. *Better Than Human – The Promise and Perils of Enhancing Ourselves*. Oxford: Oxford University Press.
- Buchanan, A. 2011b. *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford: Oxford University Press.
- Buchanan, A., D. W. Brock, N. Daniels, and D. Wikler. 2000. *From Chance to Choice*. Cambridge: Cambridge University Press.
- Buss, D. (ed). 2005. *The Handbook of Evolutionary Psychology*. 2<sup>nd</sup> edition. Volume 1, Foundations. Wiley, Hoboken.
- Buss, D. 2007. *Evolutionary Psychology - The New Science of the Mind*. 3<sup>rd</sup> edition. Pearson Education, Boston.
- Campbell, F. 2009. *Contours of ableism: The production of disability and abledness*. New York: Palgrave Macmillan.
- Canton, J. 2003. The impact of convergent technologies and the future of business and the economy. In *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, M.C. Roco, and W.S. Bainbridge, Kluwer Academic Publishers: 71–78.
- Carman, M. 2021. The limits of direct modulation of emotion for moral enhancement. *Bioethics* 35: 192–198.
- Carter, S. 2017. A Kantian ethics approach to moral bioenhancement. *Bioethics* 31: 683–690.
- Casal, P. 2015. On not taking men as they are: reflections on moral bioenhancement. *Journal of Medical Ethics* 41 (4): 340–342.
- Christiansen, J. B., Leigh, I. 2002. *Cochlear implants in children: Ethics and choices*. Washington DC: Gallaudet University Press.
- Clarke, S. 2016. A Brave New World: Understanding the Ethics of Human Enhancement. *Australian Quarterly* Oct-Dec: 29–33.
- Clewis, R.R. 2017. Does Kantian Ethics Condone Mood and Cognitive Enhancement?. *Neuroethics* 10: 349–361.
- Confer, J.C., Easton, J. A., Fleischman, D. S., Goetz, C. D., Lewis, D. M. G., Perilloux, C., and Buss, D. M. 2010. Evolutionary psychology: Controversies, questions, prospects, and limitations. *American Psychologist* 65(2): 110–126.

- Cosmides, L, Tooby J. 2000. Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (eds), *Handbook of emotions*, 2nd ed. The Guilford Press, New York: 91–115.
- Crockett, M.J. 2014. Moral bioenhancement: a neuroscientific perspective. *Journal of Medical Ethics* 40(6): 370–371.
- Crockett, M. J., Apergis-Schoute, A., Herrmann, B., Lieberman, M.D., Müller, U., Robbins, T. W., and L. Clark 2013. Serotonin Modulates Striatal Responses to Fairness and Retaliation in Humans. *The Journal of Neuroscience* 33(8): 3505–3513.
- Crutchfield, P. 2021. *Moral Enhancement and the Public Good*. New York, NY, USA: Routledge.
- Cullity, G. 2003. Asking Too Much. *The Monist* 86(3): 402–418.
- Cushman, F., Young, L., Greene, J.D. 2010. Multi-System Moral Psychology. In J.M. Doris (ed.), *The Moral Psychology Handbook*. Oxford University Press, Oxford.
- Danaher, J. 2018. Moral Enhancement and Moral Freedom: A Critique of the Little Alex Problem. *Royal Institute of Philosophy Supplement* 83: 233–250.
- Daniels, N. 2000. Normal Functioning and the Treatment-Enhancement Distinction. *Cambridge Quarterly of Healthcare Ethics* 9 (3): 309–322.
- Daniels, N. 2009. Can anyone really be talking about ethically modifying human nature? In *Human Enhancement*, ed. N. Bostrom and J. Savulescu, Oxford University Press, Oxford: 25–42.
- De Dreu, C.K.W. 2011. Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1262–1266.
- De Dreu, C.K.W. et al. 2010. The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328: 1408–1411.
- DeGrazia, D. 2012. Genetic enhancement, post-persons and moral status: a reply to Buchanan. *Journal of Medical Ethics* 38: 135–139.
- DeGrazia, D. 2014. Moral enhancement, freedom, and what we (should) value in moral behavior. *Journal of Medical Ethics* 40(6): 361–368.
- Diamond, C. 1982. Anything but Argument. *Philosophical Investigations* 5 (1): 23–41.
- Doris, J.M. (ed.) 2010. *The Moral Psychology Handbook*. Oxford University Press, Oxford.



- Douglas, T. 2008. Moral Enhancement. *Journal of Applied Philosophy* 25 (3): 228-245.
- Douglas, T. 2013. Moral enhancement via direct emotion modulation: A reply to John Harris. *Bioethics* 27(3): 160-168.
- Drake, N. 2016. Is moral bioenhancement dangerous?. *Journal of Medical Ethics* 42(1): 3-6.
- Earp, B. D., Douglas, T., and Savulescu, J. 2018. Moral neuroenhancement. In S. Johnson & K. Rommelfanger (eds.), *Routledge Handbook of Neuroethics*. Routledge, New York: 166-184.
- Ebels-Duggan, K. 2019. Beyond Words: Inarticulable Reasons and Reasonable Commitments. *Philosophy and Phenomenological Research* 98 (3): 623-641.
- Engelbart, D.C. 1962. Augmenting Human Intellect: A Conceptual Framework. *Summary Report AFOSR-3223*. Available online <http://www.bootstrap.org/augdocs/friedewaldo30402/augmentinghumanintellect/iintroduction.html> [Accessed April 20 2019].
- Fabiano, J. 2021. Virtue Theory for Moral Enhancement. *AJOB Neuroscience* 12(2-3): 89-102.
- Fahmy, M. S. 2018. Kantian Perspectives on Paternalism. In Jason Hanna & Kalle Grill (eds.), *The Routledge Handbook of the Philosophy of Paternalism*. Routledge, New York: 96-107.
- Fröding, B. 2013. *Virtue Ethics and Human Enhancement*. Springer Briefs in Ethics.
- Fukuyama, F. 2002. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York: Farrar, Straus, and Giroux.
- Gibson, R.B. 2021. The epidemiology of moral bioenhancement. *Medicine, Health Care and Philosophy* 24: 45-54.
- Ginsberg, Y., Långström, N., Larsson, H., and Lichtenstein, P. 2013. ADHD and criminality: Could treatment benefit prisoners with ADHD who are at higher risk of reoffending? *Expert Review of Neurotherapeutics* 13(4): 345-348.
- Giubilini, A., Savulescu, J. 2018. The Artificial Moral Advisor. The Ideal Observer Meets Artificial Intelligence. *Philosophy of Technology* 31: 169-188.
- Glannon, W. 2018. Moral Enhancement as a Collective Action Problem. *Royal Institute of Philosophy Supplement* 83: 59-85.
- Gligorov, N. 2018. The Cognitive Basis of Commonsense Morality. *Journal of Cognitive Enhancement* 2: 369-376.

- Greene, J. D. 2007. Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences* 11: 322–323.
- Habermas, J. 2003. *The Future of Human Nature*. Cambridge: Polity.
- Haidt, J. 2012. *The righteous mind: why good people are divided by politics and religion*. New York: Random House.
- Hardcastle, V. G. 2018. Lone Wolf Terrorists and the Impotence of Moral Enhancement. *Royal Institute of Philosophy Supplement* 83: 271-291.
- Harris, J. 2007. *Enhancing Evolution: The Ethical Case for Making Better People*. Princeton: Princeton University Press.
- Harris, J. 2009. Enhancements are a moral obligation. In *Human Enhancement*, eds. N. Bostrom and J. Savulescu, Oxford University Press, Oxford: 131-154.
- Harris, J. 2011. Moral enhancement and freedom. *Bioethics* 25(2): 102–111.
- Harris, J. 2012. 'Ethics is for bad guys!' Putting the 'Moral' into Moral Enhancement. *Bioethics* online, 1467-8519.
- Harris, J. 2013. Moral Progress and Moral Enhancement. *Bioethics* 27(5): 285-290.
- Harris, J. 2016. *How to be good – The possibility of Moral Enhancement*. Oxford: Oxford University Press.
- Harman E. 2015. Morality Within the Realm of the Morally Permissible. In: *Oxford Studies in Normative Ethics, Volume 5*. Oxford University Press.
- Hauskeller, M. 2013. *Better Humans? Understanding the Enhancement Project*. Durham: Acumen.
- Hauskeller, M. 2017. Moral bioenhancement and the little Alex problem. *Cambridge Quarterly of Healthcare Ethics* 26 (3): 365–376.
- Heinrichs, J.H., Stake, M. 2018. Enhancement: Consequentialist Arguments. *Zeitschrift für Ethik und Moralphilosophie* 1: 321–342.
- Hickey, C. 2017. Biomedical Enhancement and the Kantian Duty to Cultivate Our Talents. *Journal of Value Inquiry* 51: 165–185.
- Holm, S. 1995. Not just autonomy--the principles of American biomedical ethics. *Journal of Medical Ethics* 21: 332-338.
- Insel, T.R., Fernald, R.D. 2004. How the brain processes social information: Searching for the social brain. *Annual Review of Neuroscience* 27: 697-722.

- Jebari, K. 2014. What to Enhance: Behaviour, Emotion or Disposition?, *Neuroethics* 7: 253–261.
- Jensen, S. R., Nagel, S., Brey, P., Ditzel T., Rodrigues, R., Broadhead, S. and D. Wright. 2018. SIENNA D3.1: State-of-the-art Review: Human Enhancement (Version V1.1).
- Jotterand, F. 2011. ‘Virtue engineering’ and moral agency: will post-humans still need the virtues?. *AJOB Neuroscience* 2(4): 3–9.
- Juengst, E.T. 1998. What does enhancement mean? In *Enhancing Human Traits: Ethical and Social Implications*, ed. E. Parens, Georgetown University Press: Washington D.C.: 29-47.
- Juengst, E.T. 2009. What’s taxonomy got to do with it? ‘Species integrity’, Human rights, and science policy. In *Human Enhancement*, eds. N. Bostrom and J. Savulescu, Oxford University Press, Oxford: 43-58.
- Kagan, S. 1984. Does Consequentialism Demand too Much? Recent Work on the Limits of Obligation. *Philosophy & Public Affairs* 13(3): 239-254.
- Kagan, S. 1998. Rethinking Intrinsic Value. *Journal of Ethics* 2: 277-297.
- Kahane, G., Everett, J.A.C., Earp, B.D., Caviola, L., Faber, N.S., Crockett, M.J., Savulescu J. 2018. Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review* 125(2): 131-164.
- Kahane, G., Savulescu, J. and R. Meulen, eds. 2011. *Enhancing Human Capacities*. Oxford: Wiley-Blackwell.
- Kamm, F. M. 2005. Is There a Problem with Enhancement? *The American Journal of Bioethics* 5 (3): 5-14.
- Kant, I. 1960. *Religion within the Limits of Reason Alone*, translated by Theodore M. Green and Hoyt H Hudson. New York: Harper Torchbooks.
- Kant, I. 1996. *Groundwork of the Metaphysics of Morals*, trans. and ed. by Mary J. Gregor, in *Practical Philosophy: The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.
- Kant, I. 2002. *Critique of Practical Reason*, trans. Werner S. Pluhar, intro. Stephen Engstrom. Hackett Publishing Company.
- Kass, L. 2003. Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection. *The New Atlantis* 1 (Spring): 9-28.
- Korsgaard, C. 1983. Two Distinctions in Goodness. *The Philosophical Review* 92 (2): 169-195.

- Korsgaard, C. 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Korsgaard, C. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E. 2005. Oxytocin increases trust in humans. *Nature* 435: 673–676.
- Kudlek, K. 2019. The Role of Emotion Modulation in the Moral Bioenhancement Debate. *Topoi* 38: 113–123.
- Kudlek, K. 2021. Is human enhancement intrinsically bad?. *Medicine, Health Care and Philosophy* 24: 269–279.
- Kudlek, K. 2022a. On the uneasy alliance between moral bioenhancement and utilitarianism. *Bioethics* 36: 210–217.
- Kudlek, K. 2022b. Challenges in the human enhancement debate: a critical review. *Techné: Research in Philosophy and Technology* 26 (2) [forthcoming].
- Kudlek, K. 2022c. Towards a systematic evaluation of moral bioenhancement. *Theoretical Medicine and Bioethics* [forthcoming].
- Kudlek, K., Smith, P.T. 2022. The Kantian promise and peril of moral bioenhancement. *Journal of Applied Philosophy*. DOI: <https://doi.org/10.1111/japp.12575>.
- Landeweerd, L. 2011. Asperger's Syndrome, Bipolar Disorder and the Relation between Mood, Cognition, and Well-Being. In *Enhancing Human Capacities*, eds. J. Savulescu, R. t. Meulen and G. Kahane, Blackwell Publishing Ltd.: 207–217.
- Lane, A., Luminet, O., Nave, G., and Mikolajczak, M. 2016. Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. *Journal of Neuroendocrinology* 28(4): 1–15.
- Lara, F. and Deckers, J. 2020. Artificial Intelligence as a Socratic Assistant for Moral Enhancement. *Neuroethics* 13: 275–287.
- Lewis, M. et al 2008. *Handbook of Emotions*. The Guilford Press, New York.
- Lindsay, R. A. 2012. The debate over enhancements. *Free Inquiry* 32 (1): 8–21.
- Margari, F., Craig, F., Margari, L., Matera, E., Lamanna, A.L., Lecce, P.A., . . . Carabellese, F. 2014. Psychopathology, symptoms of attention-deficit/hyperactivity disorder, and risk factors in juvenile offenders. *Neuropsychiatric Disease and Treatment* 11(1): 343–352.

- Miah, A. 2012. Ethics issues raised by human enhancement. Available online <https://www.bbvaopenmind.com/en/articles/ethics-issues-raised-by-human-enhancement/>. [Accessed April 20 2019].
- Mill, J.S. 1864. *Utilitarianism*. Second edition. Longman, Green, Longman, Roberts, and Green: London.
- Moore, G. E. 1993. *Principia Ethica*. Cambridge: Cambridge University Press.
- Mukerji, N. and Nida-Rümelin, J. 2014. Towards a Moderate Stance on Human Enhancement. *Humana.Mente Journal of Philosophical Studies* 26: 17–33.
- Murphy, M. 2019. The Natural Law Tradition in Ethics. In *The Stanford Encyclopedia of Philosophy* (Summer 2019 ed.), ed. Edward N. Zalta. <https://plato.stanford.edu/archives/sum2019/entries/natural-law-ethics/>. Accessed 12 December 2020.
- Nave, G., Camerer, C., and McCullough, M. 2015. Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science* 10(6): 772–789.
- Nesse, R.M. 1990. Evolutionary Explanations of Emotions. *Human Nature* 1(3):261-289
- Nesse, R.M. 1991. What good is feeling bad?: The evolutionary benefits of psychic pain. *The Sciences* 30-37.
- Nesse, R.M., Williams, G.C. 1994. *Why we get sick*. New York: Times Books Random House.
- Pacholczyk, A. 2011. Moral enhancement: what is it and do we want it?. *Law, Innovation and Technology* 3(2): 251-277.
- Panksepp, J., Panksepp, J.B. 2000. The Seven Sins of Evolutionary Psychology. *Evolution and Cognition* 6(2): 108-131.
- Parens, E. 1998. Is better always good? The Enhancement Project. *Hastings Center Report* 28: s1-s17.
- Paxton, J. M., Bruni, T., and Greene, J. D. 2014. Are ‘counterintuitive’ deontological judgments really counterintuitive? An empirical reply to. *Social Cognitive and Affective Neuroscience* 9: 1368–1371.
- Pellegrino, E.D. 2004. Biotechnology, Human Enhancement, and the Ends of Medicine. In *The Center for Bioethics and Human Dignity Website*. Available online from <http://www.cbhd.org/resources/biotech/pellegrino2004-11-30.htm>. [Accessed April 20 2019].

- Persson, I., Savulescu, J. 2008. The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy* 25(3): 162-177.
- Persson, I., Savulescu, J. 2011. Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement. *Bioethics* 27(3): 124-131.
- Persson, I., Savulescu, J. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.
- Persson, I., Savulescu, J. 2015a. Summary of *Unfit for the future*. *Journal of Medical Ethics* 41(4): 338-9.
- Persson, I., Savulescu, J. 2015b. Reply to commentators on *Unfit for the future*. *Journal of Medical Ethics* 41(4): 348-352.
- Persson, I., Savulescu, J. 2018. The Moral Importance of Reflective Empathy. *Neuroethics* 11: 183-193.
- Persson, I., Savulescu, J. 2019. The Duty to be Morally Enhanced. *Topoi* 38: 7-14.
- Portmore, D. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.
- President's Council on Bioethics. 2003. *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. Washington DC: President's Council on Bioethics.
- Prinz J.J., Nichols, S. 2010. Moral Emotions. In J.M. Doris (ed.), *The Moral Psychology Handbook*. Oxford University Press, Oxford.
- Rachels, J. 1991. When philosophers shoot from the hip. *Bioethics* 5: 67-71.
- Rachels, J. 2003. *The Elements of Moral Philosophy*. Boston: McGraw-Hill.
- Radke, S., de Bruijn E.R.A. 2012. The other side of the coin: oxytocin decreases the adherence to fairness norms. *Frontiers in Human Neuroscience* 6(193): 1-7.
- Rakić, V. 2015. We Must Create Beings with Moral Standing Superior to Our Own. *Cambridge Quarterly of Healthcare Ethics* 24: 58-65 .
- Raus, K., Focquaert, F., Schermer, M., Specker, J., Sterckx, S. 2014. On defining moral enhancement: a clarificatory taxonomy. *Neuroethics* 7: 263-273.
- Rawls, J. 1971. *A Theory of Justice – Revised Edition*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Rønnow-Rasmussen, T. and M.J. Zimmerman, eds. 2005. *Recent Work on Intrinsic Value*. Dordrecht: Springer.

- Rueda, J. 2020. Climate Change, Moral Bioenhancement and the Ultimate Mostropic. *Ramon Llull Journal of Applied Ethics* 1(11): 277-03.
- Rueda, J., Lara, F. 2020. Virtual reality and empathy enhancement: Ethical Aspects. *Frontiers in Robotics and AI* 7: 506984.
- Rumbold, B. 2017. Tying oneself to the mast: One necessary cost to morally enhancing oneself biomedically. *Bioethics* 31: 543– 551.
- Sandberg, A. 2011. Cognition Enhancement–Upgrading the Brain. In J. Savulescu, R. ter Meulen and G. Kahane (Eds.) *Enhancing Human Capacities*, Wiley-Blackwell, Oxford: 71-91.
- Sandel, M.J. 2004. The case against perfection. *The Atlantic Monthly* 293(3): 50–65.
- Sandel, M.J. 2007. *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge: Harvard University Press.
- Sandel, M.J. 2009. The case against perfection: What’s wrong with designer children, bionic athletes, and genetic engineering. In *Human Enhancement*, eds. N. Bostrom and J. Savulescu, Oxford University Press, Oxford: 71-89.
- Savulescu, J. 2006. Justice, Fairness, and Enhancement. *Annals of the New York Academy of Sciences* 1093(1): 321–338.
- Savulescu, J. 2009. Genetic Interventions and The Ethics of Enhancement of Human Beings. In *The Oxford Handbook of Bioethics*, ed. B. Steinbock, Oxford University Press, Oxford: 517-535.
- Savulescu, J. 2010. Human liberation: Removing biological and psychological barriers to freedom. *Monash Bioethics Review* 29 (1): 4.1-4.18.
- Savulescu, J., Persson, I. 2012. Moral enhancement, freedom, and the god machine. *Monist* 95(3): 399–421.
- Savulescu, J., Sandberg, A., and G. Kahane. 2011. Well-Being and Enhancement. In *Enhancing Human Capacities*, ed. Julian Savulescu, Ruud ter Meulen, and Guy Kahane, Blackwell Publishing Ltd. Hoboken: 1-18.
- Schaefer, G. O. 2015. Direct vs. indirect moral enhancement. *Kennedy Institute of Ethics Journal* 25(3): 261-289.
- Schermer, M., Bolt, I. 2011. What’s in a Name? ADHD and the Gray Area between Treatment and Enhancement. In *Enhancing Human Capacities*, eds. R. ter Meulen, J. Savulescu and G. Kahane, Blackwell Publishing Ltd.: 179-193.
- Schnall, S., Haidt, J., Clore, G. 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34: 1096–1109.

- Sen, A. K. 1979. Utilitarianism and Welfarism. *The Journal of Philosophy* 76: 463–89.
- Shafer-Landau, R. 2012. *The fundamentals of ethics*. New York: Oxford University Press.
- Shiffrin, S. V. 2000. Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy & Public Affairs* 29: 205–250.
- Shiffrin, S. V. 2017. The Moral Neglect of Negligence. In *Oxford Studies in Political Philosophy*, Volume 3, edited by David Sobel, Peter Vallentyne, and Steven Wall. Oxford University Press, Oxford: 197–228.
- Shook, J.R. 2012. Neuroethics and the possible types of moral enhancement, *AJOB Neuroscience* 3(4): 3–14.
- Shook, J.R. and Giordano, J.J. 2017. Moral Bioenhancement for Social Welfare: Are Civic Institutions Ready? *Frontiers in Sociology* 2: 21.
- Sidgwick, H. 1907. *The Methods of Ethics*. London: MacMillan.
- Singer, P. 1972. Famine, Affluence and Morality. *Philosophy and Public Affairs* 1(3): 229–243.
- Singer, P. 2005. Ethics and Intuitions. *The Journal of Ethics* 9(3/4): 331–352.
- Sorell, T. 2011. The Limits of Principlism and Recourse to Theory: The Example of Telecare. *Ethical Theory and Moral Practice* 14: 369–382.
- Sparrow, R. 2014a. Better living through chemistry? A reply to Savulescu and Persson on ‘moral enhancement’. *Journal of Applied Philosophy* 31(1): 23–32.
- Sparrow, R. 2014b. Egalitarianism and moral bioenhancement. *The American Journal of Bioethics* 14(4): 20–28.
- Specker, J., Focquaert, F., Raus, K., Sterckx, S., Schermer, M. 2014. The ethical desirability of moral bioenhancement: a review of reasons. *BMC Medical Ethics* 15: 67.
- Stich, S., Doris, J.M., Roedder, E. 2010. Evolutionary vs. Psychological Altruism. In J.M. Doris (ed.), *The Moral Psychology Handbook*. Oxford: Oxford University Press.
- Sussman, D. 2001. *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics* New York: Routledge.
- Sverdlik, S. 2011. *Motives and Rightness*. Oxford: Oxford University Press.



- Timmons, M. 2012. *Moral Theory: An Introduction* (2nd ed). Rowman and Littlefield publishers, inc. UK.
- Tooby, J. 1985. The emergence of evolutionary psychology. In D. Pines (ed), *Emerging syntheses in science: Proceedings of the founding workshops of the Santa Fe Institute*. NM: Santa Fe Institute, Santa Fe: 1–6.
- Tooby, J., Cosmides, L. 1990. The past explains the present: emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* 11: 375–424.
- Tooby, J., Cosmides, L. 2008. The Evolutionary Psychology of the Emotions and Their relationship to Internal Regulatory Variables in: Lewis, M., J. M. Haviland-Jones & L. Feldman Barrett (eds) *Handbook of Emotions*. The Guilford Press, New York: 114–137.
- Triviño, J.L.P. 2013. On the need of moral enhancement. A critical comment of ‘Unfit for the future’ of I. Persson and J. Savulescu. *Dilemata* 12: 261–269.
- Tse, W.S., Bond, A.J. 2002. Serotonergic intervention affects both social dominance and affiliative behaviour. *Psychopharmacology* 161(3): 324–330.
- Vedder, A. 2019. An Obligation to Enhance? *Topoi* 38(1): 49–52.
- Vujošević, M. 2019. Kant’s account of moral weakness. *European Journal of Philosophy* 27(1): 40–54.
- Vujošević, M. 2020. Kant’s Conception of Moral Strength. *Canadian Journal of Philosophy* 50(4): 539–553.
- Watson, G. 1975. Free Agency, *Journal of Philosophy* 72 (8): 205–220.
- Wilks, A. 2018. Kantian Challenges for the Bioenhancement of Moral Autonomy. *Royal Institute of Philosophy Supplement* 83: 121–143.
- Wood, R. M., Rilling, J. K., Sanfey, A. G., Bhagwagar, Z., & Rogers, R. D. 2006. Effects of tryptophan depletion on the performance of an iterated prisoner’s dilemma game in healthy adults. *Neuropsychopharmacology* 31(5): 1075–1084.
- Woodard, C. 2019. *Taking Utilitarianism Seriously*. Oxford: Oxford University Press.
- Zak, P.J., Kurzban, R., Matznera, W.T. 2004. The Neurobiology of Trust. *Annals of the New York Academy of Sciences* 1032: 224–227.
- Zimmerman, M.J. 2001. *The nature of Intrinsic value*. Rowman and Littlefield publishers, Inc. USA.



# Summary

The debate on moral bioenhancement (MBE) – broadly understood as a set of emerging technologies aimed at improving human morality – has mainly focused on the potential dangers and benefits of these interventions. However, there is little to no consensus as to what constitutes MBE and how to define it (is it boosting our altruism or diminishing violent aggression?); whether we urgently need it (to combat climate change and threats of mass destruction); and, most importantly, whether we should allow it even if it were to become feasible and safe (is it *morally* permissible)? This thesis is not concerned with MBE’s defining features or biotechnological feasibility. Instead, it attempts to shed some new light on the debate by addressing several theoretical and normative challenges closely related to MBE’s moral permissibility. These challenges are addressed through the lens of ethical theory, i.e., by placing bioethical debate on MBE into an explicit dialogue with ethical theory.

This dissertation is generally concerned with *how can we use ethical analysis to improve the bioethical debate on MBE and our understanding of what is or is not morally permissible?* It attempts to prove that employing ethical analysis can bring us closer to resolving some of the pertinent issues, and it does so through exploring three main agendas:

- i) What are the main challenges in the MBE debate and how can we move towards a systematic ethical evaluation of enhancement projects?
- ii) What does a close theoretical examination of enhancement’s intrinsic and extrinsic properties imply about its coherence and feasibility, and how is this relevant for its moral permissibility?
- iii) How does MBE align with prevailing moral norms and does it conflict with basic moral values?

These questions are addressed across six stand-alone but jointly related discussions (articles) that contribute to the philosophical and bioethical debate on MBE.

The thesis identifies (where) there is discourse polarization and misunderstanding and argues that employing ethical analysis (relying on normative ethics and metaethics) will minimize vagueness and arbitrariness and provide more

systematic guidance in assessing moral permissibility of MBE (Chapters 1 and 2). It further challenges the claim that enhancements may be intrinsically bad and, by extension, morally impermissible by looking into philosophical theories of intrinsic value (Chapter 3). Employing theories of intrinsic value shows that there are no decisive reasons to think enhancements are intrinsically bad and morally impermissible by virtue of their intrinsic properties. However, even if enhancements are not bad in themselves, some types could have undesirable or even harmful effects. For example, the thesis challenges the claim that MBE by emotion modulation may be feasible and effective (Chapter 4). By relying on neuroscientific and evolutionary-psychological views of emotions, it is argued that MBE cannot be based (solely) on emotion modulation because this may have counterproductive or undesirable effects. The dissertation demonstrates how the application of normative ethical theories can help us further illuminate moral permissibility in the MBE debate. It shows that, despite popular assumptions, MBE is not so obviously supported by utilitarianism (Chapter 5) and opposed by Kantianism (Chapter 6). Although both accounts hold important promises and perils, it is argued that MBE does not explicitly contradict fundamental moral values which are typically supported by utilitarian and Kantian ethical doctrines.

This dissertation is a proof of concept that the theoretical and normative analyses are indispensable tools for developing the MBE debate further and helping us understand what is and is not permissible. It suggests that, on net, the permissibility of MBE will depend on the context but we currently do not have decisive reasons to think it is utterly impermissible under realistic scenarios.

# Samenvatting

## De Ethische Analyse van Morele Mensverbetering: Theoretische en Normatieve Perspectieven

Het bio-ethische debat omtrent *morele mensverbetering* (MMV) – gedefinieerd als een set opkomende technologieën die tot doel hebben de menselijke morele eigenschappen te verbeteren – heeft zich met name gespitst op de potentiële voordelen en de risico's ervan. Echter, er is tot nog toe weinig tot geen consensus over wat MMV precies inhoudt (het bevorderen van altruïsme of juist het onderdrukken van agressie?); de noodzakelijkheid ervan als effectief middel tegen bijvoorbeeld klimaatverandering; en, bovenal, de *morele* toelaatbaarheid ervan, zelfs in het geval dat de technologieën effectief en veilig zijn bevonden. Dit proefschrift focust niet op de essentiële eigenschappen van MMV noch haar technologische haalbaarheid, maar probeert juist een vernieuwende bijdrage aan het debat te leveren door verschillende theoretische en normatieve vraagstukken te behandelen die samenhangen met de morele toelaatbaarheid van MMV. Deze kwesties worden geanalyseerd vanuit een ethisch-theoretisch kader, door het bio-ethische debat omtrent MMV in een dialoog te plaatsen met verschillende ethische theorieën.

Dit proefschrift stelt de volgende vraag centraal: Hoe kunnen we ethische analyse inzetten om een bijdrage te leveren aan het bio-ethische debat omtrent MMV en een beter begrip te verkrijgen van wat wel en niet moreel toelaatbaar is? Het tracht te aan te tonen dat de inzet van ethische theorieën ons kan helpen bij het oplossen van een aantal belangrijke kwesties en het doet dat door antwoord te geven op de volgende drie deelvragen.

- i) Wat zijn de belangrijkste uitdagingen in het debat omtrent MMV en hoe kunnen we een systematische ethische evaluatie van MMV-projecten tot stand laten komen?
- ii) Wat vertelt een theoretische bestudering van de intrinsieke en extrinsieke eigenschappen van MMV over de coherentie en haalbaarheid van MMV (en hoe is dit relevant voor de morele toelaatbaarheid van MMV)?
- iii) Hoe verhoudt MMV zich tot gangbare morele normen en conflicteert het met fundamentele morele waarden?

Deze vragen worden in dit proefschrift beantwoord in zes opzichzelfstaande, maar samenhangende verhandelingen (gebaseerd op zes door de auteur geschreven wetenschappelijke artikelen) die allemaal een bijdrage leveren aan het filosofische en bio-ethische debat omtrent MMV.

Dit proefschrift constateert dat er sprake is van polarisatie en misverstanden in het debat en bepleit het gebruik van normatieve en meta-ethische analyses om onduidelijkheden en willekeur te beperken en voor meer systematische begeleiding te zorgen in de beoordeling van de morele toelaatbaarheid van MMV (hoofdstukken 1 en 3). Verder bestrijdt dit proefschrift de bewering dat mensverbeteringen per definitie slecht en daarom moreel onacceptabel zijn door te kijken naar theorieën omtrent intrinsieke waarde (hoofdstuk 2). Het toont aan dat er geen doorslaggevende redenen zijn om te stellen dat mensverbeteringen intrinsiek slecht en moreel onacceptabel zijn op grond van deze theorieën. Echter, zelfs als mensverbeteringen niet per definitie slecht zijn, kunnen sommige soorten mensverbetering ongewenste en schadelijke effecten hebben. Dit proefschrift haalt als voorbeeld emotiemodulatie aan en bestrijdt de bewering dat dit een effectieve en haalbare vorm van MMV zou kunnen zijn (hoofdstuk 4). Aan de hand van inzichten uit de neurowetenschap en evolutionaire psychologie toont het aan dat emoties niet voldoen aan bepaalde noodzakelijke vereisten en dat modulatie ervan contraproductieve effecten zal hebben. Dit proefschrift laat verder zien hoe verschillende normatieve ethische theorieën tot verschillende conclusies kunnen leiden ten aanzien van de morele toelaatbaarheid van MMV. Het laat zien dat, in tegenstelling tot wat vaak wordt aangenomen, de morele toelaatbaarheid van MMV niet zo vanzelfsprekend is binnen utilistische theorieën (hoofdstuk 5) en dat Kantiaanse theorieën de toepassing van MMV regelrecht afwijzen (hoofdstuk 6). Hoewel beide perspectieven belangrijke beloften en gevaren bevatten, kan worden aangetoond dat MMV niet leidt tot expliciete schending van fundamentele morele waarden (zoals die worden beschreven door utilistische en Kantiaanse ethische theorieën).

Dit proefschrift toont als 'proof of concept' aan dat theoretische en normatieve analyses onmisbare hulpmiddelen zijn in de ontwikkeling van het bio-ethisch debat omtrent MMV en in de evaluatie van wat wel en wat niet aanvaardbaar is. Het laat zien dat, per saldo, de morele toelaatbaarheid van MMV zal afhangen van de specifieke context en er momenteel geen doorslaggevende redenen zijn om te aan te nemen dat MMV regelrecht ontoelaatbaar is onder realistische omstandigheden.

# Sažetak

## Etička Analiza Moralnog Biopoboljšanja: Teorijske i Normativne Perspektive

Rasprava o moralnom biopoboljšanju (MBP) – koje možemo u širem smislu shvatiti kao skup nastajućih tehnologija za unaprijeđenje ljudske moralnosti – je uglavnom usmjerena na moguće prednosti i nedostatke takvih intervencija. Međutim, postoji veliko neslaganje što točno MBP podrazumijeva i kako ga definirati (je li to pojačanje altruizma ili smanjenje agresije?); postoji li prijeka potreba za njim (u borbi protiv klimatskih promjena i prijetnji masovnog uništenja?); te, posebice, je li *moralno* dopustivo, čak i ako postane izvedivo i sigurno? Temeljna određenja i biotehnološka izvedivost MBP ne predstavljaju središnja pitanja ove disertacije. Umjesto toga, ona nastoji unijeti novu perspektivu u raspravu istražujući niz teorijskih i normativnih poteškoća vezanih uz moralnu dopustivost MBP. Tim poteškoćama se pristupa kroz prizmu etičke teorije, tj. stavljajući bioetičku raspravu u izravan dijalog s etičkom teorijom.

Općenito uzevši, ova disertacija istražuje *kako etičku analizu možemo koristiti za unaprijeđenje bioetičke rasprave o MBP i razumijevanja što jest, a što nije moralno dopustivo?* Nastoji pokazati kako nas etička analiza približava rješavanju nekih temeljnih problema, i to kroz tri osnovna smjera istraživanja:

- i) Koji su temeljni izazovi u raspravi o MBP i kako možemo ostvariti napredak ka sustavnom etičkom vrednovanju poboljšanja?
- ii) Što podrobna teorijska analiza intrinzičnih i ekstrinzičnih svojstava MBP otkriva o njegovoj koherentnosti i izvedivosti, te kako ti zaključci utječu na njegovu moralnu dopustivost?
- iii) U kakvom je MBP odnosu s prevladavajućim moralnim normama, te je li u sukobu s temeljnim moralnim vrijednostima?

Prethodna pitanja obrađujem kroz šest samostalnih, ali međusobno povezanih članaka koji pridonose filozofskoj i bioetičkoj raspravi o MBP.

Disertacija identificira dijelove rasprave obilježene polarizacijom i međusobnim nerazumijevanjem, te nastoji pokazati da će upotreba etičke analize (koja se

oslanja na normativnu etiku i metaetiku) smanjiti neodorečenost i arbitrarnost te pružiti sistematičan okvir za procjenu moralne dopustivosti MBP (Poglavlja 1 i 2). Nadalje, disertacija se oslanja na filozofske teorije intrinzične vrijednosti s ciljem propitivanja tvrdnje da poboljšanja mogu biti intrinzično loša i, samim time, moralno nedopustiva (Poglavlje 3). Uporaba teorija intrinzične vrijednosti ukazuje kako ne postoje čvrsti razlozi za vjerovanje da su poboljšanja intrinzično loša i moralno nedopustiva isključivo na temelju svojih intrinzičnih svojstava. Međutim, čak i ako poboljšanja nisu loša po sebi, neki oblici bi mogli imati nepoželjne ili čak štetne posljedice. Primjerice, disertacija nadalje kritički propituje tvrdnju da bi modulacija emocija mogla biti izvedivo i učinkovito sredstvo (Poglavlje 4). Oslanjajući se na neuroznanstvena i evolucijsko-psihološka shvaćanja emocija, tvrdim da nije moguće MBP (isključivo) temeljiti na modulaciji emocija jer bi to moglo imati neželjene učinke. Disertacija potom pokazuje kako nam primjena normativnih etičkih teorija može pomoći u rasvijetljavanju problema moralne dopustivosti unutar rasprave o MBP. Pokazuje da, usprkos popularnim vjerovanjima, MBP nije u *očitome* skladu s utilitarizmom (Poglavlje 5), niti sukobu s kantovskom etikom (Poglavlje 6). Iako obje analize ukazuju na obećavajuće aspekte, a tako i teškoće, argumentiram u prilog tezi da MBP ne predstavlja izrazitu opasnost temeljnim moralnim vrijednostima koje se primarno promiču kroz utilitarizam i kantovsku etiku.

U cijelosti, disertacija demonstrira nužnost upotrebe teorijske i normativne analize kao neizostavnih sredstava za unaprijeđenje daljnjeg razvoja rasprave o MBP i razumijevanja što jest, a što nije moralno dopustivo. Ovo istraživanje pokazuje da će dopustivost MBP na koncu ovisiti o kontekstu, ali u ovome trenutku nemamo uvjerljive razloge u prilog njegove potpune nedopustivosti.



# Acknowledgments

Like many others, this PhD journey has not been without its struggles, but it would take another book to tell this story. In large, I am proud of my perseverance and I am thankful to all of those who supported me along the way. It is a privilege to express my gratitude to them here, in the final chapter of my book and of my PhD life.

Professionally, I am thankful to ALL of my supervisors – and, oh, the list is long! Therefore, it is only fair to proceed in **chronological** order.

Although this thesis was completed at the University of Twente, my research on moral enhancement started at the University of Zagreb. Tomislav, you were the first person who inspired me to work in philosophy and pursue a PhD in moral enhancement. You thought me so much about academia, academic integrity (and lack thereof), writing, presenting, organizing events... And all of that sprinkled with your inherent cynicism – reminding me never to take myself or this work too seriously. You showed me patience, support and structure when I was only starting, but without ever sugarcoating it. Thanks to you I met many truly great philosophers (like Peter Singer and John Harris, to name a few) whom we hosted at our conferences in Zagreb and who further inspired me to pursue a PhD. And when things got rough, you helped me to move forward and continue my research elsewhere. Hvala na svemu!

Javier, you welcomed me at the University of Granada with the immeasurable, or better said, *incommensurable* warmth. You gave academia a human, friendly face which, honestly, set the standards impossibly high for everyone else. I am thankful for your impeccable feedback on my early writings, for all the hours we spent chatting, for your tour guides around Albaicín, and for an endless amount of jokes and positivity. I am thankful that you supported me to move forward when funding ran out and believed in me when other opportunities appeared. I will always cherish the memories from Granada. ¡Gracias por todo!

Philip B. You kindly invited me to visit the University of Twente back in 2018, when neither of us could even imagine how this will unravel. It meant that I will move to the Netherlands, transfer to UT, acquire external funding, and, finally complete my dissertation with you as my promoter. Not many things have been constant over the past four years, but you were one of them. Thank you for your spot-on, concise feedback, and for always pushing for “high standards” and clarity. I learned a lot about Dutch efficiency from you, and although I occasionally lacked different approaches, yours did bring valuable results. Dank je wel!

Matej, your role in hiring me for the HIRe (!) project, hosted by the Institute of Philosophy in Zagreb, was probably a crucial point for completing this dissertation. Back then, the funding struggles were slowly getting to me, and acquiring this research position meant that I could finally focus on my PhD more steadily. Thank you for trusting me with this position and believing that I will be a good addition to your team. Thank you for giving me all the freedom to pursue my goals however I thought was best. In retrospect, I even wish you had restrained it a little. And, finally, thank you for having my back at the end, when things got especially difficult. Hopefully, we will enjoy many more (small) Greenery meetings.

Per, your role as a daily supervisor was perfectly rounded between giving me professional guidance and being genuinely interested in my work and ideas; while at the same time being incredibly friendly and attentive. Thank you for all that and for always reminding me to take care of myself. I am immensely grateful that you stayed in touch – reading, commenting and checking up on me – long after your formal role has ended. My only regret is that the timing meant we couldn’t make more out of our professional collaboration.

Patrick, I won’t lie, working with you was a rollercoaster, but certainly one memorable ride! We worked through our differences and we made it count. Professionally and intellectually, you are probably the person to whom I’m indebted the most. After all, the paper we co-authored is a substantial part of this thesis – and it is a wonderful piece of work. Thank you for all the feedback and intense discussions. Thank you for teaching me how to channel my disagreement with other scholars and focus on the pragmatic aspects of my academic career: “The best dissertation is a finished dissertation!”. Thanks for all the movie tips, football matches, and witches’ hammers. But most of all, thank you for bearing with me

through all the insecurities and rough patches in my professional and personal life, and we both know there were plenty.

To continue, I must thank all of my colleagues who made this journey more bearable. I first thank my colleagues at the University of Twente, as this is where I spent most of the time writing this dissertation. Apart from Philip, Patrick, and Per, there were many others including Kevin, Owen, Michael, Bas, Ciano, Yasemin, etc., who affected my work in one way or the other. I thank you for all your contributions, be it a hallway chat or constructive criticism. I was lucky to meet many wonderful PhDs coming and going through the department: Philip, Margoth, Sean, Rosalie, Sage, Isaac, Brian, Kishore, Patty, Elisa... to mention just a few. Thanks for all the fun times! Special thanks goes to Ada and Ciano for their efforts in removing many bureaucratic obstacles, and to Seeta for simultaneously cheering me on and rushing me out of the office late at night. I am also thankful to the department and BMS faculty for generously funding numerous conferences and workshops I attended, and providing open access funding for my articles.

I'm thankful to my colleagues at the Institute of Philosophy in Zagreb. Especially to Tvrtko for relentlessly providing answers whenever I needed them and all the dinners he didn't let me pay for. To Viktor for proofreading almost this entire thesis, encouraging me and keeping me in the loop all the time. To Matej and Tomislav for the things already said, and all the others who supported me in any way. Special thanks goes to the Croatian Science Foundation for providing a grant which was essential in carrying out this research. I also want to mention colleagues from the University of Rijeka here, Luca and Marko, for lifting my spirit, especially in the last stretch of this journey. Lovro, thanks for skyping during lockdowns. From my Spanish adventures, I fondly remember Lilian, Javier, Paco, David, Alberto, later on, Jon and Pablo – I am so glad to know you all.

To finalize these professional acknowledgments, I must thank the members of my graduation committee. Although I am momentarily in the dark about what awaits me at the defense, I am immensely grateful to all of you for accepting our invitation to join the graduation committee. So, thank you to Julian, Michael H., Birgit, Ciano, and Michael N. for finding the time in your busy schedules to read these pages and participate in this event. It is truly a privilege to receive your feedback.

I am also grateful to numerous anonymous reviewers and editorial teams who made my work better and worthy of publishing in these prestigious journals. I thank all the scholars whose work I have cited throughout this dissertation, all of those I have met at academic events, and who made an impact on me or my work. You are the true heroes and thank you for your service.

Now, it is time to move to the second category of these acknowledgments, which is by no means a lesser category. Many people have been there for me professionally, but also personally. I will only mention some of them here.

Petar, I know we said no clichés, but I truly could not have done this without you. Staying side by side through a PhD should be higher on the list of social commitments people make. And although our PhD journeys were polar opposites, we kept each other going. It made me very proud to see you become a doctor, although I've never doubted it. On the other hand, I doubted myself a lot, but you provided unconditional support even in the darkest moments. I can't even begin to express my gratitude for all the hours you put into reading my thesis, discussing it with me, and suggesting how to improve it. To this day, I still believe you know it better than I do. And this is why you would make a perfect paronymph, if only you knew how to organize a party. So, sorry, but: priorities.

Philip J. As much as we like to joke about you being the “small Philip”, your role in this PhD journey was anything but small. From the very first day, when I was cluelessly roaming the halls of Cubicus, you made me feel welcome and less alone in this mayo-obsessed culture. You helped me to understand so many of the Dutch quirks and eased my transition into this new life. We quickly evolved from office buddies to genuinely good friends and I hope we stay this way, no matter what happens in the future. Thanks for sticking with me and for patiently listening to my complaints about almost anything. Thanks for all the dinners, the movies, and the travels we've taken together. Hopefully, I will be more available in the future.

Jasmina, you are truly a rockstar and I'm so lucky to have you as a friend. Thanks for all the emotional and moral support you provided in person and virtually, engaging vigorously in the philosophical soap operas I was telling you about. Thanks for your inexhaustible energy, coming over from Berlin only to make my parties feel homier. I will compensate for my absence, promise.

Mom, I want to finish this tractatus with you because, in the end, it all comes down to you. You are my biggest fan, just like you are biologically predisposed to be. Joking aside, I know you are extremely proud of me, although you could never grasp why am I doing any of this. You are right because I do not know either, yet here we are. Let's hope you can find some comfort in telling your friends that your daughter is a doctor now (disclaimer: not the “real” one). I love you, mom, and thanks for your selfless support and understanding. *Mama, završit ću ovaj traktat s tobom jer, na koncu, sve se svodi na tebe. Ti si moj najvjerniji obožavatelj, upravo kako je i biološki zamišljeno. Šalu na stranu, znam da si izuzetno ponosna na mene, iako nikada nisi mogla razumjeti zašto radim išta od ovoga. U pravu si jer ni ja sama ne znam, ali što je tu je. Nadam se da ćeš pronaći utjehu u činjenici da sada možeš reći prijateljicama da ti je kći doktorica (iako ne ona “prava”). Volim te, majko, i hvala na nesebičnoj potpori i razumijevanju.*

Final thanks goes to the rest of my family, biological and extended (Ziggy, this one is for you). I am truly grateful to all of you who were part of this process, even if I haven't mentioned you by name. It has been overwhelming.

*So long, and thanks for all the fish!*



## About the Author

Karolina Kudlek (1986) was born in Sisak, Croatia. She obtained a bachelor's degree in Philosophy and Croatology (2008) and a master's degree in Philosophy (2011) from the University of Zagreb. In her master thesis, she conducted a comparative analysis of ethical egoism and utilitarianism, showing that several reasons for dismissing ethical egoism as a legitimate normative theory apply with equal force to utilitarianism. During her studies, she worked for the Society for the Advancement of Philosophy in Zagreb, where she also managed a philosophical journal *Prolegomena*. During her PhD study, Karolina conducted research at the University of Zagreb and the University of Granada. She completed her PhD dissertation at the University of Twente (2018-2022). In the dissertation she investigated how ethical theory can improve the bioethical debate on biomedical moral enhancement and facilitate our understanding of what is or is not morally permissible.

Karolina was also engaged in several research projects: *Artificial Intelligence and Biotechnology of Moral Enhancement: Ethical Aspects* (at the University of Granada); SIENNA – *Stakeholder-Informed Ethics for New Technologies with High Socio-Economic and Human Rights Impact* (at the University of Twente); and she worked as a research assistant in the project *Harm, Intentions, and Responsibility* (HIRE) hosted by the Institute of Philosophy in Zagreb (2018-2022).

So far, she has presented at over 25 international academic events and her work has appeared in peer-reviewed philosophical and bioethical journals – including *Topoi*; *Medicine, Health Care and Philosophy*; *Bioethics*; *Journal of Applied Philosophy*, *Theoretical Medicine and Bioethics*, etc. Her general research interests are normative and applied ethics, bioethics, philosophy of technology (especially ethics of human enhancement), and moral psychology.

In her free time, she enjoys good movies, stand-up comedy, craft beer, dancing, and city breaks.

# Simon Stevin Series in Ethics of Technology

Delft University of Technology, Eindhoven University of Technology,  
University of Twente & Wageningen University

Editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

## *Books and Dissertations*

- Volume 1: Lotte Asveld, *Respect for Autonomy and Technology Risks*, 2008
- Volume 2: Mechteld-Hanna Derksen, *Engineering Flesh: Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering*, 2008
- Volume 3: Govert Valkenburg, *Politics by All Means: An Enquiry into Technological Liberalism*, 2009
- Volume 4: Noëmi Manders-Huits, *Designing for Moral Identity in Information Technology*, 2010
- Volume 5: Behnam Taebi, *Nuclear Power and Justice between Generations: A Moral Analysis of Fuel Cycles*, 2010
- Volume 6: Daan Schuurbiers, *Social Responsibility in Research Practice: Engaging Applied Scientists with the Socio-Ethical Context of Their Work*, 2010
- Volume 7: Neelke Doorn, *Moral Responsibility in R&D Networks: A Procedural Approach to Distributing Responsibilities*, 2011
- Volume 8: Ilse Oosterlaken, *Taking a Capability Approach to Technology and Its Design: A Philosophical Exploration*, 2013
- Volume 9: Christine van Burken, *Moral Decision Making in Network Enabled Operations*, 2014
- Volume 10: Faridun F. Sattarov, *Technology and Power in a Globalising World: A Political Philosophical Analysis*, 2015
- Volume 11: Gwendolyn Bax, *Safety in Large-Scale Socio-Technological Systems: Insights Gained from a Series of Military System Studies*, 2016



- Volume 12: Zoë Houda Robaey, *Seeding Moral Responsibility in Ownership: How to Deal with Uncertain Risks of GMOs*, 2016
- Volume 13: Shannon Lydia Spruit, *Managing the Uncertain Risks of Nanoparticles: Aligning Responsibility and Relationships*, 2017
- Volume 14: Jan Peter Bergen, *Reflections on the Reversibility of Nuclear Energy Technologies*, 2017
- Volume 15: Jilles Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, 2018
- Volume 16: Taylor William Stone, *Designing for Darkness: Urban Nighttime Lighting and Environmental Values*, 2019
- Volume 17: Cornelis Antonie Zweistra, *Closing the Empathy Gap: Technology, Ethics, and the Other*, 2019
- Volume 18: Ching Hung, *Design for Green: Ethics and Politics for Behavior-Steering Technology*, 2019
- Volume 19: Marjolein Lanzing, *The Transparent Self: a Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*, 2019
- Volume 20: Koen Bruynseels, *Responsible Innovation in Data-Driven Biotechnology*, 2021
- Volume 21: Melis Baş, *Technological Mediation of Politics. An Arendtian Critique of Political Philosophy of Technology*, 2022
- Volume 22: Mandi Astola, *Collective Virtues. A Response to Mandevillian Morality*, 2022
- Volume 23: Karolina Kudlek, *The Ethical Analysis of Moral Bioenhancement. Theoretical and Normative Perspectives*, 2022



## Simon Stevin (1548-1620)

‘Wonder en is gheen Wonder’

This series in the philosophy and ethics of technology is named after the Dutch/Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinary versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, however mysterious, ultimately had a scientific explanation. Hence his dictum ‘Wonder is no Wonder’, which he used on the cover of several of his own books.

What if we could use biomedical science not only to make ourselves stronger, smarter and more beautiful but also to improve our *moral* character? The debate on moral bioenhancement (MBE) – a set of emerging technologies aimed at improving human morality – has been mainly focused on the potential dangers and benefits of these interventions. However, there is little to no consensus as to what constitutes MBE and how to define it; whether we urgently need it; and, whether we should allow it even if it were to become feasible and safe. This dissertation attempts to shed some new light on the debate by addressing several theoretical and normative challenges closely related to MBE’s moral permissibility. It does so by placing the bioethical debate on MBE into an explicit dialogue with ethical theory. This dialogue between ethical theory and bioethics is meant to give us insight into the moral status of MBE technologies, provide methodological guidelines, and help us clear the path for future research. On the whole, the thesis is a proof of concept that the theoretical and normative analyses are indispensable tools for developing the MBE debate further and helping us understand what is and is not morally permissible. It suggests that, on net, the permissibility of MBE will depend on various contingent factors but we currently do not have decisive reasons to think it is utterly impermissible under realistic scenarios.

*‘Wonder en is  
gheen wonder’*

**4TU** ● Centre for Ethics  
and Technology