HUMAN ENHANCEMENT AND TECHNOLOGICAL UNCERTAINTY: ESSAYS ON THE PROMISE AND PERIL OF EMERGING TECHNOLOGY

KARIM A. JEBARI

Doctoral Thesis
Stockholm, Sweden 2014

Essay I explores brain machine interface (BMI) technologies. These make direct communication between the brain and a machine possible by means of electrical stimuli. This essay reviews the existing and emerging technologies in this field and offers an inquiry into the ethical problems that are likely to emerge.

Essay II, co-written with professor Sven-Ove Hansson, presents a novel procedure to engage the public in deliberations on the potential impacts of technology. This procedure, convergence seminar, is a form of scenario-based discussion that is founded on the idea of hypothetical retrospection. The theoretical background and the results of the five seminars are presented.

Essay III discusses moral bioenhancement, an instance of human enhancement that alters a person’s dispositions, emotions or behavior. Moral bioenhancement could be carried out in three different ways. The first strategy is behavioral enhancement. The second strategy, favored by prominent defenders of moral enhancement, is emotional enhancement. The third strategy is the enhancement of moral dispositions, such as empathy and inequity aversion. I argue that we ought to implement a combination of the second and third strategies.

Essay IV considers the possibility and potential desirability of sensory enhancement. It is proposed that existing sensory modalities in vertebrate animals are proof of concept of what is biologically possible to create in humans. Three considerations on the normative aspects of sensory enhancement are also presented in this essay.

Essay V rejects disease prioritarianism, the idea that the healthcare system ought to prioritize the treatment of diseases. Instead, an approach that focuses on what medicine can accomplish is proposed.

Essay VI argues that from the idea that species have an intrinsic value and that humanity has a collective responsibility to protect animal species from extinction, the conclusion that we ought to recreate species follows.

Essay VII argues that unknown existential risks have not been properly addressed. It proposes a heuristic for doing so, and a concrete strategy. This strategy consists in building refuges that could withstand a large number of catastrophic events.
This doctoral thesis consists of an introduction and the following essays:


Jebari, Karim A., “Disease Prioritarianism: A Flawed Principle”. *Submitted manuscript*

Jebari, Karim A., “Should Extinction be Forever?” *Submitted manuscript*


© 2014 by Karim A. Jebari

ISSN 1650-8831

ISBN 978-91-7595-341-0

Printed in Stockholm, Sweden by E-print AB 2014
# CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGMENTS</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1. The History of Transhumanism</td>
<td>2</td>
</tr>
<tr>
<td>2. Transhumanism and Normative Essentialism</td>
<td>6</td>
</tr>
<tr>
<td>3. Summary of the Essays</td>
<td>13</td>
</tr>
<tr>
<td>4. Summary in Swedish</td>
<td>22</td>
</tr>
</tbody>
</table>

## ESSAYS

I. Brain Machine Interface and Human Enhancement - An Ethical Review

II. European Public Deliberation on Brain Machine Interface technology: Five Convergence Seminars

III. What to Enhance: Behavior, Emotion or Disposition?

IV. Sensory Enhancement

V. Should Extinction be Forever? (under review)

VI. Disease Prioritarianism: A Flawed Principle (under review)

VII. Existential Risks: Exploring a Robust Risk Reduction Strategy
ACKNOWLEDGMENTS

I would like to thank my supervisors, Sven-Ove Hansson and Barbro Fröding. I would also like to thank the participants at the philosophy seminar at the Royal Institute of Technology as well as the participants at the CHE-seminar. In particular I would like to thank Niklas Juth, Per Sandin, Seth Baum, Aron Vallinder, Till Grüne-Yanhoff, John Cantwell, Tor Sandqvist, Olle Häggström, Björn Lundgren, Payam Moula, William Bülow, and Niklas Möller for their extensive commentary and ample patience. Joakim Lundborg deserves a special mention for the many hours spent over beers discussing the ideas presented here. I would also like to thank my colleagues that kept me going. Mika Lindblad, my beloved girlfriend, deserves the greatest gratitude for her probing questions, brilliant comments and generosity in the darkest moments of these five years. Finally, I would like to show my immense gratitude and admiration to the many anonymous reviewers who have helped me more than anyone in clarifying my thoughts on these topics.
INTRODUCTION

Humanity has sought ways to overcome its biological limitations since time immemorial. Today, we may finally be entering an age when technology can realise this ambition. Biotechnology, computer science and nanotechnology could radically alter the human condition for better or worse. Known as human enhancement technologies when applied to overcoming human biological limitations, these technologies hold great promise but also pose risks. The challenges associated with the increasing power of technology need to be addressed, particularly with regard to the potential impact of human enhancement. Although traditional probabilistic risk analysis is an important tool in this endeavour, due to the uncertainty associated with technological disruption, it is poorly suited to some aspects of this task (Hansson, 1993). This study uses applied ethics, an interdisciplinary method that seeks to combine an informed scientific review and technology assessment with a rigorous normative analysis. Accordingly, the essays in this study formulate scenarios about possible future developments and aim to clarify and analyse some of the associated ethical concerns.

It is in the nature of technological development that it is both difficult to predict and has the potential to revolutionise the human condition. Due to the potential impact of future human enhancement technology, we cannot afford the luxury of waiting for the technology to emerge before thinking about the ethical consequences of its widespread adoption. This insight motivated Carl Sagan and others to discuss and investigate the potential consequences of nuclear weapons, which contributed to a more comprehensive understanding of their risks (Sagan, 1983). The seemingly diverse topics treated in this dissertation are part of a similar project, namely to participate in a research programme that aims to increase the understanding of the risks and possibilities of human enhancement technology. In particular, this dissertation seeks to explore the ethical and social ramifications associated with this technology from a non-essentialist perspective. This perspective has been inspired to a significant degree by thoughts and ideas from transhumanism, an intellectual tradition that has become increasingly prominent in both academic and lay discourse on applied ethics.

This introduction aims to give the reader an intellectual context for the essays that follow. The first part includes a brief review of the history of the transhumanist project. As transhumanism is advocated by a large and disparate number of organisations and people, this review is not exhaustive. It introduces the ideological context that preceded and formed part of the
movement. The transhumanist project, as understood here, connects two seemingly disparate viewpoints which are described below.

First, transhumanism involves a systematic and radical rejection of normative essentialism and the notion that human nature needs to be preserved. This essentialist notion still pervades philosophical and lay discussions in the realm of applied ethics. Secondly, transhumanism explores and seeks to revise public perceptions of technological risk. By setting aside normative essentialism, transhumanists argue that the main risk of technology is not erosion of human dignity but a threat to our very existence.

The second part of the introduction explores these two transhumanist viewpoints. It also introduces some of the topics to be covered in the later essays, and connects these with the transhumanist project outlined here. The third part of the introduction gives a short summary of each essay. The final part gives a summary of the essays in Swedish.

1. The History of Transhumanism

Transhumanism is both a set of normative claims and a movement seeking to achieve its ideals. At the core of the transhumanist project is the rejection of human exceptionalism, the idea that there is a specifically human essence and that humans represent the pinnacle of creation. Transhumanists argue that emerging technologies may fundamentally transform the human condition and that this transformation does not need to be morally wrong although it risks becoming so. As an intellectual movement, transhumanism has its roots in secular humanism and futurism. Transhumanists argue that although secular humanism has undermined many essentialist beliefs in the natural science discourse and promoted a science-based world view, it has failed to extend this analysis to the human condition, which retains its exalted status in humanist ethics. Transhumanist ideas were first proposed in 1923 by the geneticist J.B.S Haldane who argued in his essay *Daedalus: Science and the Future* (Haldane, 1923) that the human condition could be greatly improved by the application of science to human biology. The term transhumanism was first coined by the biologist Julian Huxley in 1957 in an article where he argued that ‘the human species can, if it wishes, transcend itself’ (Huxley, 1957). However, many of the most prominent ideas behind transhumanism came into disrepute after World War II, most notably eugenics. Since then computer scientists and mathematicians have argued that the human condition could be transcended

---

1 By risk, I refer here to the probability multiplied with the harm of a possible event summed over the range of possible events.
by integrating with or by being assimilated into intelligent machines (Good, 1965). In particular, the idea of an intelligent machine with the power to recursively improve its own intelligence has been seen as leading to a technological singularity. It has been suggested that this event, also referred to as an intelligence explosion, would violently usher in a new transhuman era where normal humans would be left far behind. Another important influence on the transhumanist movement was the psychedelic subculture, which contributed a distinctly contrarian perspective and a libertarian outlook.

The first systematic philosophical enquiries into transhumanism and human enhancement were formulated in the last decade of the twentieth century. The British philosopher Max More is credited with having presented the first systematic, philosophical defence of human enhancement and transhumanism in the early 1990s (Hughes, 2004). He proposed that:

Transhumanism is a class of philosophies that seek to guide us towards a posthuman condition. Transhumanism shares many elements of humanism, including a respect for reason and science, a commitment to progress, and a valuing of human (or transhuman) existence in this life (More and Vita-More, 2013).

However, the work of Eric Drexler, which discussed nanotechnology and advocated cryonics, was at least as influential in bringing together the disparate subcultures and organisations into a cohesive movement. His ideas coalesced around the Extropy Institute. These early defenders of human enhancement and transhumanism argued from the radically individualist position of morphological freedom, stating that an individual’s desire to alter their body or mind is a private concern, regardless of risk (More and Vita-More, 2013). These radical pro-enhancement ideas are often combined with the proactionary principle which states that:

Our freedom to innovate technologically is valuable to humanity. The burden of proof therefore belongs to those who propose restrictive measures (More and Vita-More, 2013).

These views remained in the fringe of academia for obvious reasons. The proposed ideas were utopian and often expressed in an inaccessible social dialect full of jargon inspired by science fiction. The early transhumanists were quirky even by the standards of academic philosophy; the culture of the transhumanist movement was insular and socially homogenous. By the beginning of the twenty-first century, however, the status of transhumanism had been transformed by the creation of the World Transhumanist Association (WTA). In contrast to the Extropy Institute, the WTA actively sought public
recognition and social acceptance of transhumanist ideas. A few years later, the Future of Humanity Institute was established at the University of Oxford, led by one of the WTA’s founders, Nick Bostrom (Bostrom, 2005). These institutions have successfully changed the profile of human enhancement and transhumanism from being on the fringe of academia to becoming a movement taken seriously in prominent, bioethical, peer-reviewed journals. The subject is now included in many courses in bioethics at well-renowned universities. The public image of transhumanism has also been improved by the notable shift in the political and ideological centre of gravity in the WTA. From being radical libertarians, the members of the WTA began emphasising social concerns such as the equal distribution of enhancement and moved to the centre-left under its executive James Hughes.

Human enhancement and transhumanism entered broader public debate after rigorous criticism by leading intellectuals including Francis Fukuyama, Leon Kass, Michael Sandel, Bill McKibben, George Annas and Jürgen Habermas. These critics were labelled bioconservatives by proponents of human enhancement. While their ideological views diverged on most other topics, they all subscribed to the bioconservative thesis:

Even 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 (Douglas, 2008).

These intellectuals have proposed a general restriction of human enhancement with remarkably similar arguments, considering their disagreement on other topics. Michael Sandel argues that human enhancement epitomises the modern desire for mastery and control, and that human enhancement is contrary to traditional virtues such as humility and openness to the unbidden. In his view, the hyperautonomy that human enhancement creates threatens to undermine all value (Sandel, 2007). These claims have been contested by Frances Kamm, and others (Kamm, 2006). She argues that Sandel’s argument rests on a significant distinction between human enhancement and other applications of modern technology to treat disease. Senile osteoporosis (bone decalcification) and unwanted pregnancies are not considered to be diseases, yet should the treatment of these problems be regarded as human enhancement? And if so, can we really conclude that these treatments are unacceptable? Jürgen Habermas argues that moral autonomy depends on not being subject to another person’s specifications. This autonomy is threatened by the genetic enhancement of embryos (Habermas, 2003). In his view, a person whose genetic predispositions are the result of a random process is more autonomous than one whose genetic makeup is partly the result of deliberate choices. Another potential consequence of human enhancement is the further
entrenching of inequalities. Sandel and McKibben argue that if cognitive enhancement was possible it would create a genetic divide between the enhanced and the natural humans (Sandel, 2007; McKibben, 2004). Huxley’s depiction of a dystopian society in *Brave New World* in which one’s social class depends entirely on genetic modifications is often cited by critics in support of this view. Francis Fukuyama suggests that human enhancement threatens to undermine the idea that all humans have equal moral worth by challenging the idea of a specifically human essence (Fukuyama, 2002). Leon Kass objects to human enhancement on the grounds that it is inconsistent with human dignity (Kass, 2003). These essentialist objections to the transhumanist project will be discussed in greater detail in the next section of this introduction.

In the last few years, less radical proponents of human enhancement have made the debate more subtle. Philosopher Nicholas Agar argues that although some instances of human enhancement are acceptable and even desirable, radical enhancement risks undermining the values that we hold dear (Agar, 2010). Julian Savulescu and Ingmar Persson have warned about the dangers of cognitive enhancement and have suggested enhancements to our moral dispositions to reduce some of the risks entailed by technological development (Persson and Savulescu, 2008). Allen Buchanan is another philosopher who has adopted a moderate and cautious approach to human enhancement. He argues that we have to take a more fine-grained approach because there is no clear answer on how to approach human enhancement. Different modes of enhancement in different contexts have different risk/benefit profiles (Buchanan, 2012). However, despite the more nuanced views on human enhancement in the philosophical community, transhumanism remains a controversial project. Its association with individuals and movements on the fringe has prevented the spread of its radical and important message. Yet, as will be argued in the next section, the ideas behind transhumanism are of great relevance to public discourse.

2. Transhumanism and Normative Essentialism

Transhumanists comprise a motley group of disparate people and movements. However, at the core of the project lies an idea that is relevant to a wide range of discussions in applied ethics, namely the rejection of normative essentialism, a view sometimes attributed to Aristotle. Even in the twenty-first century, we are still struggling with his legacy. Arguably among the most influential thinkers, Aristotle’s thoughts have shaped systematic thinking in the world like few others. The sheer ubiquity of his thoughts makes it difficult to scrutinise or reject them as they seamlessly blend into our common sense intuitions.
Normative essentialism retains its appeal particularly when applied to bioethics.²

Normative essentialism with regards to a particular kind, such as a species, is a combination of two claims: (1) that the members of this kind are members by virtue of some intrinsic property (or set of properties); and (2) that this membership has normative significance. In other words, it combines a claim about the essential properties of a kind with a claim that there are some normative facts that pertain to instances of a kind by virtue of being of that kind. In this context, essential property means a property that is shared by all proper members/instances of a kind and by these members and instances alone, i.e. necessary and sufficient properties.

Allow me to exemplify the first claim. Gold is a paradigmatic example of a kind whose instances share essential properties. It is believed that the number and composition of elementary particles in a gold molecule are unique to all proper instances of such molecules. Furthermore, these features of gold explain other features of this substance, such as its density and electrical conductivity. Thus, the atomic number of gold seems to be an essential property of its instances. If we accept these claims, we also accept essentialism with regards to gold.

Normative essentialism simply adds the claim that being a member of a certain kind is normatively significant, i.e. that membership of some kinds counts as a moral reason. For example, Carl Linnaeus devised an essentialist taxonomy of the different human races. In his view, each race corresponded to each of the four elements; membership of one race implied certain duties and obligations by virtue of this membership (Brace, 2005). While racial essentialism is no longer widely accepted, normative essentialism is still prevalent in other domains.

Normative essentialism can be challenged on two levels. First, the essentialist assumption can be undermined by showing that the alleged essential features of a particular kind are shared by non-members of that kind or by showing that a presumed member of that kind lacks one of the alleged essential properties. For example, if giving birth to live offspring is claimed to be an essential property of mammals, one could undermine this claim by pointing out that the platypus is an egg-laying mammal.

² It should be noted that although normative essentialism is often ascribed to and associated with Aristotle, other interpretations of his ethics are certainly possible. See, for example Nussbaum (1998).
Secondly, since normative essentialism seeks to derive a normative claim from naturalist premises, it can also be confronted with the naturalist fallacy. For example, Eric Katz argues that natural ecosystems have a value that restored ecosystems lack (Katz, 1996). He believes that this depends on the distinction between natural objects and artefacts that are objects created by humans with a purpose in mind. Here, a normative claim is derived from a set of naturalistic features (not being man-made). This claim is discussed in greater detail in Essay V.

Numerous critics have rejected normative essentialism with regards to race, gender and sex. Transhumanism, however, offers a more general rejection of normative essentialism. In some domains, such as its claims about the essential nature of humans, the transhumanist rejection of essentialism is not controversial. Biologists and philosophers of biology abandoned classical species essentialism long ago (Hull, 1965). Attempts to classify organisms in discrete taxonomic units by virtue of their traits, also known as phenetics, has been largely superseded by the cladistics approach, which organises species in genealogical relationships.

Modern biological essentialism, which defines species in terms of reproductive compatibility, has also met with significant challenges. Reproductive compatibility means that population A and B are of the same species if and only if A and B can create fertile offspring. The problem with this essentialist view is that whereas species identity is supposed to be transitive, reproductive compatibility is not. This is best illustrated by so called ring-species, where geographically adjacent populations organised in a ring around the planet can interbreed with other populations in their vicinity, but not with populations that are farther removed. Unfortunately, misguided notions of a human essence or of humans as the most evolved animal still pervade contemporary discussions on bioethics. Normative essentialism also has a strong influence on our intuitions regarding the appropriateness of human enhancement technologies. However, no such essence has been identified, nor is it likely to be.

Francis Fukuyama argues that political liberalism and its egalitarian ethos presupposes the idea of human essentialism and an idea of humans as exceptional. He writes:

Underlying this idea of the equality of rights is the belief that we all possess a human essence that dwarfs manifest differences in skin color, beauty, and even intelligence. This essence, and the view that individuals therefore have inherent value, is at the heart of political
liberalism. But modifying that essence is the core of the transhumanist project (Fukuyama, 2004).

Without human exceptionalism, would we continue to pay heed to the needs of the dispossessed? Racial exceptionalism was a prevalent political ideology in the days of colonial Europe. Yet this did not prevent the ruling elite from abusing their white subjects. Nor has human essentialism prevented widespread destitution. If Fukuyama is correct, abandoning racial essentialism should have led to increased exploitation of poor whites. Yet, while some whites did lose some of their racial privileges after desegregation in the American south, their rights and legal status were barely challenged.

Human essentialism is sometimes expressed as a concern for human dignity. For example, in the debate on human enhancement, detractors of transhumanism, most notably Leon Kass, have claimed that human dignity is threatened by human enhancement (Kass, 2004). For example, he writes:

Human nature itself lies on the operating table, ready for alteration, for eugenic and psychic "enhancement," for wholesale re-design. … Our immediate ancestors, taking up the challenge of their time, rose to the occasion and rescued the human future from the cruel dehumanizations of Nazi and Soviet tyranny. It is our more difficult task to find ways to preserve it from the soft dehumanizations of well-meaning but hubristic biotechnical "re-creationism"—and to do it without undermining biomedical science or rejecting its genuine contributions to human welfare (Kass, 2004, pp. 2–3).

But what does dignity mean? Since Kass believes that human dignity would be violated by, for example, enhancing cognition, he cannot share Kant’s idea of dignity, which is associated with our ability to reason. Rather, Kass seems to hold the belief that dignity is a feature possessed by all members of the human kind and only by members of the human kind, and that this feature confers moral obligations on those who possess it. This feature is somehow harmed or undermined by human enhancement. In this view, some things are considered immoral simply because they violate this essential feature. For example, the essential property of a musical instrument is its ability to produce harmonious sounds and this feature makes any use of such an instrument that does not involve producing harmonious sounds a wrongful act. We may think that it would be wrong to use a violin to dig a hole in the dirt, because such an act would be contrary to the essential nature of the object. Kass seems to hold similar beliefs about humans. Since there are essential features that all humans share, acts that run contrary to these features are wrong. When such acts are wrong, they can be characterised as violating human dignity. These ideas could
be challenged as we become less squeamish about biotechnology and, in particular, if human enhancement technology allows for increased diversity among humans, a topic discussed in Essay V.

Ingmar Persson and Julian Savulescu argue that, rather than trying to preserve our human essence, we should strive towards that which we value, being humane (Persson and Savulescu, 2010). If moral bioenhancement makes us choose between being more human and being more humane, we ought to choose the latter. Moral bioenhancement is even more controversial than cognitive enhancement. John Harris has for example argued that moral bioenhancement is inconsistent with free will (Harris, 2011). In Essay III, I argue that these concerns are primarily directed against one form of moral bioenhancement, behavioural enhancement. Other forms of bioenhancement are not vulnerable to this objection.

Normative essentialism is not restricted to the classification of individuals but forms part of our mainstream beliefs in medical ethics. Here, the concepts of health and disease are central in the prevailing discourse on what medical practice is, or should be. It is widely believed that these concepts refer to natural kinds, analogous to the chemical elements, and that whatever condition is a true disease is a bad thing for anyone having this condition. Many philosophers and medical practitioners have sought to formulate an intuitive, coherent and practical definition of this concept. Even critics of the medical establishment have accepted this implicit assumption, criticising the medical treatments of conditions that are not real diseases (Szasz, 2010). Considerable efforts have been made to find consistent, plausible and practical definitions of disease and health that appeals to some intrinsic feature (Boorse, 2011). Despite repeated failures, the belief in the underlying assumption that normative essentialism about disease must be true is still so strong that the enthusiasm has not yet faltered. The transhumanist outlook serves to reject this assumption. As I argue in Essay VI, ‘Disease Prioritarianism: A Flawed Principle’, normative essentialism about disease is deeply misguided.

By discarding various instances of normative essentialism, transhumanists have been in a position to investigate technological risk and uncertainties from an unconventional perspective. Transhumanism is often conflated with general optimism about technology. To be sure, such optimism is not difficult to find among transhumanists. Yet transhumanism should rather be associated with the belief in the transformative power of technology, both as an enabler and as a great peril. However, if transhumanists dismiss the essentialist concerns of Kass and Fukuyama, what perils remain? To properly address this question it is worth remembering the close association between the transhumanist and humanist traditions. In particular, humanist political activism against perceived
threats to the continued existence of mankind, as so powerfully expressed by Bertrand Russell and Carl Sagan, has inspired many transhumanists to understand, explore and inform the public about existential risks. This task is complicated by a number of popular narratives and perceptions that seem to be largely influenced by the concerns formulated by Kass rather than those that Sagan articulated.

During the writing of these essays, and in particular Essay II, which describes convergence seminars, I had the privilege to take part in a large number of discussions with people outside the academic world. These discussions were centred on human enhancement, technological risk and the future of humanity. A convergence seminar (described in more detail in Essay II) is a form of scenario-based group discussion technique. One of the ideas underlying this technique is to draw out moral intuitions and allow the participants to critically reflect on them. Three broad reactions to human enhancement and technological risk were prominent in the intuitions that affected the judgment of those taking part in the discussions.

First, disgust was clearly communicated by some sceptical individuals. Secondly there was frequent use of the narrative of science fiction dystopia, often used by bioconservatives as an argument against human enhancement and transhumanism. The trope of societal decline, decadence and degeneration has always been part of the conservative narrative, so it should not be a surprise that it appears as part of a general aversion to human enhancement technologies. Thirdly, the myth of Icarus, where human hubris, but also ingenuity, is punished by fate was also notable in the thoughts formulated in these discussions. The appeal to disgust as an emotion that ought to guide moral thinking and jurisprudence is not new or limited to popular opinion. Lord Patrick Devlin wrote:

I do not think one can ignore disgust if it is deeply felt and not manufactured. Its presence is a good indication that the bounds of toleration are being reached. Not everything is to be tolerated (Devlin, 1965).

This sentiment has been echoed in modern times by conservative scholars such as Justice Antonin Scalia (Strong, 1997). Although disgust, just as any other emotion, may be a valid starting point in an ethical discussion, it ought not to be the final word. However, this is often the case. In my experience, no other emotional reaction is as powerful as disgust when it comes to the distortion of rational deliberation. Individuals who feel disgust tend to refuse to engage critically with their intuitions or to rationalise *ad absurdum* their initial reactions. Jonathan Haidt describes this effect in some detail and how it affects
our moral deliberations and risk perceptions (Haidt, 2001). According to Haidt, much of our moral reasoning is a *post hoc* construction that justifies our initial gut reaction. Many applications of human enhancement technology have in my experience roused such reactions and such reasoning. However, when the medical and therapeutic uses of this technology were mentioned, fewer adverse emotional reactions were evoked. Although some bioethicists argue that there is wisdom in repugnance (Kass, 2004), the arguments against this view are compelling. Martha Nussbaum, for example, has noted that disgust has been used as a justification for persecution (Nussbaum, 2006). Palaeontologist Stephen Jay Gould has also remarked that reactions of disgust undermine critical and rational reflections, and are thus contrary to wisdom (Gould, 1996). The philosopher John Harris has also rejected the notion that there is wisdom in repugnance by arguing that ‘there is no necessary connection between phenomena, attitudes, or actions that make us uneasy or even those that disgust us, and those phenomena, attitudes, and actions that there are good reasons for judging unethical’ (Harris, 1998 p.37).

A second concern relies on the narrative aspects of science fiction films and novels. While thought experiments and scenarios are widely used both by philosophers and the public to better understand risks, fiction is different, as it has an inbuilt narrative bias. A scenario has to be credible and plausible. In contrast, a novel must tell a good story, even if the unfolding of events in the story is very implausible. Dramatic logic is not logic. Storytellers are routinely warned that an event is not necessarily dramatically credible just because it happened in real life. Conversely, we ought to reject the notion that an event is likely just because it has been vividly portrayed in a fictional work. The prevalence of fiction in our thinking about the ethics of technology may lead us to both underestimate and overestimate risks (Kahneman, 2003). In the context of human enhancement and technological risk, the narrative in dystopian fiction is particularly apt at evoking images of totalitarian governments using this technology to control its citizens or of the mindless drones in *Star Trek’s* Borg community. Less attention is directed at the risks of unsafe medical practices, rejection of brain implants and other long-term adverse health effects. As risks in movies and books are often related to specific (evil) agents, not being aware that one’s risk assessment is anchored by such narratives may lead us to overlook risks created by mistakes, natural disasters or by benevolent agents with the best intentions (Lassiter et al., 2002). Thus fictional representations in moral and ethical discourse combine availability bias, agency bias and narrative bias in a powerful mix that distorts clear thinking on these issues. When the risks of technological innovations discussed in public discourse resemble those used in science fiction narratives, this should be noted. Not all science fiction scenarios are dystopian, however,
and the allure of fantastic future projections may be just as misleading in making us believe that we can really predict the future with great accuracy.

The third narrative thread prevalent in the discussion groups was based on the myth of Icarus. This story and its moral, about technological hubris and the inevitable punishment of the gods, resounds profoundly in our heuristic toolbox. The idea that technology is something inherently dangerous has since been replicated in countless narratives, from Mary Shelly’s *Frankenstein* to Aldous Huxley’s *Brave New World* and modern works such as the *Terminator* series. While it is of course reasonable to be cautious when making important decisions about technology, no story about the opportunity costs of delayed technological development seems available to us. Although we tend to appreciate (some) well-known technologies, future technologies are mostly portrayed as menacing, dehumanising and alien. However, there may be more risk in failing to implement some technologies. It seems to be much more difficult to imagine and tell a compelling story about this risk.

I believe that philosophers of risk and applied ethics may engage fruitfully in this discussion if they seek to avoid these biases. While technological risk is formidable, the current state of affairs, with widespread destitution and environmental stress, cries out for greater acceptance of existing technology to address some of these problems. Meanwhile, the risk of an existential disaster, perhaps too large and obscure to focus our minds on, needs to be explored more aggressively.

Existential risks are likely anthropogenic. While non-anthropogenic risks are certainly possible, events of the magnitude required to pose such risks are extremely unlikely. For example, while there are a number of large asteroids in the solar system that could cause significant damage to a city, no asteroid capable of killing a significant fraction of the current population on this planet is likely to collide with earth within a century. Similarly, natural pandemics are either very contagious or very virulent. Even extremely virulent microbes, such as *Yersinia pestis*, killed only a third of the population in Europe during the Black Death. Anthropogenic risks and in particular risks caused by existing and possible future technology are much more likely to pose an existential threat. Nuclear weapons, bio-engineered microbes and unforeseen consequences of geoengineering are some of the risks that figure prominently in these discussions. Yet, technological change is inherently difficult to predict and the possibility space of risky technologies may be dominated by unknown unknowns. This adds further difficulties to assessing and communicating risk. In Essay VII, such a strategy is outlined together with a practical strategy to reduce existential risk.
information that could theoretically be extracted through BMI devices is of concern. Here there is work to be done, both for philosophers and lawmakers; a plausible definition of privacy and a comprehensive and transparent regulatory framework must be agreed. Threats to autonomy may come in the future if less risky technology for DBS is developed. A potential example of this could be advanced transcranial direct current stimulation that modulates brain activity through the application of short bursts of current from an external device (Cohen Kadosh et al., 2012). Since brain stimulation can powerfully alter emotional states and change our desires and dispositions, such technology clearly has an enormous potential for abuse. However, it could also be used as a tool for moral enhancement and to enhance autonomy in some cases.

ii. European Public Deliberation on Brain Machine Interface (BMI) Technology: Five Convergence Seminars

The second essay, co-written with Professor Sven-Ove Hansson, was published in Science and Engineering Ethics. This essay describes a novel procedure to engage lay people in deliberations on risk, ethics and technology. These deliberations, referred to as convergence seminars are a form of scenario-based group discussion which is founded on the idea of hypothetical retrospection formulated by Sven-Ove Hansson (Hansson, 2007). Convergence seminars enable the systematic application of a pattern of argument that is prevalent in non-philosophical discussions. One of the most common types of argument about future possibilities consists of referring to how, in the future, one might come to evaluate the actions one takes now. These arguments are often stated in terms of predicted regret: ‘Do not do that. You may come to regret it’. Just as we can improve our decisions by considering them from the perspectives of other concerned individuals, we can also improve them by considering them from alternative future perspectives. We can hypothetically see them as we will see them retrospectively in the future. We consider this methodology to be particularly useful in areas where considerable uncertainty exists and where standard quantitative methods for risk assessment are less suitable such as the future development of BMI technology. A set of concrete scenarios were developed for the discussions. These scenarios all led to some future point in time, but each scenario led to a different branch of future development. The focus was on a decision in the present or the near future that the participants were asked to evaluate from the viewpoint of their scenario. The different scenarios were also constructed so that they represented branches in which different alternate decisions gave rise to problems that made them difficult to defend in hypothetical retrospection. To make the procedure easy to apply in a few hours, only three scenarios were
used. This was the first use of convergence seminars on this topic. The method functioned well, both logistically and more importantly, by giving rise to the type of discussions that we aimed for, namely discussions on how today’s decisions might be influenced by different possible future developments. As expected, the methodology was well suited for discussions on the future of BMI with its many uncertainties. The responses provided by the participants in discussions and questionnaires indicated that their advice regarding what decisions should be made about BMI development was influenced both by different possible future developments and by the points of view of their co-participants. It is also worth noting that most participants agreed that BMI technology was beneficial when and if used for medical purposes, whereas some were sceptical about the use of this technology for commercial and military purposes.

iii. What to Enhance: Behaviour, Emotion or Disposition?

The third essay, published in Neuroethics, explores the idea of moral bioenhancement, a controversial example of human enhancement that has been widely discussed in recent years. Although opponents of human enhancement emphasise the risks associated with tampering with the complex nature of our evolved psychology, these risks ought to be balanced by the potential rewards that any such tampering could achieve. While the rewards from increasing our physical strength or our perceptual apparatus are likely to be limited, the rewards from improving our moral sensibilities should be clear to anyone in favour of moral education or any other kind of traditional moral improvement.

Here I distinguish between three kinds of moral bioenhancement. Behavioural enhancement, known from science-fiction films such as A Clockwork Orange, consists in restricting or promoting certain behaviour. This could be carried out, for example, by drug-induced hypersensitivity to alcohol or by some implant that could modulate behaviour by electrical stimulation. Emotional enhancement, as proposed by Thomas Douglas in his article Moral Enhancement, consists in promoting or restricting specific emotions (Douglas, 2008). Aggression and xenophobia are likely candidates on the list of problematic emotions that could be reduced. The third possibility is to enhance dispositions to feel in certain ways in certain contexts. The most plausible candidates, I argue, are empathy and inequity aversion. First, empathy is closely linked to pro-social behaviour on a conceptual level. Although a definitive empirical demonstration of the causal role of empathy and pro-social behaviour has yet to be carried out, there is consistent evidence that empathy motivates altruistic behaviour (Batson and Shaw, 1991; Eisenberg, 2000). For
example, empathetic people avoid harming others, are more willing to cooperate with strangers, and are more willing to help others. Secondly, as Michael Slote argues in his book *Moral Sentimentalism*, empathetic people are not less rational, contrary to cultural stereotypes (Slote, 2010, pp. 141–159). Thirdly, we seem to have a relatively good idea of how empathy works when compared to what we know about xenophobia, sadism or other complex emotions and desires that it might be desirable to alter. Some selective serotonin reuptake inhibitors (SSRIs) seem to work reasonably well in altering emotional responses (Crockett et al., 2010).

Our sense of fairness, or inequity aversion, is a preference for fairness and resistance to incidental inequalities (Fehr and Schmidt, 1999). Although there are substantial cultural variations in its particulars, a sense of fairness is probably a human universal (Brosnan and de Waal, 2003). Inequity aversion seems to be associated with, but independent from, our ability to empathise (Bird et al., 2010). In particular, people with Asperger’s syndrome and high-functioning autism syndrome appear to be highly motivated to be moral (Kennett, 2002; McGreer, 2008). This implies that, although empathy plays a central role in moral thinking, it is not the only important disposition. While a sense of fairness is present in (some) autistic people (although this is still poorly understood), it seems that some neurological correlates have been identified (Güroğlu et al., 2011).

I argue that some arguments made against moral bioenhancement are only relevant against behavioural enhancement, and that general arguments against moral bioenhancement are less powerful when directed against the second and third kind of moral enhancement. I also argue that we ought to adopt a combined strategy, where dispositional enhancement is supplemented with some emotional enhancement.

iv. Sensory Enhancement

The fourth essay, published in the *Handbook of Neuroethics*, considers the possibility and potential desirability of sensory enhancement. Sensory enhancement is in this context an instance of human enhancement. Thus, a correction of deficient hearing or sight is not covered here. At the moment, glasses and cochlea implants are not sensory enhancement technologies. Should telescopes and microscopes be defined as sensory enhancement technologies? Here the distinction is more tenuous and based on what we perceive to be part of a person’s abilities. For something to count as an enhancement this must add or extend a person’s functioning or ability to do something. Clearly we do not believe that a person who owns a telescope has
and Martha Nussbaum, and sketches out the position that we may be deprived in comparison to people in a possible future, even if we do not regret being so.

v. Should Extinction be Forever? (under review)

This essay explores a problem which is related to our moral obligations towards non-human animal species. In its *Earth Charter*, the United Nations Educational, Scientific and Cultural Organization (UNESCO) asserts that ‘every form of life has value regardless of its worth to human beings’. This sentiment seems to be widely shared, notably in the field of environmental ethics. One widely-used method of conceptualising the moral wrongness of anthropogenic animal extinction is to formulate it as a negative *prima facie* duty. The notion that we can have duties towards species or other abstract entities is somewhat at odds with traditional deontology, which has tended to the view that a duty requires at least one rights holder. In environmental ethics, however, it is often argued that duties are to be understood as general principles of proper conduct, in that the preservation of species can be seen as the practical expression of respect for those species, rather than as a traditional rights-based duty.

It is often also claimed that humanity has a collective responsibility to non-human animal species. In contemporary political philosophy, collective responsibility is a contested notion, set against the idea of moral individualism, which holds that people only have obligations as individuals. Moral communitarianism, by contrast, holds that we are not unencumbered selves, but participate in a community, which serves to define not only who we are but also our responsibilities. This explains why we are entitled to feel pride for the achievements of past generations, but also why we must bear the burden of their crimes (Sandel, 2009). Although moral communitarianism remains controversial in moral and political philosophy, it is a mainstream view in the environmentalist community and in environmental ethics.

Within a deontological framework, it is argued that when humanity causes the extinction of an animal species, this is a moral transgression, entailing a residual obligation. Such an obligation implies a positive duty to mitigate any harm caused by our moral failure. To be able to apologise or make reparation requires someone to whom apologies or reparations can be made. For any species that has been exterminated, no such compensation is possible. A further question, then, concerns whether the creation of a viable population of organisms, genetically identical to an extinct species, would constitute a genuine re-creation of that species, or whether it would merely be creating a copy. This echoes an issue raised in current discussions of personal identity as to whether creating a qualitatively identical replica of a deceased person would
recreate that same person or only a copy. The way in which we answer this question is important for how we assign value to the continued existence of a particular species, just as the question of personal identity is important for how we assign value to individual survival (Parfit, 1984).

Until recently, little could be done to redress the harm done to extinct species; now, or in the near future, it will be possible to recreate them. This possibility brings residual obligations into play. If species are to be understood as entities that are spatiotemporally restricted, rather than as atemporal classes, not all of the available technological means for species re-creation would constitute a genuine reconstruction. This argument has some important implications. First, if genetic material is placed in a preserving medium (liquid nitrogen, for example), this may reduce the risk of extinction for at least some species with few living members. This safe-fail (rather than failsafe) conservation strategy seems worth pursuing more aggressively, particularly in light of repeated failures to protect endangered species. Secondly, the mere existence of biological material from a species means that the very same species has the potential to exist. Thirdly, the moral transgression of extermination can be at least partially mitigated by reanimating the species. This idea becomes starkly controversial when we consider the possibility that scientists’ plans to recreate the mammoth might be used to restore life to the Neanderthals, and by very similar means.³ There is some less than certain evidence that the death and eventual assimilation of the last Neanderthals was caused by Homo sapiens. However, the evidence is less equivocal concerning mankind’s involvement in eliminating the last members of other species (Lowe et al., 2012). If we accept the notion of collective responsibility, then we have residual obligations towards dodos and Pinta Island tortoises among other species whose numbers have been fatally reduced by human activity.

vi. Disease Prioritarianism: A Flawed Principle (under review)

The healthcare system accounts for a large and increasing share of the welfare state’s commitment to its citizens. It is therefore pertinent to ask what ought to fall within its domain. Much of the debate on medicalisation, waged between psychiatrists and sociologists of medicine, (Illich, 1982) has explicitly assumed that being able to define the term disease is important to guide our healthcare priorities (Parens, 2013). This notion is also the standard approach to prioritisation in the healthcare system. This notion is referred to as disease prioritarianism in this essay, a principle that is often implicitly or explicitly employed in the realm of healthcare prioritisation. This principle states that the

³ Depending on how species is defined, Neanderthals are either a species or a subspecies.
healthcare system ought to prioritise the treatment of disease before the treatment of any other human problem. This essay argues that disease prioritarianism ought to be rejected. Instead, I argue that we should adopt “The problem-oriented heuristic” when making prioritisations in the healthcare system. According to this idea, we ought to focus on specific problems and whether or not it is possible and efficient to address them with medical means. This has radical implications for the extension of the healthcare system.

First, getting rid of the binary disease/not disease dichotomy implicit in disease prioritarianism would improve the ability of the healthcare system to address chronic diseases and disabilities, which often defy easy classification. This would be of major importance in assisting those who suffer from mental illness, as recent research suggests that such conditions lie along a spectrum and are seldom possible to organise in discrete categories (Adam, 2013). For example, insomnia increases the risk of illness and accidents. Yet although this condition may be caused by a disease, such as depression, this is not always easy to determine as insomnia can also cause depression (Fleishman, 2012).

Secondly, a problem-oriented approach could empower medical practitioners to address social problems without the need to pathologise these conditions. New research shows that medical treatment shows great promise in preventing crime (Lichtenstein et al., 2012) and improving educational attainment (Scheffler et al., 2009). If this holds true, a rethinking of medical practice could yield massive societal benefits.

Thirdly, a problem-oriented approach clearly states that what we choose to treat is a normative consideration. Under this assumption, we can engage in a discussion on demedicalisation without distorting preconceptions. For example, it could also help decision makers and the general public distinguish between normative considerations, such as whether drug addicts should be held responsible for the effects of their substance abuse and empirical considerations, such as whether harm-reduction policies create a moral hazard and increase the number of drug addicts.

Fourthly, this pragmatic and decompartmentalising approach should allow us to reconsider the term efficiency. By thinking of medical practice as a tool that can be applied wherever it is potentially effective, the boundaries of the healthcare system could be dramatically redrawn. If we adopt this analysis, it may prove significant to the extent that sectors that are today non-medical could become medicalised. Other considerations become salient when focusing on specific problems and whether or not these could be successfully addressed by the healthcare system. We should systematically investigate which
societal problems could successfully and efficiently be addressed by medical means. We should also consider the costs for the public sector as a whole. For any plausible normative theory about what the healthcare system should be doing, interventions that have negative costs can plausibly be offered to the public at taxpayer expense.

vii. Existential Risks: Exploring a Robust Risk Reduction Strategy

There is a significant risk that mankind will not survive this century. An existential catastrophe refers to a category of possible events that permanently and drastically reduce the ability of earth-originating intelligence to create or sustain value. The seventh essay of this dissertation, published in *Science and Engineering Ethics*, discusses this problem. The common sense assessment of the negative value of such an event is to equate the extermination of mankind with the disvalue of a single individual dying prematurely multiplied by the number of humans that exist at the moment of the catastrophe. According to Derek Parfit and others, this notion vastly underestimates the disvalue of such an outcome since it neglects the contributive value of future generations (Parfit, 1984). Population ethics is a novel field in moral philosophy; most positions with regards to the importance of people in the future are associated with absurd conclusions and paradoxes (Arrhenius, 2000). Yet, axiological actualism, or the view that only existing (actual) people matter, remains a minority view. Making some assumptions about the long-term prospects of earth-originating intelligent life, the opportunity cost of extinction could be even more staggering according to Nick Bostrom. He argues that, according to most prominent strands of consequentialism, the potential value lost to human extinction is equal to the aggregate value of all lives that could have existed in a non-extinction scenario (Bostrom, 2003).

A small but growing number of studies have aimed to understand, assess and reduce existential risks, or risks that threaten the continued existence of mankind. However, most attention has been focused on known and tangible risks. This essay proposes a heuristic for reducing the risk of stochastic, unexpected or black swan extinction events. Neither probabilistic risk analysis nor decision theory, based on a fixed model of possible outcomes, can properly deal with this kind of event. Poor theoretical understanding of the climate system may, for example, lead to a disastrous result from efforts to reduce greenhouse gases by geoengineering. Alternatively, a nuclear war could have unexpected consequences if volcanic activity was at that moment more active than usual. Human errors, such as failure to predict how consumers will use a certain product, are also potential sources of catastrophic failure. In general, the heuristics are characterised by sacrificing some performance in exchange for less vulnerability to failed assumptions. This essay argues that the
As mentioned before, transhumanism is neither optimistic nor pessimistic with regards to technology, but rather possibilistic. While technology may be able to achieve much more than we can imagine, this is both a great opportunity and a great peril. Mankind has given itself godlike powers to shape its future. These powers are however not wielded by wise men or saints, but by the crooked timber of humanity. It is our urgent task to confront this challenge.

3. Summary of the Essays

i. Brain Machine Interface and Human Enhancement – An Ethical Review

The first essay, previously published in Neuroethics, reviews existing and experimental applications of brain-machine interface (BMI) technology. This technology essentially consists of various ways to connect the nervous system to a machine via electrodes, thereby making the direct exchange of information between these two possible. BMI technology can extract information from the brain or spinal cord and direct prostheses, computers and accessories, making this technology very promising for people with disabilities such as paralysis. Although steering prostheses currently require advanced arrays of micro electrodes, cheaper and simpler BMI devices suffice to direct characters in computer games. For example, simple BMI devices based on electroencephalography (EEG) that can be attached directly to the scalp have been used to analyse reactions to marketing input. Thus, this technology has already entered the commercial realm. This breakthrough is important because it is non-invasive and relatively cheap in contrast with earlier more invasive BMI devices. Accordingly, it has a potential for commercial use that the highly invasive intracranial electrocorticography (ECoG) interfaces lacked.

BMI devices can also be used to feed information to the brain. This allows implants to provide hearing for deaf people or rudimentary visual orientation for the blind. Deep brain stimulation (DBS) is routinely used to reduce motor symptoms for patients with Parkinson’s disease and other neurodegenerative diseases as well as chronic pain and major depressive disorder. However, as long as this treatment relies on an invasive procedure, it is likely to be restricted to use in cases of very serious diseases. The future development of BMI is a moral concern. While this technology provides help for disabled and sick people, its commercialisation could potentially undermine both privacy and autonomy. Advertising agencies, employers and the government are all interested in effective ways to know how we feel, think and respond to stimuli. If it is important to preserve privacy, as is generally accepted, the kind of
the ability to see the Galilean moons, even if that person is in some more
general sense able to do so. Human enhancement technologies ought therefore
to be seen as artefacts that are to a certain degree integrated with our persona.
According to a narrow view, only things that are physically integrated, i.e.
attached or assimilated to the body ought to count. A clear example of this
would be a vaccine. In a more inclusive view, devices that form part of a
person's mental self-representation should also be included. In this view, if a
person perceives glasses, clothes, contact lenses and (perhaps) a smartphone to
be part of his or her body, these devices count as human enhancements. This
characterisation will vary in different contexts and across generations, but it
will suffice for the purposes of this discussion. Another distinction that ought
to be considered is between a sensory enhancement that allows us to perceive
a sensory input without giving us a new sensory quality and an enhancement
that would literally change how we see the world. For example, a
thermographic camera forms an image of infrared light using visible light. An
observer can therefore indirectly see infrared light. This sensory experience
probably differs from what it might be like to be able to perceive infrared light
directly, which of course may differ radically between different individuals and
species. However, an indirect experience of the world may be just as useful for
practical purposes as a direct one. We should therefore include technologies
that allow people to see indirectly in our definition of sensory enhancement.
Another relevant distinction is between enhancement of our sensory capacities
and the enhancement of our perception. Whereas the first changes what we
perceive, the second changes how we perceive. Examples of perceptual
enhancement include drugs that allegedly enhance the way we perceive music
and allow us to distinguish between subtle nuances. The difference here is that
whereas sensory enhancement is primarily informational, perceptual
enhancement is primarily phenomenological.

This essay argues that sensory enhancement could be implemented in two
main ways: via the application of digital technology or by genetic engineering
of the human body. The potential of augmented reality (AR) and brain-
computer interface (BCI) technology is also explored in the section on digital
enhancement. The section on genetic engineering is mainly concerned with the
potential of horizontal gene transfer (HGT). To illustrate how to anchor our
expectations of what may be possible, the essay suggests that existing sensory
modalities in animals belonging to the chordate group function as proof of
concept of what may be possible given the right technology. Three arguments
on the normative aspects of sensory enhancement are also presented in this
essay. The first considers the instrumental value of being able to perceive new
forms of artistic expression. The second concerns the idea of diversity and
whether sensory enhancement could increase human diversity. The third
argument departs from the capabilities approach, formulated by Amartya Sen
engineering safety approach could be applied to reduce the risk from black swan extinction events.

It also proposes a conceptual sketch of how such a strategy might be implemented using isolated, self-sufficient and continuously-manned underground refuges. Some characteristics of such refuges are also described, in particular the psychosocial aspects. Furthermore, it is argued that this implementation of engineering safety strategy would be effective and plausible and could reduce the risk of an extinction event in a wide range of possible (known and unknown) scenarios. While building such a doomsday shelter is less glamorous than colonising the moon, it may give us much greater risk reduction for the money invested. The conceptual sketch of the project in this paper should be further developed in an interdisciplinary research project, which could benefit from the extensive literature on isolated, self-containing habitats. Architecture, engineering, social psychology and decision theory would probably be needed to fully assess the costs and social and technological challenges. Considering the staggering opportunity cost of an existential catastrophe, this strategy ought to be explored more vigorously.

4. Summary in Swedish (Sammanfattning på svenska)

Teknologisk utveckling är till sin natur svår att förutsäga och teknologier som har en potential att revolutionera vår tillvaro måste diskuteras och granskas innan de har förändrat samhället på ett irreversibelt sätt. Den här avhandlingen diskutrar denna utveckling, dess risker och möjligheter.


Två trender inom teknikutvecklingen kan sägas påskynda denna utvecklingen inom detta område. Den ena är att avancerade datorer blir allt mindre, mer kraftfulla och billigare. Med hjälp av ren beräkningskraft kan mycket mer information från hjärnan processas på ett effektivt sätt. Det har gjort att icke-invasiva gränssnitt, alltså sådana som kan användas utan att utsätta användaren

Min andra artikel, ”European Public Deliberation on Brain Machine Interface Technology”, är resultatet av ett årlångt fältarbete där jag besökte fyra olika europeiska städer (inklusive Stockholm) och testade en metod för att engagera allmänheten i etiska diskussioner om hjärnmaskingrännssnitt och relaterade teknologier. Metoden som testades kallas för ”konvergensseminarier” och har utvecklats av professor Sven-Ove Hansson.

Konvergensseminariet är ett gruppsamtal som utgår från specifika scenarier om möjliga framtida utvecklingar. Idén bakom den här metoden är en systematisk tillämpning av ett sätt att argumentera i icke-filosofiska sammanhang. Ofta uttrycks ett resonemang i termer av hur en bedömare skulle se på ens handlingar nu från ett framtida perspektiv. ”Kan du nu ha skäl att tro att du kommer att ångra ditt val i framtiden?” är den centrala frågan i den här typen av reflektion.


Urvalet av deltagare kunde av resursskäl inte vara representativt, men vår ambition var att hämta deltagare från vitt skilda regioner och intressegrupper för att åtminstone försöka få till en viss spridning. De fem grupperna som
deltog var lärare från en liten by i södra Spanien, medlemmar i Parkinsonföreningen i Stockholm, teologistudenter i London, medicinsk personal från Warszawa och filosofistudenter vid Stockholms universitet.

Seminariedeltagarna delades upp i tre mindre grupper där varje grupp läste ett scenario. Därefter fick deltagarna diskutera sina scenarier i sin grupp. Efter en tid delades grupperna upp så att varje scenario fanns representerad i varje ny grupp. Här ombuds deltagarna att presentera sitt scenario och sina tankar kring det för de andra gruppendeltagarna. Slutförde hade vi en diskussion i stor grupp med alla deltagare. Diskussionerna som denna övning resulterade i höll, anser jag, en mycket hög nivå, och fokusrade på hur och i vilken utsträckning vi idag kan påverka den framtidiga samhällsutvecklingen. Även om deltagarernas synpunkter inte kan ses som representativa för den europeiska allmänheten, är det värt att notera att de allra flesta var ense om att hjärn-maskinläggningar var till godo i den mån tekniken användes för medicinska ändamål. Fler studier av den här typen behövs, dels för att engagera medborgare i dessa viktiga diskussioner, dels för att stärka dialogen mellan medborgare, lagstiftare och vetenskapssamhället.


Jag argumenterar för att en del av invändningarna som riktats mot moralisk förbättring visar att man missat en viktig distinktion mellan olika möjliga förbättringsstrategier. Jag argumenterar vidare för att moralisk förbättring kan utföras på följande tre olika sätt:

(1) Beteendeförbättring, där vissa beteenden blockeras eller framkallas, till exempel genom att skapa en stark känsla av obehag när individen ägnar sig åt beteendet. Exempel på detta skildras i filmen A Clockwork Orange där
huvudpersonen Alex utsätts för en behandling som betingar honom till att må illa när han kommer i kontakt med våldsamma känslor.

(2) Emotionell förbättring går ut på att reducera eller stärka kraften i vissa emotioner. Det kan handla om att minska aggressivitet eller xenofobi. Till skillnad från beteendeförbättring påverkar emotionell förbättring personligheten direkt, på gott och ont.

(3) Dispositionell förbättring går ut på att förbättra människors förmåga till empati och aversion mot ojämlika uttalande. Dessa är inte emotioner i sig, utan snarare en benägenhet att reagera på vissa situationer på särskilda sätt. Empati innebär, i det här sammanhanget (a) förmågan att uppleva en känsla, (b) som är lik en annan persons känsla, (c) som väcktes av att identifiera eller föreställa sig en annan persons känsla och (d) när vi vet att vår känsla orsakas av den andres känsla. En känsla för rättvisa är, liksom empati, en benägenhet att känna vissa emotioner i vissa specifika situationer. Det kan röra sig om vrede i en situation av upplevd orättvisa, eller tillfredsställelse när ett utfall uppfattas som rättvist.

Den främsta invändningen som riktats mot moralisk förbättring handlar om att en sådan förbättring innebär en inskränkning i personers frihet och möjlighet till självbestämmande. Den distinktion jag presenterar visar att denna invändning bara är tillämplig på den första typen av moralisk förbättring. Endast beteendeförbättring reducerar individens handlingsmöjligheter. Andra former av moralisk förbättring påverkar inte handlingsutrymmet, utan snarare individens personlighet och benägenhet att handla på vissa sätt.


Traditionella argument mot andra former av förbättring, att de skulle förvärra ekonomisk ojämlikhet eller att det skulle handla om ett nollsummespel, tycks inte gälla för moralisk förbättring. Det är inte individen själv som blir den största vinnaren av moralisk förbättring, utan de som den här individen interagerar med. Inte heller blir det svårare för mig att vara moralisk om du är mer moralisk. Vissa, främst kristna, dygdetiker bekymrar sig för att den rätta
moraliska karaktären måste byggas upp genom ansträngning och dedikation. Men det här resonemanget tycks implicera att de för vilka moraliskt beteende är lätt borde försämra sin empati för att därmed genom ansträngning kunna bli mer moraliska.

Den fjärde artikeln, ”Sensory Enhancement”, är en diskussion om en typ av förbättringar som knappt nämns i den akademiska debatten, förbättringar av våra sinnen. Artikeln redar ut begreppet mänsklig förbättring, som i detta sammanhang innebär att begränsningar hos den mänskliga kroppen övervinns med hjälp av teknologi som är integrerad med den mänskliga kroppen. Om jag till exempel äger ett teleskop, kan jag i en bemärkelse sägas ha förmågan att se Jupiters månar. Men det är inte i den här bemärkelsen som mänsklig förbättring avser. Men vad menas med ”integrerad med kroppen”? Jag föreslår två tolkningar: (1) en teknologisk produkt måste vara fysiskt inbäddad i vår vävnad, som ett chochlea-implantat eller ett vaccin, eller (2) en teknologisk produkt måste vara en del av vår mentala självrepresentation, som till exempel mina kontaktlinser eller kläder. Jag uppfattar att tolkning (1) är onödigt begränsad, då vi i vardagsligt tal uppfattar att vi kan göra de saker som förutsätter att vi har tillgång till kläder, glasögon och andra icke-inbäddade artefakter.

Med sinnesförbättringar menar jag både sådana som förstärker ett existerande sinne (t ex skarpare syn), sådana som innebär att ett helt nytt sinne tillkommer (förmågan att känna magnetfält) och sådana som förändrar hur ett existerande sinne används för att kunna göra något helt annorlunda (sonar). Jag föreslår i artikeln en heuristisk modell för att fundera kring vad som är rimliga förbättringar inom detta område. Enligt denna kan vi anta att de sensoriska modaliteter som återfinns inom djurriket, och i synnerhet de som återfinns bland ryggradsdjur, rimligen kan bli förbättringar hos människor i framtiden. Det faktum att naturen har producerat modaliteten blir därmed ett proof of concept.

Artikeln diskuterar tre tekniker som idag tycks vara de rimligaste för att åstadkomma detta. Sensorer, som till exempel värmeljuskameror, blir allt billigare och mindre. Om dessa kan integreras i kroppsburna datorer, som digitala glasögon eller kontaktlinser, så skulle de i ett avseende ge användaren infrasyn. En annan möjlighet beror på utvecklingen inom hjärnmaskinering. Om sensorer kunde kopplas direkt till dessa, skulle en användare ha möjlighet att uppleva världen direkt, utan att låta sinnesintrycken ”tolkas” av sina existerande sinnen. En tredje möjlighet är att använda sig av ”horisontell genöverföring”, det vill säga att överföra vissa gener från en icke-mänsklig organism till en människa. Denna teknik har framgångsrikt använts.
för att producera självlysande fiskar, men mycket återstår innan den kan användas för att göra den typen av genomgripande förändringar som till exempel mörkersyn skulle innebära.

Artikeln diskuterar även normativa överväganden som är av särskild relevans för sinnesförbättringar. Även om dessa knappast har samma potential att förändra samhället som kognitiv förbättring, finns det skäl att inte avfärda värdet i sinnesförbättringar.

I den etiska diskussionen kring värdet av att inte vara döv framhålls ibland att döva saknar förmåga att uppskatta vissa typer av värdefull konst. Om konst som förutsätter vissa, för oss främmande, sensoriska modaliteter görs i framtiden, skulle vi då inte i på ett motsvarande sätt vara ”döva” inför den konstformen? Ett annat skäl att tillåta vissa former av sinnesförbättringar är att de, till skillnad från andra förbättringsmöjligheter, sannolikt skulle öka diversiteten i mänskliga perspektiv på världen. Minskad diversitet är något som kritiker av transhumanismen, med rätta, oroar sig över. Ett sätt att öka denna diversitet vore att tillåta sinnesförbättring. Ett tredje skäl att omvärdera sinnesförbätträrring vore att utsträcka Amartya Sens och Martha Nussbaums välfärdsanalys till att omfatta icke ännu existerande sinnen. Enligt Sen och Nussbaum är välfärd inte bara en subjektiv erfarenhet, det är också att ha vissa möjligheter och förmågor för att kunna interagera med världen. Men varför bör vi begränsa oss till det som är ”normalt” idag?


Inom miljörörelsen är uppfattningen att mänskligheten som kollektiv kan anses ha ett gemensamt ansvar för våra handlingar. Den här uppfattningen, som ibland kallas för ”moralisk kommunitarianism”, är inte okontroversiell. Idén innebär till exempel att moraliskt ansvar inte endast handlar om individuella handlingar och ställningstaganden, utan också hur kollektiv agerar. En fotbollsspelare kan till exempel vara berättigad att känna stolthet för sitt lags seger i en match trots att han inte själv spelade under den matchen. En person
kan på ett motsvarande sätt känna rättfärdigad skuld för de brott mot mänskligheten som hens nation begick, även om personen i fråga inte var född när brotten begicks. Givet dessa kontroversiella men i vissa kretsar väl etablerade uppfattningar, så argumenterar jag för att vi har skyldigheter gentemot de arter vars utrotning vi har orsakat genom historien.


1. Har den här personen ett problem?
2. Kan medicinsk expertis hjälpa den här personen?
3. Är det effektivt och/eller kostnadseffektivt att använda sig av medicinsk expertis?

Om svaret är ”ja” på dessa frågor, så behöver vi inte längre veta om personen är ”sjuk” eller inte. Jag menar att denna modell kan dels utöka sjukvårdens möjligheter att adressera problem som idag beskrivs som ”sociala” eller ”moraliska”, dels användas för att begränsa behandling av vissa tillstånd där medicinska interventioner är relativt ineffektiva.

generationer. Givet att vi antar att icke-födda (och därmed blott möjliga) personer bör tilldelas ett värde i våra etiska överväganden, så blir den etiska magnituden i en existentiell katastrof många magnituder större, givet det stora antal människor som skulle kunna komma att existera i frånvaron av en sådan katastrof. I ljuset av detta kan åtgärder som minskar en sådan risk vara kostnadseffektiva, även om minskningen är liten.


5. References


Scheffler, R.M., Brown, T.T., Fulton, B.D., Hinshaw, S.P., Levine, P., Stone, S., 2009. Positive association between attention-deficit/ hyperactivity...