The Value of Technological Idealisations to Future Ethical Thinking

By

Adam Kelly

A thesis submitted to Victoria University of Wellington in fulfilment of the requirements for the degree of Master of Arts in Philosophy

Victoria University of Wellington

2016 Year of Submission

<u>Acknowledgements</u>

My work in this thesis owes much credit to my sole supervisor Nick Agar and fellow student Johnny McDonald. I would like to thank Nick for his encouragement, guidance and criticism. I would like to thank Nick and Johnny both for many enjoyable discussions, some of which were even related to this thesis.

Abstract

This thesis focuses on what I have called "technological idealisations", and how they are valuable to many current and future ethical debates. Technological idealisations refer to a methodology of using technology thought experiments to contribute to ethical debates. I do not claim this to be a new idea, and in fact will go on to give many examples of technological idealisation that already exist in the philosophical literature. The term describes the purposeful effort to collate these examples into a specific methodological framework; one which gives a particular kind of evidence which can ignore concerns of practicality and critically focus on the theoretical issues in a given debate.

In order to explore this idea I will first be looking at past, better known, examples of idealisations to facilitate understanding of my own. I will look at Rawlsian ideal theory as a template for my own idealisations, as well as to explain how they can be valuable in contributing to debates (in Rawls' case political and in my case ethical). Rawls' split up the field of political theory into ideal and non-ideal theory. Non-ideal theory is practical and works within the constraints of current political reality. Ideal theory idealises the political conditions to allow theorising regarding perfect political theory. The same can be done for ethics and for technology as it relates to ethics, as is my goal. Following on from this, I also examine Johann Roduit's use of ideal theory in the closely related field of human enhancement, in which he develops an interesting methodology of using ideals to guide human enhancement programmes.

However, rather than being concerned with Roduit's practical aim, my goal is theoretical. I want to take the ethical principles and theories themselves as ideals for technological development; in doing so technologies will be created, through the use of thought experiments, which agree with the theoretical aims of the theory or principle. These technologies can then be ethically examined and the resulting evidence can contribute (and has in the past contributed) to the ethical debate of those concepts and theories. The kind of evidence I see technological idealisations as offering ignores practical concerns and in doing so is also immune to criticisms of impracticality. This allows for more closely focused scrutiny of the ethical theories and principles themselves, undistracted by appeals to practicality which either argue for accepting a theory due to its utility or argue for rejecting a theory due to its impracticality.

Table of Contents

1. An Introduction to Technological Idealisations and their Value

This thesis focuses on what I have called "technological idealisations", and how they are valuable to many current and future ethical debates. "Technological idealisations" refer to a methodology of using technology thought experiments to contribute to ethical debates. I do not claim this to be a new idea, and in fact will go on to give many examples of technological idealisation that already exist in the philosophical literature. The term describes the purposeful effort to collate these examples into a specific methodological framework; one which gives a particular kind of evidence which can ignore concerns of practicality and critically focus on the theoretical issues in a given debate.

In order to explore this idea I will first be looking at past, better known, examples of idealisations to facilitate understanding of my own. I will look at Rawlsian ideal theory as a template for my own idealisations, as well as to explain how they can be valuable in contributing to debates (in Rawls case political and in my case ethical). Rawls split up the field of political theory into ideal and non-ideal theory. Non-ideal theory is practical and works within the constraints of current political reality. Ideal theory idealises the political conditions to allow theorising regarding perfect political theory. The same can be done for ethics and for technology as it relates to ethics, as is my goal. Following on from this, I also examine Johann Roduit's use of ideal theory in the closely related field of human enhancement, in which he develops an interesting methodology of using ideals to guide human enhancement programmes.

However, rather than being concerned with Roduit's practical aim of guiding human enhancement, my goal is theoretical. I want to take the ethical principles and theories themselves as ideals for technological development; in doing so technologies will be created, through the use of thought experiments, which agree with the theoretical aims of the theory or principle. These technologies can then be ethically examined and the resulting evidence can contribute (and has in the past contributed) to the ethical debate of those concepts and theories. The kind of evidence I see technological idealisations as offering ignores practical concerns and in doing so is also immune to criticisms of impracticality. This allows for more closely focused scrutiny of the ethical theories and principles themselves,

undistracted by appeals to practicality which either argue for accepting a theory due to its utility or argue for rejecting a theory due to its impracticality.

This is all best explained through the use of examples and fortunately there are many to be found in the philosophical literature already, so I have many to present my reader in this thesis. A significant portion of my examples regard enhancement and come from the enhancement literature. I will go on to explain more regarding the enhancement debate, however here I will say that the reason for this is that enhancement technologies tend to be ethically interesting in a way that other technologies often aren't; cognitive enhancement through genetic modification tends to make for a better debate than hands free segways. However, other examples are simply forms of technology which do not manipulate the human body. For this reason I didn't limit myself to a term which solely referred to enhancement and instead settled on one that encompasses all the potential examples: "Technological idealisations".

My first example concerns utilitarianism and the moral beta-blocker. Following on from the work of Nicholas Agar, I examine his technological idealisation of the ethical theory of utilitarianism and show how he arrives at the technology of the "moral beta-blocker"; a morally enhancing pill which supresses one's conscience when it interferes with that utilitarian's ability to maximise utility. Continuing with Agar's work I explain that this technological idealisation leads to abhorrent actions, and departing from Agar's company, show that this allows us to more acutely recognise the well know issues with utilitarianism demand to maximise utility.

My second example concerns hedonism and the experience machine. This well-known example by Robert Nozick focuses on the hedonistic theory of value and the technological idealisation of the experience machine. This machine offers a simulated and hedonistically superior life but is disconnected from reality. I agree with Nozick that this thought experiment allows for focused criticism of the hedonistic thesis that only pleasure is valuable, whether connected to reality or not.

My third example concerns autonomy and the God machine. Here I examine the work of Savulescu and Persson and their thought experiment creation of the God machine which

examines the concept of autonomy, more specifically the sacrifice of autonomy for safety. In this case I disagree with their conclusion that the technological idealisation of the God machine supports the sacrifice of autonomy, instead claiming that the world of the God machine is one that lacks any agency other than that of the God machine itself and should make us suspicious of the idea of sacrificing autonomy for safety.

My fourth example comes for the work of Anna Smajdor and concerns sex equality and ectogenesis; sex equality referring here to the principle of justice and ectogenesis being the technological idealisation in this case; artificial womb technology that removes the need from a female to gestate the foetus and give birth. In this case I agree with Smajdor that it shows that not all inequality between the sexes is socially constructed and that despite an inequality being biological in nature, the obligation to remedy it is not lifted.

My fifth and final example also concerns reproduction. This time the focus is on reproductive rights and male pregnancy. This time I look at the work of Robert Sparrow as he examines the technological idealisation of theoretical male pregnancy through the use of reproductive technology. Sparrow takes his thought experiment to show that the current characterisation of positive reproductive rights is problematic as it would lead to the endorsement of male pregnancy. Contrary to this, I believe Sparrow shows that sex equality once again demands the treatment of like desires alike which could include the desire to be pregnant.

After I have presented all of these examples I believe I will be justified in concluding two things. The first is that "technological idealisations" is a useful term which collates together many different thought experiments which already exist in the philosophical literature. The second is that these technological idealisations are valuable, and will continue to be valuable, to our ethical thinking. As you can see, the purpose of these examples is to convince the reader of the value of the method, with the conclusions which I come to in each example being less important. As I have briefly described there is disagreement regarding some of these examples, and though I believe I argue my case well I am willing to accept disagreement. What is important is that the debate itself takes place which itself is a recognition of the value of the method I present here.

1.1 An Introduction to Future Ethical Thinking

In this small section I wish to introduce the idea of how technological thought experiments affect ethical thinking. I claim that these thought experiments reach out into the moral content which is outside of current human capacities; whereas other thought experiments such as the famous trolley problem remain strictly within the realm of current human capacities. Both kinds are valuable, but I believe that technological thought experiments (especially those regarding significantly more advanced technology than is currently available) are less easily understood and often undervalued.

The moral can be sorted into three categories which can be thought of as three circles. The first circle is the smallest and is contained within the other two. This circle represents the actions with moral content which are currently possible by human beings; those goods and ills which we can achieve now, given our current nature and current technology. The second circle encompasses the first. This circle represents the actions with moral content which are logically possible; all those goods and ills which there is not logical barrier to achieving. These are the counterfactual arguments we imagine when making statements such as: "It would be good if there were no useless suffering in the world". Circumscribing the second circle is the third. This circle wall is the barrier to the others and represents logical impossibility, which contains within it all those actions with moral content which are logically impossible; these are goods which break the laws of logic and are thereby impossible. Examples are more difficult to come by here so let us for the sake of argument assume that those who claim backwards time travel is logically impossible are correct. We might still say that being a backwards time travelling Good Samaritan and, to use a tired but usefully uncontroversial example, stopping Adolf Hitler from rising to power in Germany would be good; giving us a logically impossible action with moral content. These are our three categories of the moral.

If we wanted to represent how technological thought experiment function with our newfound circles, we should show our first circle of actions with moral content currently possible by human beings expanding outwards, slowly encroaching further into the space of our second circle of the logically possible. This is what technology does; by increasing human capacities it expands the moral actions which are attainable which in turn expands

our responsibility as moral agents. In this thesis, I want to recognise that these thought experiments bring a more concentrated focus to old moral problems. How the circle of human possibility expands is logically determined by our ethical presuppositions. The technologies which these thought experiments reference will be those which are consistent with those presuppositions. Given this, it is possible to relate those technologies back to the ethical axioms from which they were derived. Therefore, how we regard those technologies will affect our understanding of the ethical axioms (in this case varying ethical theories or principles). This may lead to disapproval or approval, and in either case the deeper understanding which is achieved is valuable.

1.2 Technological Advancement and Enhancement

As I have mentioned, much of my discussion draws either upon examples from the enhancement literature or philosophers who are active in the enhancement debate. For this reason, in this section and the following section, I will be pre-emptively clarifying terms and concepts used in the current enhancement debate. As I have noted, the terms discussed here are themselves a matter of debate and I don't claim to have solved the debate in these small sections but my conclusions should be taken as definitive within the context of this thesis. I also believe that much of what I write in these sections applies to technology. In this section I discuss definitions of enhancement, however, much of this also has application when it comes to describing the advancement of technology, which I will note as I continue. In the following section I discuss the different sides of the enhancement debate, though this too can be applicable to the different views regarding the desirability of technological advancement. However, I will be focusing chiefly on the enhancement debate as it informs much of my work and requires a more detailed introduction than a discussion purely regarding technology.

Before I proceed I feel that it is prudent in any discussion referencing enhancement to first clarify what can be meant by enhancement and what I will mean when I talk of enhancement. There exist three popular definitions of human enhancement that are worth discussing. The first contrasts enhancement with therapy, the second associates enhancement with qualitative change and the third associates enhancement with quantitative change. I will argue that the first two definitions are "politically" motivated and

problematic, whereas quantitative change allows for a useful understanding of enhancement without the theoretical baggage of the others.

First let's look at enhancement defined as being beyond therapy. Therapy involves either reparation or restoration to the human average or to healthy functioning and any intervention which leads to functioning beyond these statistical averages is on this definition considered enhancement. This can be most clearly understood when looking at the same technology as both therapy and enhancement; one example is the use of anabolic steroids: its use for hormone replacement in cases of low testosterone is on this definition therapy but its use by athletes to improve their performance beyond their natural ability is an enhancement. This definition is favoured by bioconservatives who want to criticise (and in some cases entirely prevent) enhancement without therapy becoming collateral damage; if the two can be separated then the known benefits of therapy can be retained without permitting the potential dangers of enhancement.

However many wonder whether the therapy and enhancement can be separated in a truly normative sense, with transhumanists claiming that they cannot be and that bioconservative attempts are ad hoc and hold no normative force. They consider the concepts of the human average and healthy functioning to be arbitrary and therefore cannot justify a moral distinction between therapy and enhancement. Nick Bostrom and Rebecca Roache have a number of criticisms of the distinction in this vein (Bostrom and Roache 2008). They point to many areas of medicine which can't be considered therapy but are still ethically permissible (and even praiseworthy) such as preventative medicine, palliative care, cosmetic surgery, contraceptive treatment, vaccination programmes and more. Preventative medicine and vaccinations programmes seem rather to be examples of enhancement of resistance to diseases as opposed to a therapy which restores an individual to normal human functioning; to succumb to the diseases which these medicines prevent is normal human functioning. They highlight the problem of relativity when it comes to the bell curve of any given human property; How does one determine what is within the acceptable normal or average range? Do we include those with radical deficiencies of excesses in our calculation of that average? Do we have to take the average of all human beings in the world or just the population we live in? Do we take into account what stage of

life individuals are at which can lead to different averages and a different understanding of normal functioning? For these problematic ambiguities I will not use the therapy-enhancement distinction. It doesn't tell us what enhancement is; only that it is not therapy.

Another popular definition, this time often favoured by transhumanists, is that of enhancement as qualitative change or improvement. "The broadest concept of human enhancement identifies it with improvement. To enhance a human being is to improve him or her" (Agar 2014, 18). A positive of this definition when compared to the therapy-enhancement distinction is that it does give a clear definition of what enhancement is: Positive qualitative change. This also allows for measurement of the degree of enhancement relative to the amount of improvement. Another positive of this definition is that enhancement as improvement does capture the common sense understanding of the word and concept of enhancement. However, this leads to the one unforgivable sin this definition commits, which is to prejudice enhancements' acceptability and desirability by defining away any moral objection to it: "Once defined as improvement, human enhancement becomes unobjectionable" (Agar 2014, 18). "Improvement" is good by definition, to associate enhancement with it is to benefit from that built in positivity. Yet opponents are put in the unfair position of having to argue against improvement. This definition conflates the semantic and the normative questions of enhancement, which is to say that when we ask "Is enhancement good?" enhancement as improvement begs the question. To quickly divert to technology again, if all technological advancement was described as good by definition it would prejudice many current debates. For example, it would not be possible to condemn the development of nuclear armaments, biological weapons, or just weapons in general. This makes it even clearer as to why such a definition of enhancement is problematic.

Ruth Chadwick has her own criticisms of enhancement as improvement. She recognises the aforementioned problem of prejudicing moral acceptability, "A sense of enhancement is needed which does not prejudge the issue of acceptability and desirability. With this in mind the 'improvement' view is not helpful" (Chadwick 2008, 31), but raises her own concerns about the ambiguity of improvement and which sort of improvement should be focused on. She gives an example of an individual who can take pill that has no effects at all other than

to improve the hearing range of the recipient. She then notes that improvement of hearing range can be accompanied by degradation of hearing discrimination. In cases such as this, where the improvement of one characteristic comes at the expense of another, can the improvement still be called an enhancement? Why is the improvement of this characteristic more or less important than the degradation of another? The definition of enhancement as improvement is not helpful here.

Having problems with both of the definitions previously mentioned here, Chadwick offers her own; enhancement as quantitative change. "The sense of enhancement, I argue, which is to be preferred is the additionality view, where an enhancement is an addition or exaggeration of a characteristic which may or may not constitute an improvement" (Chadwick 2008, 31). This definition retains some of the same positives of enhancement as improvement while avoiding some of the negatives. It gives a clear definition of what enhancement is: Addition to a property or quantitative change. It can also be measured by the extent of the quantitate change. Additionally, it avoids the moral prejudice of enhancement as improvement; quantitative change can be good or bad (the popular adage "bigger is better" notwithstanding), and so the definition is morally open to interpretation. Unfortunately, for this very reason, enhancement as quantitative change doesn't capture the common sense understanding of enhancement in the way improvement does. Results such as "enhancing" someone's quantity of cancer doesn't capture what enhancement is generally understood to mean. There is also the recurrent matter of ambiguity with quantitative change as well. In keeping with the cancer theme, Chadwick touches on the case of preventative mastectomy; here we have a quantitative change of subtraction of an individual's breast tissue but a quantitative change of addition in the individual's defence against breast cancer. Here we have the same kind of ambiguity which Chadwick criticised enhancement as improvement for. However, to return to technology once more, I believe Chadwick's definition is useful. One may beg to differ given our current obsession with making our technology small and compact, but despite this the magnitude of technology is still increasing in these small packages and this is what is important regarding technological advancement. Similarly, this definition would not prejudice technological advancement; for example, we could have technological advancement in the form of ever increasingly

powerful nuclear weapons and while this would undeniable be an increase in magnitude it would not have to be accompanied by approval.

It is clear that none of the definitions of enhancement here are definitive, but I find Chadwick's enhancement as quantitative change to be the least problematic. It has the same advantages of a clear definition, and measurability as enhancement as improvement, but avoids the moral prejudice that comes with that definition. They both suffer from ambiguities. However, in the case of enhancement as quantitative change, I think it is possible to solve this problem by splitting the enhancement questions into two; what is enhancement? And, which enhancements are desirable? What enhancement is can be sufficiently answered by enhancement as quantitative change. Which specific enhancements are desirable is a question of prudential value and ethical debate.

1.3 The Supposed Bioconservative-Transhumanist Dichotomy

For most of its existence the enhancement debate has been ruled by a two party system, a split between bioconservatives and transhumanists. Given this, I think it's important, as I am discussing enhancement, to mention this debate and where this thesis stands in relation to it. In order to achieve this I'll outline both positions, and briefly describe the emerging moderate position.

In the western tradition of going from left to right, let me start by discussing transhumanism. It might be unfair of me to put transhumanism on the left side of the political spectrum. In theory transhumanism could be compatible with personal beliefs across the spectrum, however as Bostrom notes, conservatives are dramatically underrepresented:

"In principle, transhumanism can be combined with a wide range of political and cultural views, and many such combinations are indeed represented, e.g. within the membership of the World Transhumanist Association. One combination that is not often found is the coupling of transhumanism to a culture-conservative outlook. Whether this is because of an irresolvable tension between the transformative agenda of transhumanism and the cultural conservative's preference for traditional arrangements is not clear. It could instead be

because nobody has yet seriously attempted to develop such a position." (Bostrom 2005, 22)

This is not surprising as what transhumanists advocate for is often quite radical. In a very broad sense transhumanism can be thought of as the pro-enhancement camp. They believe that human enhancement technologies should be developed and used, though they will differ on which technologies should be adopted and for what reasons. The only further trait that can be attributed to transhumanists overall is their lack of concern for crossing the species boundary, and in some cases outright advocacy for it. Douglas mentions this point, "Transhumanists hold that we should sometimes use biomedical enhancements to radically transform ourselves, even if this means that we will no longer qualify as human" (Douglas 2013, 1633). They disregard species in general as having any moral content, and in so doing see no moral issue with crossing species boundaries. Going forward in this thesis we can understand transhumanism and transhumanists as referring to a pro-enhancement position, which either finds no moral issue with, or claims it is morally good (perhaps even required) that humans become post-humans.

It is perhaps also unfair that the anti-enhancement position has come to be known as bioconservatism, as although most are conservative there is also diversity to be found in the movement. As Bostrom also notes:

"Instead, they [cultural conservatives] have gravitated towards transhumanism's opposite, bioconservatism, which opposes the use of technology to expand human capacities or to modify aspects of our biological nature. People drawn to bioconservatism come from groups that traditionally have had little in common. Right-wing religious conservatives and left-wing environmentalists and anti-globalists have found common causes, for example in their opposition to the genetic modification of humans." (Bostrom 2005, 23)

The thing these groups are united by is a desire to maintain the current status quo of the human species. As mentioned, this is sometimes for religious reasons such as human beings supposedly being created in the image of a god, or for more secular reasons such as the value of human nature, either politically, socially or both. What can be understood as my use of the term "bioconservatism" beyond this point is that it loosely refers to the anti-

enhancement position which opposes both crossing the species boundary and using human enhancement technologies any way other than therapeutically (a point I will soon expand upon).

Finally, we have what I am sure the advocates of would like me to call the golden mean in the debate. Supporters of moderate enhancement (which I will have to refer to the position as for now until a more elegant name is invented for them) do sit in the middle of the debate, taking aspects from both transhumanism and bioconservatism. Agar describes moderate enhancement as follows:

"Moderate enhancement improves significant attributes and abilities to levels within or close to what is currently possible for human beings." (Agar 2014, 2)

The influence from bioconservatism is clear here and shows itself in the respect for human beings' attributes and abilities and perceived value in maintaining them within the species' boundaries. However, it is also clear that influence from the claims of transhumanists about the benefits of enhancement technologies have also been felt here. Enhancement technologies are not universally decried and are permissible (some may be encouraged) within the species boundary of Homo sapiens. We can think of moderate-enhancers as supporting enhancement while placing importance on not crossing the species boundary to become post-human.

1.4 'Ought Implies Can' and Inevitable Technological Advancement

Before I jump into the thesis proper, here I want to first anticipate some critiques that you might already be thinking to yourself. The first is that this may all be irrelevant as technology may not continue to advance, and if it does not then the moral problems that I claim will arise in the future will not. The second is that "ought implies can" saves us from having to worry about any of the absurdity outlandish future technologies which I will be talking about. The first critique can be avoided by appealing to the philosophical literature on enhancement, which reflects the defence of ideal theory in the political field. The second can be avoided by argument for the weak interpretation of "ought implies can" that allows for the existence of moral obligation outside of the current limits of human capabilities.

In this thesis I will be assuming that technology will continue advancing, not necessarily at the same fast rate that it is now but advancing nonetheless. There are a couple reasons why I feel comfortable making such an assumption: The first is that human history is the history of technological advancement, recent history even more so, which gives me little reason to think that will suddenly change now[1]. The second, more philosophically justifiable reason is that the literature and tradition of enhancement which I draw upon makes this assumption. It is true that there are many interesting discussions had about currently existing enhancement technologies, a lot of the debate surrounding enhancement technologies regards hypothetical technologies and the ethical debate over whether they should be research and developed or prevented. Given this, I feel content to argue with sources I draw from on their own ground and assume like they do that technology will continue to advance.

Additionally, much of my discussion will focus around Ideal theory. I do not want to anticipate too much of that discussion here and risk boring the reader with repetition later, so for now I will just say a little about how ideal theory justifies my focus on advanced technologies. Examine how Rawls' discusses ideal theory, "nonideal theory presupposes that ideal theory is already on hand. For until the ideal is identified, at least in outline - and that is all we should expect - nonideal theory lacks an objective, an aim, by reference to which its queries can be answered" (Rawls 2003, 89-90). Here we see that nonideal theory, which in this context would represent our current technology, requires some kind of identified ideal against which to measure it, which might be the highly advanced hypothetical technologies which I'll be examining. The value of ideal theory here is that it offers theoretical guidance regarding the nonideal theory associated with it. The hypothetical technology I look at offers this kind of theoretical interest, giving us interesting information about current ethical theories. This will become clearer when I examine my chosen examples but one thing to remember is that I believe the examples I look at are valuable whether or not the technologies referenced are developed or not.

---

[1] If, however, technological advancement were to stagnate or decline the inaccuracy of my thesis would be the least of our problems and given the prevalence of digital media my thesis may not even be accessible. So I write with the security that if my thesis is still accessible its assumptions will likely be correct.

Moving on to the issue of "ought implies can"; it is a common sense moral principle attributed to Kant. The fact that many dictionaries of philosophy which I consulted before writing this section did not include the principle can be seen as a testament to its common sensical nature. In one dictionary in which I found mention of the principle it was described as follows:

"A formula in Kant's ethics, meaning that correctly judging that a given agent is morally obliged to perform a certain action logically presupposes that the agent can perform it. He can perform it not just if he wants, prefers, or wills to, but in some absolute sense." (Nicholas and Jiyuan Yu 2004, 497)

As described here the principle appears straightforward and unambiguous; you cannot be obligated to do that which it would be impossible for you to do. No human is morally obligated to breathe underwater unassisted by technology or prevent the heat death of the universe. Such as obligation would seem absurd, along with any moral theory from which the obligation was derived.

However, like all things in philosophy, even the most common sensical, it is not free from controversy. Robert Stern in his paper "Does 'Ought Imply Can'? And Did Kant Think It Does?" examines multiple different uses of "ought implies can", and implies that the disagreements about the principle are born out of different interpretations of what the principle means. The interpretations can be divided into two broad categories; strong and weak. Stern sums up the difference between these two positions when talking about Kant's use of the principle:

"Kant is using the principle of "ought implies can" in a weak sense, by arguing that the moral law only has its status of being obligatory for us because we are able to act upon it, and that we can thus only explain this obligatoriness by accepting certain claims about our capacities and their conditions ('we ought implies we can'). But this is distinct from the claim that no act can be right (rather than just obligatory for us) unless we are able to perform it, which is how the principle is understood in the strong sense." (Stern 2004, 57)

The weak sense is described here as meaning that morality is obligatory only if it is possible to act on it. The stronger sense is described as human capacities circumscribing what is moral; an act can be right only if we can perform it. Clearly these are two very different interpretations. The weak sense is also interesting in that it leaves open the possibility of there being moral content which it is not possible for humans to currently achieve; for when talking about Kant's moral law Stern claims that it is only obligatory because it happens to be possible to act upon it, presumably it would not be obligatory (though importantly it would still be moral) if it were impossible for humans to act upon it. As an example, let us examine Kant's famous prohibition against lying. Kant thought it was possible for human beings to live their lives without lying (if Stern's interpretation is correct) but let us more realistically assert that such a thing is impossible. For the strong interpreter this would mean that a moral law such as do not lie (ever) would be nonsensical, it would have no moral content good or bad because such a law is outside the boundaries of human possibility and therefore outside the boundaries of the moral. Of course this would also mean that it is not obligatory. For the weak interpreter such a law could still be considered good, and desirable if somehow the state of affairs was organised in such a way as for it to be possible, but it would also not be obligatory as it is not possible for us to act on such a law. Equally, it would be good for you to end all suffering in the universe immediately, but you are under no obligation to do so.

The weak and strong senses of "ought implies can" have different implications when applied to the field of enhancement and technology in general. On the weaker interpretation of "ought implies can" enhancement and technology is much more at home. In this case we can think of enhancement and technology as simply allowing us to access the moral content which currently falls outside of our current capacities. A very simple minded example could be that it is currently impossible to save a person falling from a tall building by flying up to them and grabbing them, but would be possible if we enhanced human beings with the power of flight or perfected personal jetpack technology. This would give us access to this moral good which we previously didn't have access to as well as widening the scope of our obligation to now include saving people falling from tall buildings. Given this more natural fit with the weak interpretation, in the reminder of this section I want to continue along

Stern's reasoning and argue for the weak interpretation, and establish it as a secure basis to continue my argumentation throughout this thesis.

I will begin by critically examining the arguments of those who endorse the strong interpretation, starting with this concern which Stern attributes to James Griffin:

"Griffin has argued for what might be called a greater degree of realism in ethics, in the sense that we should begin by understanding ourselves and our capacities, as a necessary first step to thinking about moral issues. He claims that moral theories have too often neglected facts about human nature and society, and as a result have become distorted and inadequate to our real needs: We have theorized in a vacuum, and so have failed to do so successfully." (Stern 2004, 44)

Clearly this is an appeal to practicality, claiming moral theories which ignore human capacities fail to achieve the goal of guiding moral action because they are impossible to act on. This is a common theme running through the arguments of advocates of the strong interpretation and critics of the weak interpretation:

"The stronger view, that a theory which argues for principles that are unrealizable by us must be wrong, in which case here "ought implies can" is once again being used in a strong sense." (Stern 2004, 46)

The above quote sums up the strong interpretation. Its criticisms of the weak position, as expressed by Griffin, are summed up by Stern thusly:

"Now, Griffin's position here rests on two assumptions. First, if a moral rule says that what is right is something we cannot do, it is pointless; and second, if a moral rule is pointless, it cannot really be a moral rule." (Stern 2004, 50)

So there are two claims, each of which is would be problematic for my work. The first is the softer claim that a moral rule or moral good which cannot be performed or obtained has no value. The second goes further, making the additional claim that pointless moral rules cannot exist.

To respond to these concerns, one need only think of some possible value that could be had for moral rules or goods that cannot be obtained. Stern himself offers a potential value: "Many of us admire certain figures or acts which we know we could not follow or even try to follow because of our own incapacities, where nonetheless this admiration gives these exemplars a kind of point" (Stern 2004, 50). Think of the pious Christians who try to live their lives by Jesus' example, and strive to be like him. They know that by his very nature they could never be perfect as they believe Jesus to have been but his ideal has inspired many to do great things[2]. This fits well with the idea that there is moral content that currently exists outside of our current capacities, and even if we cannot attain it right now or ever it serves as an ideal. This not only makes them valuable but recognises the contingency of those goods being unobtainable, and can allow for obligations to work towards obtaining them in the future.

Another reply challenges the human relative nature of the strong interpretation: "It is harder to see why, to be right, an act must be a duty for us, something we are obliged to perform, any more than it must be a duty for a dog or a monkey" (Stern 2004, 48). This simply makes clear the absurdity of having human capacities circumscribe morality. I can take Stern's thought further and imagine that we were visited by aliens which had superior capacities that were far greater than our own. The story of Superman is a good example of this: with his greatly enhanced capacities of speed, strength, flight, etc. he has the abilities to do more moral good than human beings. He is able to perform moral acts that no human being can perform. What is a supporter of the strong interpretation to say about Superman singlehandedly stopping a plane from crashing and saving all those on-board? Can that not be a moral act because it's impossible for a human to do the same? Would they also deny that Superman might be obliged, may have a moral duty, to perform the act? This is not to say that we as humans have the same moral obligations as Superman. Given that he is able to perform the act and we aren't, he has an obligation where we do not. However, to try to claim that as Superman's action lay outside the boundaries of human capacities they cannot have moral content, or that we cannot claim it would be good to have the same abilities as Superman from a moral standpoint, is absurd. It is obvious that the only reasonable response is to recognise that what is morally good (not prudentially good) is in no way

---

[2] And terrible things, but let's leave that aside for the sake of this example.

determined by human capacities, only the good that we can perform is, and even that is contingent.[3]

Finally, a supporter of the strong interpretation might respond to Stern's claims and my only, by claiming that moral acts need not be relative to human beings as a whole but to the capacities of each individual agent. This would mean that superman could have different duties to a regular human without completely abandoning the relativism of the Strong interpretation. Superman ought to do only what he can, though that is a lot more than any human could. However, this strategy would commit would anyone who proposed it to relativisms about all forms of value and valuable acts:

"Moreover, unless one embraces relativism, there must be some way of assessing the value of an action that is independent of the capacities of the agent, as when we judge that it would be better if the child could read Shakespeare rather than just Harry Potter books, or that we could act more impartially rather than less so. But surely this requires us to judge the value of the act as right on its own merits, regardless of the capacities of particular agents?" (Stern 2004, 51)

This is just one example, in this case moral relativism regarding the relativism of moral acts bleeds over to literary value and relativism regarding the value of literary acts. It is important to note here that this is not to claim that a support of the weaker interpretation would claim that children have an obligation to read Shakespeare over Harry Potter, only that Shakespeare is of greater value as a literary experience than Harry Potter and if children did have the capacities necessary to understand Shakespeare it would be more valuable for them to read it over Harry Potter.

In summation, the weak interpretation of "ought implies can" is certainly the least problematic interpretation of the two presented. Not only is it not pointless, in that the moral good beyond our capacities can serve as ideals to strive towards and can guide our enhancement projects and technological advancement, but it doesn't suffer from being overly human centric, or from collapsing into general relativism regarding value. What this means for technological idealisation is that I can firmly establish (or at least justifiably

---

[3] Of course there might still be goods which are logically impossible as previously mentioned.

assume) that there is moral content outside the realm of current human possibility which they can draw upon. Also, it can be understood that though these idealisations may never be achieved they can still be valuable.

Recognition of this inevitable widening of the scope of the moral content beyond human limitations and the potential increase in our obligations can already be found in the enhancement literature. However, much of the dialogue in the enhancement literature concerning moral obligation has so far focused on the potentially problematic enhancement of the moral status of post-humans[4]. The point raised here, namely the enhancement of moral obligation itself, has received far less attention. Despite this there are still examples to be found. Michael Sandel in his book "The Case Against Perfection" gives a somewhat pessimistic and bioconservative account of the enhancement of moral obligation:

"Whatever one believes about which, if any, genetic conditions warrant terminating a pregnancy (or selecting against an embryo, in the case of preimplantation genetic diagnosis), the advent of genetic testing creates a burden of decision that did not exist before. Prospective parents remain free to choose whether to use prenatal testing and whether to act on the results. But they are not free to escape the burden of choice that the new technology creates. Nor can they avoid being implicated in the enlarged frame of moral responsibility that accompanies new habits of control." (Sandel 2007, 88-89)

Here we see this expansion of the moral described as a "burden", noting that the advancement of technology brings with it new moral problems which can no longer be avoided as fantasies. Sandel characterises this as something lamentable but unavoidable, put upon us by our own hubris. A more optimistic account is given by Buchanan et al. in "From Chance to Choice", which relates to genetic interventions and how it is opening up new moral possibilities. The fine phrase they come up with to describe the process is called "the colonization of the natural by the just".

"To explore how the prospect of genetic interventions with human beings challenge existing ethical theory. The challenge takes two distinct forms. First, the prospect of vastly increased powers of genetic intervention brings with it the inevitability of new choices, the

---

[4] See Agar 2014, Buchanan 2009, and Fukuyama 2002 for some examples of this debate.

contemplation of which stimulates us to articulate existing ethical theories in greater detail[5]. Second, by placing within human control features of our condition that we have heretofore regarded as given and unalterable (the fate assigned to us by natural lottery), the prospect of genetic interventions forces us to rethink the boundary we have traditionally drawn between misfortune and injustice, and indeed between the natural and the social." (Buchanan et al. 2000, 18)

This quote describes the issues which technological idealisations try to address. As I have already described we must rethink the boundary of the moral; recognise that it is changeable and that through technology it is changing. And we need to "articulate existing ethical theories in greater detail", which for me is a process that starts by putting our current ethical theories and concepts to the test through technological idealisation. Before I begin to do that myself, in the next chapter I will first explain exactly how the methodology will do just that.

---

[5] In this case Buchanan et al. are distinguishing different variants of level playing field theories of equal opportunity, which appear to have different practical implications.

2. Technological Idealisations

In talking about "technological idealisations", a phrase I expect to cause some bemusement, I think it helps my case to look at past concepts of idealisations and in particular past concepts of idealisations involving technology and human enhancement. Rawls' distinction ideal and non-ideal theories provides the best example for the former, and for the latter I can look to Johann Roduit who explicitly discusses idealisations and ideal theory as it relates to enhancement in a co-authored paper "Evaluating human enhancements: The importance of ideals" (with H. Baumann, and J.-C. Heilinger 2015), and individually expands the discussion in his paper "Ideas of perfection and the ethics of human enhancement". After discussing these works I think there will be strong groundwork to understanding the concept of idealising ethical theories through technological thought experiments.

2.1 Rawls' Ideal and Non-Ideal Theory

The distinction between ideal and non-ideal theory which this particular discussion will be founded on comes from the work of John Rawls. It can broadly be described as a distinction between pragmatic theories and idealised theories. More specifically, Laura Valentini describes three ways that the ideal vs. non-ideal distinction has been used in the literature of political philosophy; the first interpretation corresponds to full-compliance vs. partial compliance theory, the second corresponds to utopian vs. realistic theory and the third regards end-state vs. transitional theory.

I start by looking at the first interpretation of ideal and non-ideal which is full-compliance vs. partial compliance theory (Valentini 2012, 655-656). This category of theories are as they sounds; full-compliance theory assumes full compliance on the part of its theoretical agents, partial compliance theory assumes (more reasonably) that many agents will fail to fully comply and theorises accordingly. The ideal theory in this case is full-compliance theory as it idealises the agents' compliance, and also the conditions necessary to obtain that compliance. Rawls' himself does this exactly in "A Theory of Justice", in which he idealises both agents and conditions for political theorising: "(1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles" (Rawls 1999, 4). This is

obviously not the case in our unideal world, and so non-ideal partial compliance does not make these assumptions and theorises within the constraints of our world. This interpretation does have relevance to technological idealisation which does assume full compliance with the ethical theories it idealises. However, we will come to see that this is not the interpretation which Roduit comes to use when discussing enhancement.

The second interpretation which Valentini examines is utopian or idealistic vs. realistic theory (Valentini 2012, 657-660). Rather than compliance, this interpretation is concerned with feasibility. The debate is between whether political theories should be idealistic, theorising towards the perfect utopia, or whether they should be realistically constrained by feasibility concerns. This is also somewhat relevant to the field of enhancement in which it is sometimes debated as to whether it is worth talking about technologies that don't only currently not exist but seem more at home in a science fiction magazine than a philosophical paper. This thesis contains examples of such technologies which might reveal which side of the divide I can be placed, however I, along with Valentini, see value in both approaches; realistic theory helps with the practical application of the ethical principles arrived at through idealised theorising. Again, Roduit does not focus on this interpretation either so I will leave further discussion of it aside.

Valentini's third and last interpretation regards end-state theory vs. transitional theory (Valentini 2012, 660-662). In the political context ideal theory corresponds to end-state theory here, focused on identifying the perfect political system, and non-ideal theory corresponds to transitional theory which can either refer to theory focused on improvement without any particular ideal end-state in mind or non-ideal improvement designed at transiting towards a particular end-state. This third, end-state, interpretation is the one which Roduit focuses on in his discussion regarding enhancement and so will also be my focus going forward.

The reason that Roduit focuses on the third interpretation is that he is concerned with the goals (or end-states) of enhancement programmes and how to work towards them. For this reason the third interpretation is ideal[6]. Roduit is not concerned with compliance to theories, he is focused on ideals rather than theories, and so the first interpretation does

---

[6] If you'll pardon the pun.

not hold his interest. The second use is also not relevant to Roduit's focus on the goal of enhancement. Although an interesting discussion regarding the feasibility of proposed enhancements can be valuable it doesn't tell us anything about the goal of enhancement. For example, if all proposed enhancements for radical life extension were unfeasible this would not tell us anything about the goal of these enhancements other than the fact that they will not be attained. It does not tell us what the goal of enhancement can or should be; as I have previously noted, goals need not be attainable to be valuable. It is not irrational to have goals that are unattainable. A scientist who strives for omniscience might be striving for the impossible but in doing so could find herself working harder and creating valuable research even if she personally finds it underwhelming. Discussions of enhancement which ignore considerations of feasibility have historically been valuable; one such example is the presumptive discussions which philosophers have had regarding human genetic modification before CRISPR. For a long time genetic modification was a haphazard and unreliable technology, but philosophers engaged in many theoretical discussions which disregarded this imagining hypothetical technologies[7] that worked perfectly. This work had its own value theoretical value but now, with the new technological developments like CRISPR (Le Cong et al. 2013), it may soon have practical value. It is my aim that this thesis will be another such example of a forward thinking piece of work, valuable enough in theory but perhaps eventually also in practice. For this reason Roduit doesn't focus on the second, feasibility concerned, interpretation either.

Roduit's favoured third use focuses specifically on goals (Roduit et al. 2015, 208-9). His idea is that we should have a particular ideal or (to use Valentini's words) end-state which guides enhancement as a whole or any particular enhancement. Ideal or end state enhancement is guided by the ideal in order to reach an end state, most likely the instantiation of the ideal itself. Non-ideal or transitional enhancement claims to remain neutral as to the overall ideal guiding enhancement and is content to compare two particular enhancements and then decide which the best option is. How the latter is even possible without some kind of objective ideal is a problem both Roduit and I will come to address later in the chapter.

---

[7] Or should I say "technological idealisations"?

It is worth nothing that this interpretation of end-state ideal theory finds support in Rawls' work in "The Laws of People", which provides much discussion of ideal and non-ideal theory.

"Nonideal theory asks how this long-term goal might be achieved, or worked toward, usually in gradual steps. It looks for polices and courses of action that are morally permissible and politically possible as well as likely to be effective. So conceived, nonideal theory presupposes that ideal theory is already on hand." (Rawls 2003, 89-90)

Here it is explicit that ideal theory is working as a goal for non-ideal theory, in the same way that Roduit wants some ideal to guide practical (non-ideal) enhancements. He calls it "forward looking" (towards the ideal) or "end state focused enhancement", and I will go on to explain how he argues for the benefits of this analysis is the form of capturing what is meant by the term "enhancement", offering a method of measurement for enhancement, and setting a goal for enhancement.

2.2 Enhancement vs. Human Enhancement

Before I critique Roduit's ideal focused analysis of enhancement I must complete its description by examining his chosen ideal. Roduit makes it very clear that he is only interested in quite literal "human enhancement"; enhancement within our species boundary. He states that he is not interest in enhancement which would take a being from "An animal to a human, from a human to a posthuman, or from a posthuman to an animal" (Roduit 2015, 5). This puts him squarely in the camp of "moderate-enhancers", rejecting the bioconservative position of absolute prohibition while also resisting the transhumanist temptation to enhance wildly beyond current human limitations.

Consider Roduit's human enhancement contrasted with Nicholas Agar's concept of "radical enhancement". Much of Roduit's work in these two papers shares similarities to Agar's work; in particular this rejection of radical enhancement and a desired "human" focused enhancement programme. Agar describes radical enhancement as "improving significant human attributes and abilities to levels that greatly exceed what is currently possible for human beings" (Agar 2010, 1). Take life extension enhancement as an example: An

enhancement that caused the average life expectancy in a given population to rise to one hundred years would not be radical enhancement as although such life spans are currently rare, they do not "greatly exceed what is currently possible for human beings". An enhancement that caused the average life expectancy to become a thousand years would and is therefore a case of radical enhancement. Both Agar and Roduit want non-radical enhancement within the species boundary of Homo sapiens.

Due to this focus on human enhancement, the ideal that Roduit selects to guide his enhancement programme is that of the ideal or perfect human being. "The forward-looking, end-state or ideal approach asks: "What would be a perfect or ideal human being?"'" (Roduit et al. 2015, 209). The answer to this question is vague but given the diversity that currently exists among human beings some vagueness seems appropriate. "Although it is not possible to agree on what an ideal human would look like in all circumstances or at all times, we can nonetheless look at what some characteristics of an ideal human would be. "I refer to these characteristics here as perfectionist assumptions of what it means to live a good human life. In other words, to be considered an ideal human, one would have a "set of ideal human properties that allows for evaluations of enhancement interventions" (Roduit 2015, p.4). This quote requires much unpacking which I feel is best done by analogy with platonic idealism.

2.3 Platonic Idealism and the Human Form

I want to stress before I continue that this is neither an endorsement of Platonic Idealism, nor a full representation or description of its complexities. Here it is simply used as what will hopefully be a clarifying analogy for fellow philosophers, especially those familiar with classical philosophy. In basic terms, Platonic Idealism refers to (among other things) the theory of the forms. The forms are unchanging abstract objects that exist outside of the material world, but are said to act as blueprints for things in the material world. The best description for my purposes comes from the platonic dialogue Parmenides:

"These Forms are like patterns set in nature, and the other things resemble them and are likenesses [replicas] of them. For the other things, this partaking of the forms turns out to be no different than being likened to them."

"Then if something resembles the form, can that Form not be like what has been likened to it, to the extent that the thing has been made like it?" (Plato 2010, 91)

The forms are the ideals or perfect forms of every object in the material world. Material objects can resemble or partake in their form to a lesser or greater degree, as the text says: "To the extent that the thing has been made like it". The example used in every philosophy classroom is that of the chair; there are many different chairs in the material world but they are all chairs by virtue of partaking in the form of a chair. The form of the chair is the perfect chair, and this allows other chairs to be judged as better or worse depending on how much they partake in the form or how "close" they are to it in metaphysical terms.

This aspect of judgement is useful for talk of enhancement. This platonic sentiment is echoed in Roduit's writing: "Assumptions about an ideal are the precondition for evaluating states of affairs regarding to their proximity to this ideal" (Roduit et al. 2015, 209). The ideal here is of course not the chair, but the human ideal. We can judge different states of affairs for humans as better or worse compared to the human ideal and this will give us a guide as to how we should go about human enhancement.

However, it is important to remember this enhancement towards the ideal human has a limit. The analogy with Platonic idealism becomes useful here again in order to understand why Roduit is only concerned with "human enhancement". Agar explains in more detail that to enhance beyond the point of humanity would not only by unethical but would also lose its prudential value for the subject. Take cognitive enhancement as an example: Improved cognitive ability within the human species boundary would be valuable to us as humans by virtue of us being humans. We would be able to tackle the typical human decisions that we make daily, with a better brain. On the contrary, a huge change in cognitive ability perhaps to the point where a human being could understand all of the casual chains leading up a decision and know with pinpoint accuracy how their plan of action would affect the future (the debate surround determinism notwithstanding) would certainly be a greater magnitude of cognitive change but it would not be valuable to a human being qua human being because they would no longer be human, or the very least Agar wants to claim is that the value would not correspond exactly with the objective greater degree of its enhancement or quantitative change.

"According to the *anthropocentric ideal*, some enhancements of greater objective magnitude are more prudentially valuable than enhancements of lesser magnitude. However, some enhancements of greater magnitude are less valuable than enhancements of lesser magnitude. Such assessments are warranted for enhancements of our capacities to levels significantly beyond human norms." (Agar 2014, 17)

To return to the case of cognitive enhancement, he claims it is wrong to assume that a billion-fold enhancement of cognitive ability must be more prudentially valuable than objectively less valuable enhancements. This is because humans typically assess the value of an enhancement in terms of what they qua humans can do with it. In the cognitive case this might include passing an exam or some other kind of academic test, however billion-fold cognitive ability would make any academic test completely irrelevant and there would be no value to be had from passing it. Therefore the enhancement fails to be prudentially valuable to the human being trying to pass their exam, despite being much great in magnitude some smaller enhancement which might only enhance an individual's IQ a mere ten points (but will important retain the desire and prudential value to passing the exam). In the same way, if, post enhancement, the enhanced are no longer human then this common method of evaluation from our human perspective will fail.

Agar refers to this process of losing humanity through radical enhancement as "transformative change", and it is this that makes radical enhancement lose its value when it pushes the human being into the territory of post-human. The analogy with platonic idealism becomes useful here again. Take the example of the chair again. A chair manufacturer might claim that a chair design could be enhanced by increasing the size of the chair for more people to be able to sit on it, by increasing the amount of legs to make it more stable, increasing the weight so that it won't be easily knocked over or blown over in the wind and so on. All of these quantitative changes of enhancement could be applied but when the design is complete the chair manufacturer is distraught at the fact that he no longer has a chair but a bench and wonders where he went wrong. In Platonic terms he went wrong by not focusing on the ideal of the chair and his vision become corrupted by the mantra bigger is better. Radical enhancement that would change us from a human to post-human makes the same mistake, by abandoning the ideal of the perfect human for

something else that we, as humans, are disconnected from and therefore cannot be valuable to us as humans. Imagine that an individual wants to radically extend their life beyond current human limitations to around a thousand years; this is something Agar considers as an example of a self-defeating enhancement (Agar 2014, 57-60). The reason is that the desire to extend one's life is tied to identity, an identity which due to the nature of such a radical transformative change will likely not survive. To fully explore the idea a complete discussion on the philosophy of identity would be necessary but it is reasonable to assert that having a radically extended lifespan would change one's personality and identity, potentially to such a degree that the connection between person-stages is not sufficiently preserved. This means these enhancements are quite lacking in value from a subjective perspective as this perspective which motivated the initial desire is not likely to survive. Other forms of radical enhancement that involve transformative change are, it is claimed, likely to suffer this same loss of value.

2.4 Which kind of perfection is best?

When talking about what is valuable, Roduit makes an important distinction between "type-perfection" and "property-perfection", with his theory embodying type-perfection. He describes type-perfection as regarding the enhancement of "the essential properties of the individual's type or species" (Roduit 2015, 6). This fits well with Roduit's species relative theory and so it is no surprise that this is the perfectionism which he favours. Property-perfection is described as enhancing "those individuals who best realise some property or properties" (Roduit 2015, 6). This is enhancement irrespective of type and so could include enhancement from human to post-human which type-perfection would rule out. The enhancement is relative to an abstract property, not the type or species concept. Roduit defends type-perfection by suggesting that property-perfection is fundamentally flawed as a self-sufficient position because it necessarily collapses into type-perfectionism. "A property is always associated with a type", he claims, offering the property of speed as such a property: "The property of speed has to be embodied in a type" (Roduit 2015, 9). He gives the example of an athlete having the property of speed, or being fast, relative to the human type, but not relative to other types such as posthuman or objects such as planes.

Roduit reiterates his strong statement that "property cannot exist without a type" (Roduit 2015, 9). Despite following along with Roduit's reasoning so far, his claim that property cannot exist without type is wrong. I am willing to grant that many properties might necessarily be type relative, for example weight and height, however he claims that all properties are type relative. Take intelligence, it is a property that does not rely on any type. It is true that we often discuss and measure intelligence in a way that is human relative but this is merely a heuristic. It is possible to conceive of optimal intelligence unconstrained by type; the concept of God is often ascribed the property of omniscience which is to say maximal intelligence. Omniscience is an example of property-perfection which is not relative to any type-perfection. I believe that there could be other properties of this kind as well but one example is enough to show that the statement "property cannot exist without type" is false.

However, despite this I agree with Roduit's decision to endorse type-perfectionism with the caveat that two changes are made. First, in order to address the problem mentioned above, I think we should retract the statement that "property cannot exist without type". Instead what we should say is "property has no value without type". Let's take the example of the property of speed. We can make sense of speed without type; we can and sometimes do say a human is faster than a snail, but slower than a cheetah. This shows that we can abstract speed away from type as a property against which to compare different types. However, abstracting away the value of the property doesn't seem to work as well. Imagine we could give a snail the speed of a cheetah; it would not be of value to the snail and could even be detrimental. It is hard to see how a snail could function with the speed of a cheetah. It would either not utilise the capacity or it would likely end up injuring itself by moving too fast. Thus we can have property without type but not value without type.

Roduit and Agar take this to show that enhancement should be anthropocentric in the case of humans, and species relative in the case of all species. I think the mistake shown here is that while we shouldn't enhance properties abstractly away from type, what is needed is a more comprehensive and epistemically justifiable type which can be offered by theory. Our value judgements are theoretically based and this is what should guide our enhancement. Theory can ground our value judgements without relying on naturalistic and epistemically

suspect concepts such as human or more generally species. Enhancement is guided by what is valuable. This is the only reason to pursue the development of any technology; either it is prudential valuable to those who seek to develop it or perhaps has moral value, for example in developing cures for terrible pathologies. In turn, what is valuable is determined by our theory of value. As we have seen, the idea of objective value that it not relative to type fails, but the type in question should not be species or the vague concept of 'human being", but theory itself. Furthermore, I claim that Roduit agrees with me, albeit not explicitly, his concept of "human being" is theoretically, not biologically, based.

2.5 Homo Sapiens and Human being

I want to begin this section by showing that Roduit could not have any biological species in mind when determining his "ideal human being". I do this in order to anticipate the predictable reply that his concept of the ideal human could be atheoretical by focusing exclusively on the empirical ideal of the human species and therefore my claim that enhancement programmes should be based on theory is wrong. To take this criticism seriously, let me examine the meaning of species in a biological setting. The first thing to note is that there is a lot of ambiguity and little agreement regarding species. If one is going to attempt to ground value in species in an atheoretical manner, they must endorse some kind of species realism (as grounding value in anti-realism regarding species would be theoretical). The problem is that species realism is not a closed case given the debate surrounding species concepts.

First one must choose amongst the many species concepts currently in use in biology. Writing on this topic, Marc Ereshefsky has this to say about species concepts; "A plethora of definitions, or what biologists call "species concept", exist. Not merely one or two prominent definitions, but at least seven well-accepted ones" (Ereshefsky 1998, 103). In an admirable attempt to settle the species controversy D.L. Hull also settles on seven "well-accepted" species concepts and attempts to establish the most promising candidate, ranking it against the others, but his investigation was regretfully unsatisfying: "I have made up charts on each of the possible permutations, and the depressing result is that all seven species concepts score about the same. This outcome may well explain why the species controversy continues unabated" (Hull 1997, 20-21).

The problem goes further than just a plurality of species concepts, the plurality itself has led to scepticism regarding the concept of species. Some philosophers of biology have taken species pluralism to support the idea of species anti-realism; that species do not exist but are simply a useful biological heuristic (Ereshefsky 1998). It would take more than my limited knowledge and research to fully examine this idea here, but its mention is still valuable as it is a real concern for any attempt at grounding value in species. However, it should also be noted that many do not follow the same line of reasoning and not all philosophers of biology are sceptical about species; "In our view, evolutionary theory lends no support to the idea that our species classification do not reflect objective features of the living world" (Sterelny and Griffiths 1999, 182). The fact that this position is still held is enough for me to be generous here, and continue this discussion assuming that this species concept problem can eventually be overcome.

Roduit does not address specifically which species concept he invokes when talking about human beings (as I claim he endorses a theoretical understanding of "human being"), but Agar does give us an endorsement of the unfortunately named "biological species concept"[8] (Agar 2010, 19). On this species concept a species is a group of organisms which are capable of interbreeding successfully and are reproductively isolated from other groups. This means that human beings, or Homo sapiens, are defined by their ability to interbreed successfully with members of their own group and not with other organisms outside that group. Agar goes on to rightly acknowledge the concept as controversial and by no means definitive. He also believes species concept pluralism is the right course of action for biologists with varying interests and focuses. As a philosopher, he concludes that the biological species concept is the one which best suits his purpose; it is useful for identifying valuable experiences. These valuable experiences are not just of sex and reproduction, as a crude initial reading might suggest. Agar claims many of our prudential values are connected to the possibility of reproduction; social experiences, friendships, love, childrearing, and so on[9]. Let us go forward and examine if there are any problems (in addition to the

---

[8] All the seven plus species concepts are "biological" species concepts but it appears that the one described here got to pick its name first.

[9] However, I do think Agar mischaracterise the biological species concept here and fails to see that it can't make the distinction between human and posthuman that he wants. You might radically enhance a human being to the greatest extent possible, but as long as you leave the reproductive organs untouched it will still be human on the biological species concept.

aforementioned contentious debate surrounding species in the field of biology) with trying to ground value in species.

The real problem with attempting to appeal to any species concept or biology in general is that it will come with all the theoretical baggage from biology, in particular the theory of evolution. This means that value and thereby enhancement will simply be guided by a biological theory rather than a normative theory. I do not personally harbour scepticism towards the theory of evolution, and therefore I am willing to grant that it has the biological story of humanity correct (which is to say that I am willing to grant it the same epistemological status I would grant any scientific theory). Even so, a biological theory is no less a theory if it is a correct theory, and therefore would still not be atheoretical. Not only that, but it is a poor theory to use to guide enhancement; being one of the few cases in which we have experiential evidence. Social Darwinism is the embodiment of the concept of using the biological theory of evolution as a normative theory. The scientific problem with Social Darwinism is that it was simply bad science, greatly biased by the racist attitudes of the late 19th and early 20th centuries. The philosophical problem with Social Darwinism was that it failed to acknowledge the distinction between fact and value, or between ought and is (which is not to claim that the Social Darwinists were correct about what "is". Again, this is the scientific problem). This is why a biological theory is not enough to guide enhancement; there must be a motivating value to be gained through enhancement, which can only come through some normative theory of value.

Thankfully, I believe we can interpret Roduit as not referring to human beings in the biological sense at all but to a theory of human well-being as outlined by Nussbaum's capabilities approach: "Nussbaum outlines ten central capabilities. These capabilities are so fundamental for humans that without them, life would be impoverished. Without them, it would therefore be impossible to think of type-perfection" (Roduit 2015, 18). What exactly those capabilities are, is not important to her. What's important is that this clearly establishes what it means to be a human being, or to have a flourishing human life. Roduit makes it clear that he is focused on the capabilities approach as only a method of understanding what it means to be the ideal human; humanity is what he is claiming to idealise, not the theory. Despite this, I believe Roduit would be better off retroactively

endorsing my interpretation, claiming that it is the capabilities approach which he is idealising here[10].

The capabilities approach is clearly a theoretical approach which most closely resembles objective list theories of wellbeing. I propose that it is possible to think of the capabilities approach as a self-sufficient theory of wellbeing, like an objective list theory, though it can also be thought of as listing the necessary conditions to facilitate any meaningful theory of wellbeing. Either way, the capabilities approach is a theory of value, whether self-sufficiently by claiming the value of the ten central capabilities or by conceding that value to the theories of wellbeing which it directly facilitates; it does not rely on some naturalistic sense of human being to derive value from but philosophical theory of value. This becomes particularly evident when we look at two of the ten capabilities of Nussbaum as listed by Roduit, which are "bodily integrity", and "other species"; how one could even attempt to derive bodily integrity, which refers to the complex philosophical value of autonomy, or the value of other species, from human biology I have no idea. It is shown here that Roduit's enhancement programme is not guided by the ideal human being. It is determined by a theory of value.

2.6 What are technological idealisations?

What Roduit has done in the case of Nussbaum's capabilities approach is exemplify how the first step of technological idealisation is supposed to work: Some normative theory is chosen and idealised, turned into an ideal theory if it is not already, which then theoretically guides our technological developments or enhancement programmes . Through the contemplation of thought experiments (many of which are to follow) we discover what technologies would be developed if such a theory was to be our guide. If after rational ethical consideration the results appear immoral then we can think of the technological idealisation as working much like a *'reductio ad absurdum'*, providing evidence against the theory from which it was derived. However, if the results seem ethically permissible or desirable after rational consideration then they pass the test and the evidence now works in favour of the theory from which the idealisation was derived.

---

[10] Doing so would not affect the conclusions of his arguments at all, but would leave them less open to criticism of the methodology used to arrive at them, as I have presented here.

Another reason that such technological idealisations are valuable is that technology is ever increasing and our ethical theories must be fit to meet the ethical challenges that they will bring. Technological idealisations are a good test for this; they try to show the most extreme forms of technology which a given theory would endorse. If a given idealised theory would endorse the development of technologies which seem morally abhorrent, even to the initial proponents of the theory, then it is a good indication that the theory is not fit to guide our ethical thinking regarding actual future technologically developments. On the other hand, if the most extreme technologies which an idealised theory would endorse do on reflection seem morally permissible (or perhaps even desirable) then this is a good indication that such a theory can be trusted to safely guide our future technological development. For if the most extreme technological endorsements it has to offer are not problematic then we have good reason to believe that those technologies which it will find permissible in the future will also not be.

Towards establishing both these ends, I have found many applicable examples in the philosophical literature surrounding technology and human enhancement which, as well as further explaining the process, show the value of technological idealisations to our ethical thinking. To start, I will look at Agar's "moral beta-blocker" which invokes a particularly interesting technological idealisation of utilitarianism. Nozick's "experience machine" is another technological idealisation, this time of hedonism. "The God machine", which comes from the work of Savulescu and Persson, concerns the sacrifice of autonomy for safety. My discussion regarding "Ectogenesis" mostly draws on the work of Anna Smajdor and idealises sex equality. Finally I discuss "male pregnancy", which, inspired by the work of Robert Sparrow, concerns positive reproductive rights. It is my goal that after the discussion of all these examples of technological idealisations and the diversity of theories and concepts they are applicable to, that I will have convinced the reader that they offer valuable contributions to ethical debate.

3. Utilitarianism and the Moral Beta-Blocker

In this chapter I examine the first technological idealisation of utilitarianism and the moral beta-blocker. I will start this chapter by briefly explaining moral enhancement as the concept is significantly referenced in Agar's paper "Moral Bioenhancement and the Utilitarian Catastrophe", which is the focus of this chapter. The bulk of this chapter will examine a thought experiment presented by Agar which involves a moral beta-blocker, allowing the recipient to act out the full demands of the principle of utility without second thought, which has troubling implications for the utilitarianism, giving one example of the value of technological idealisations in giving us important evidence regarding ethical debates.

3.1 What is moral enhancement?

Let me begin by first introducing moral enhancement. Being a relatively recent subcategory, in a field of applied ethics regarding enhancement (which is itself a relatively recent field of study), moral enhancement warrants an explanation: If "human enhancement" is the umbrella term referring to quantitatively changing any human capacities by artificial means (as I have defined it in the introduction), then moral enhancement refers specifically to quantitatively changing those capacities which are said to be moral. So we see that enhancement is still defined as quantitative change, but that the subcategory of enhancement is determined by the reasonable intention or motivation of the enhancement; in this case a reasonable intention to morally enhance determines moral enhancement. As is the case with all forms of enhancement the method need not necessarily be biomedical; cognitive enhancement could come in the form of a complex computer chip being surgically implanted to interface with the human brain, or it could come in the form of maths classes. Moral enhancement could come in the form of oxytocin to improve the chemical bonding process between human beings, or it could come in the form of sending children to bed without dessert (Savulescu and Persson 2008, 163-165). However, as is the case in all areas of human enhancement, the non-biomedical means are universally evaluated to be good and are therefore only interesting when used as a point of comparison and consistency with current and future biomedical technologies that could be used for the purposes of moral enhancement, which are what everybody finds interesting, either out of expectation or fear.

The idea of moral enhancement is less intuitive then many other forms of enhancement. To return again to the comparison with cognitive enhancement, it is easier to understand how cognitive enhancement could work through biomedical means by improving concentration, memory, I.Q., etc. Morality is often linked with the spiritual which, although fatuous, means the idea of bio-medically improving it is instantly more suspicious for many people. Scientifically it is also less clear if it is possible to link certain process in the brain to morality in the same way as memory or intelligence. Given this difficulty, proponents for moral enhancement such as Thomas Douglas try to focus on a simple method for quantifying the moral capacities of human beings and calculating their enhancement. Douglas attempts to quantify moral improvement by appealing to motivation: "I will take it as a suggestion that we cause ourselves to have, morally better motives. I understand motives to be the psychological - mental or neural - states or processes that will, given the absence of opposing motives, cause a person to act" (Douglas 2008, 229). Douglas here offers a seemingly Kantian description of moral enhancement grounded in motives[11]. The moral person presented here is the sum of their motives, and given this it is clear that they must therefore be morally enhanced by enhancing their motives: "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" (Douglas 2008, 229). If motives are what matter then any process that improves the motives of an individual will morally enhance that individual. This is purely a theoretical description of how moral enhancement might work and how the term "moral enhancement" might make sense. If you object to Douglas' Kantian interpretation, you could replace the idea of future motives with the sum total of virtue and imagine the possibility of biomedically increasing it, or you could do what Julian Savulescu and Ingmar Persson do and focus on the future consequences. The point here is not to endorse Douglas' interpretation but simply to understand the concept and imagine its possibility.

---

[11] Douglas, however, claims to do so for Ecumenical reasons, stating that motive are a common concern to many moral theories even if those theories focus primary on the actions/behaviours derived from those motives.

Savulescu and Persson were the pioneers of the concept of moral enhancement[12]. They take a more consequentialist approach to moral enhancement, appealing to the potential for disastrous consequences if the world continues to rapidly technologically advance without similarly rapid moral advancement. They offer a few examples (among many) of how biomedical technology can be used to manipulate characteristics that are often associated with morality:

- Oxytocin has been shown to promote trust.
- SSRIs to increase co-operation/reduce aggression.
- Ritalin, given to children with Attention Deficit Disorder, reduces violent aggression.
- Antisocial personality disorder may have a biological basis.
- Criminality has been linked to MAO mutation on the X chromosome, especially when coupled with social deprivation (Persson and Savulescu 2008, 172).

I add this list only to show that the concept of moral enhancement through biomedical means is not a fanciful impossibility dreamed up by bored philosophers. Given this list, if there is any link between morality and aggression for example, it is clear that aggression and therefore morality could in some sense be affected (negatively or positively) by biomedical intervention. For those who are still sceptical, let me touch on Douglas work regarding chemical castration (Douglas 2013) (Douglas 2014). Child abuse, in particular sexual child abuse, is almost universally considered morally wrong. Any person who engages in such activity is rightfully condemned as immoral. Given this, we can say with certainty that any therapy or technology which prevented those individuals from engaging in that activity, or prevented them reoffending, would morally improve those individuals or at the very least improve the moral state of affairs in the world by preventing immoral acts. Chemical castration appears to be one such technology. The details of whether it actually works is irrelevant (though Douglas examines some evidence that it does), it is enough to simply stipulate that is does hypothetically and if you are willing to accept that if it were to work the moral situation would be improved then you are willing to accept a technology possibly

---

[12] At least in popularising the concept in enhancement literature, for as they go on to say, the idea of morally enhancing has existed probably as long as human beings have existed.

morally improving an individual or the their situation. Once this is admitted then we can now go on to examine Agar's more detailed example of moral enhancement.

What can be seen from this small discussion is that here is a small divide here between those for are interested in moral enhancement as a practical possibility to solve current ethical problems, Savulescu and Persson, and those interested in moral enhancement for its theoretical contributions to current ethical debates, here represented by Douglas. In keeping with this thesis' theme of ignoring practicality concerns, I place myself in the latter category. I will be going on to discuss the moral enhancement of the moral beta-blocker, not to claim that we should be developing it to solve moral problems, but because it acutely focuses attention on a theoretical problem with utilitarianism in what it demands of its adherents.

3.2 Moral Enhancement and Utilitarianism

One aspect of moral enhancement that is not frequently addressed is what it means for our current ethical theories. Nicholas Agar addresses this concern in his papers "Enhancing Genetic Virtue?", and "Moral Bioenhancement and the Utilitarian Catastrophe". The papers focus mainly (although not always exclusively) on utilitarianism, and what moral enhancement might mean for this well-known moral theory. He worries that moral enhancement will make utilitarianism too realisable, by allowing us to perform acts which utilitarianism would approve of (and actually require) but which would drastically conflict with common sense morality.

The first example which Agar gives comes from "Enhancing Genetic Virtue?" and regards the fictional charter Dexter: "The central character of the ShowTime's series Dexter is a psychopath who resolves to do the morally correct thing. He restricts his killing to evil types who would otherwise escape justice. It's possible that Dexter's actions make the world a happier place. He enjoys killing and we can presume that the individuals who become his victims would otherwise long continue their murdering ways. Is Dexter a morally enhanced being? Utilitarian moral enhancers might say so" (Agar 2010, 74). There are a number of

problems here, the first being that act utilitarianism[13] approves of the actions of a psychopathic murderer because his actions likely produce more overall utility; a troubling result for any moral theory. Of course this is a well-documented flaw of utilitarian theory and not exclusive to moral enhancement criticisms. What moral enhancement does is compound the problem; it adds the additional problem that not only does utilitarianism approve of Dexter's actions but also demands that we enhance others to act more like Dexter. I believe that this is much more devastating a point against the theory, particularly when it comes to using the theory to guide our future ethical thinking. The fact that a theory may approve of anomalous utility enhancing psychopaths is a problem but not a huge problem as few such people are likely to ever exist. However, the idea that a theory demands we purposefully bring such people into existent through the use of biomedical technology is much more troubling.

Agar focuses on an original and more developed example to make this point in his paper "Moral Bioenhancement and the Utilitarian Catastrophe". It regards a fictional Dr Angela, a medical researcher for HIV. Agar references the fact that, "According to the World Health Organization, approximately 1.7 million people died of AIDS-related illnesses in 2011" (Agar 2014, 7). In her research, Angela develops a new treatment for the virus which she calls "therapy X". It is radically different from existing treatments and, based on Angela's vast knowledge on the subject, she expects dramatic improvements in the treatment of HIV/AIDS patients. She is, of course, aware that many experimental therapies fail and must therefore test the treatment; however going through the bureaucratic process will take a very long time. At the cost of approximately 5,000 lives per day lost to HIV/AIDS (according to Agar), Angela hopes for a more expedient testing alternative. Luckily, she is a mother, and could test therapy X on her daughter. In order to test it though, she must first infect her daughter with HIV. The question Agar poses to us is: "Should Angela bring therapy X home and test it on her child?" (Agar 2014, 7).

Here the problem is obvious: Utilitarianism demands Angela harm her child in order to potentially save thousands of lives (depending on how long the alternative bureaucratic

---

[13] I will be focusing solely on Act utilitarianism so as to avoid the complications and word count which comes with trying to reference all variations of utilitarian theory. Beyond this point any reference to utilitarian theory can be read as referring to act utilitarianism.

process would take). It is important to note that it is not necessary that the experiment be successful; even if it fails the abandonment of useless research sooner than abandonment would be achieved through the bureaucratic method will also lead to overall greater utility and alternative and possibly fruitful research into a cure being started sooner[14]. This all clearly conflicts with common sense moral principles, such as refraining from harming the innocent, and parental duties.

As Agar notes, the principle of "ought implies can" allows the utilitarian to escape endorsement of Angela testing on her child. He points to the fact that Angela presumably loves her child; even if she attempted to do the experiment she would probably be unsuccessful and suffer without producing any valuable research. The act will likely be psychologically impossible for her if she has a normal human psychology. This means that the utilitarian can refrain from demanding that Angela test therapy X on her child, as she cannot be required to do what is impossible to do.

However with moral enhancement, what is psychologically impossible now could be very possible in the near future. In the theoretical realm we need not wait for this to happen, but we can imagine a modification to the thought experiment to take moral enhancement into account: imagine that following on from the previous scenario Angela has also developed a moral enhancement therapy which will lower her empathy (or should I say bias?) towards her child and allow her to experiment on her child. Agar calls this the "moral beta-blocker". Now the question becomes; should Angela take the moral beta-blocker? The significant difference in this technologically idealised scenario is that the utilitarian can now longer use the principle of "ought implies can" as a means of escape; the circle of what is now humanly possible has been expanded. The moral beta-blocker will no longer make the experiment psychologically impossible for Angela to undertake and so she can now be required to perform it.

Agar's conclusion from this is not to find fault with utilitarianism but with the project of moral enhancement. He goes on to say: "I think that there is a way of understanding utilitarianism that does not lead to the conclusion that Angela should take the moral beta-blocker and perform the experiment. We should acknowledge utilitarianism as a

---

[14] Even so I could simply restate the example to guarantee the cure. The result would be the same.

distinctively human morality. It gives moral advice to individuals with normal human psychologies. It says nothing about how beings who are emotionally and psychologically very different from us should act. It says nothing about how we should modify our moral psychologies" (Agar 2014, 9). The defence here is twofold. First the claim is that utilitarianism is a distinctly human morality and should not be used to modify our psychologies is a potential defence against the problems that come with moral enhancement. The second, further claim, is that the theory of utilitarianism itself does not claim to give moral advice to individuals with radically different psychologies from ours, or that it says nothing about how we should modify our psychology. In the next section I want to examine these responses in more depth, arguing that they both fail to save utilitarianism from itself.

3.3 The Utilitarian Catastrophe

As part of his defence Agar argues by analogy, comparing ethical theories with exercise programmes: They must be tailored to the kinds of beings that are to undertake them. If an exercise programme required human beings to run thousands of miles each day it would be a bad exercise programme. A personal trainer tailors their prescribed exercise programmes to the beings they are advising. Completing the analogy; a utilitarian tailors their moral prescriptions to the beings they are morally advising. Agar finishes by concluding that we should:

"Accept utilitarianism as a good human morality while rejecting suggestions that the theory become a template for the redesign of our moral psychologies. There is no inconsistency in recommending utilitarianism as a morality for human beings while rejecting the suggestion that human psychology be systematically modified so as to permit outcomes that seem superior in utilitarian terms" (Agar 2014, 10).

Here, a clear distinction is made between morality, which is anthropocentric, and modifying or enhancing our moral psychologies, which is not anthropocentric as in doing so we would become different kinds of beings; in some sense post-human if Agar's distinction is to be successful (as if we were to remain human beings it would seem we could still use the moral theories that Agar claims are fit for human beings). This distinction might be a successful

one; however, the problem is that his chosen theory of utilitarianism, by the nature of the theory itself, doesn't respect this distinction. This can be made clearer by examining the three propositions at play here:

(1) Morality is, or should be, anthropocentric; concerned with moral guidance specifically for human beings.

(2) Modifying the moral psychologies of human beings is not within the scope of morality as it is not anthropocentric.[15]

(3) Utilitarianism; maximise utility by any means necessary.

The disagreement here is that Utilitarianism is a morality which is not anthropocentric and which demands maximisation of utility by any means; which includes the modifying of moral psychologies if such a thing would increase utility on the theory (which Agar's examples show it could).

Utilitarianism is universal in two ways: Firstly, it applies to all moral agents regardless of what kind of beings they happen to be. Secondly, it applies to all actions as they all in some way affect the balance of utility in the universe. Both of these kinds of universality cause problems for Agar's argument. The first kind contradicts his claim that utilitarianism is a specifically human morality. The second kind rejects the idea that utilitarianism should not be used as a guide to modify of our psychologies.

Regarding the first kind of utilitarian universalism, its application to all moral agents; the demands of utilitarianism apply to all moral agents, there is no distinction made between different sentient beings (something often championed in its favour) and it does not claim to be an ethical theory that only applies to human moral agents. It is true that utilitarianism takes inspiration from the human experience in its value of pleasure, happiness, preferences or whatever other kind of utility any given brand of utilitarianism may focus on. It's possible to conceive of beings which do not value or which completely lack these concepts, but this is simply a failure of utilitarianism to have universal appeal and reveals its anthropocentric

---

[15] This is because it concerns changing beings for one kind to another, in this case from human to post-human, and therefore by nature of the principle cannot be concerned exclusively with any one kind of being.

origins. Those beings, if they fulfil the cognitive requirements to be moral agents, are still called to maximise utility.

This outright contradicts Agar's claim that utilitarianism is a specifically human morality. It is just not true that utilitarianism claims to solely hold authority over human beings. Its claims are universal; they apply to humans, post-humans, aliens and even artificial intelligence. A good comparison can be made with Kantian ethics which, despite their many differences, is also universal in the same way; it applies to all rational beings regardless of what other features those beings may have. Good contrasts to this kind of universalism, and theories which actually are specifically human in the way that Agar thinks ethical theories should be are virtue ethics and contract theories. Virtue ethics incorporates a perfectionism specific to human beings, and contract theories utilise facts (or perceived facts) about human nature and human psychologies as axioms. This raises the question of why Agar would choose utilitarianism over one of these other theories, which explicitly fulfil the requirements of being a specifically human morality.

The second kind of utilitarian universality is its application to all actions. As I touched on before, all actions when compared with their alternatives will affect the overall utility created in the universe in some way. Something as insignificant as the colour of the shirt you decide to wear to work on any given day is normative on utilitarianism; the different colour will slightly affect the overall utility derived from individuals seeing you in one colour or another, meaning there is one particular colour that will maximise overall utility and this beings so it is morally required for you to wear that colour. Given such an all-encompassing theory as this, it is not possible to be consistent and propose utilitarian theory to guide moral action for humans but reject the idea that it be used to modify our psychologies. Utilitarian theory does claim authority over our psychologies and over all psychologies, and to finish Agar's example; it demands that Angela take the moral beta-blocker.

This criticism of Agar's work doesn't mean that I think there is nothing of value here. It fact, his thought experiment does a wonderful job of illustrating many of the criticisms of utilitarian theory. In his book "Value Judgement: Improving Our Ethical Beliefs", James Griffin argues the almost exact opposite position to Agar. Instead of claiming that utilitarianism is a human specific morality, he argues that utilitarianism is ill-fitted to guide

human action. In his own words: "Utilitarianism seems to turn ethics into a project that fits badly the agents who are meant to carry it out. The ambitions of utilitarian rationality seem too great" (Griffin 1996, 107). He says this in reference to the above point that the utilitarian scope of moral action is universal, all-encompassing, or "all-sanctioning", to use Griffins' word. The demands of utilitarianism are so great that they are constantly beyond the powers of human beings. To return to the exercise programme analogy, utilitarianism is the exercise programme that demands you run thousands of miles each day. So even if as Agar claims, ethical theories, like exercise programmes, have the purpose of guiding action for the human beings then this only further condemns utilitarianism.

Agar uses rather vague language in his article as to whether he believes that utilitarianism is a specifically human morality or that utilitarianism should be thought of (and perhaps reconstructed to be) a specifically human morality. Until now I have worked under the former assumption but there is also value to be had in discussing the later possibility and so for the sake of completeness I will now do so.

As hinted at before, I think that Agar's paper and in particular his thought experiment Angela the medical researcher do a very good job of illustrating the problems with utilitarian theory. They demonstrate how ill-fitted utilitarianism is to human beings; it is far too demanding, and it would lead to grossly immoral acts if it were strictly adhered to. This is the value of enhancement thought experiments; they allow us to vividly imagine the ultimate implications of our theories without explaining them away with an appeal to "ought implies can". Agar also sees the value in these thought experiments, though in a less ambitious way: "I propose that moral bioenhancement scenarios belong in thought experiments whose purpose is to support or challenge philosophical claims about enhancement" (Agar 2014, 10). He needn't narrow the scope of their usefulness to only philosophical claims about enhancement, but all philosophical claims in general. His own thought experiment challenges the philosophical claims of utilitarianism.

If Agar meant to reconstruct utilitarianism in the light of these problematic revelations of his thought experiment, then I can also add that to the further use of these kinds of thought experiments, however I do have to challenge his attempted reconstruction. Agar recognises that the problems arise for utilitarianism when moral enhancement through the

manipulation of moral psychology is introduced. His solution is simple; to claim that we should not modify our moral psychologies guided by utilitarianism and that we should modify utilitarianism so that it doesn't demand that we do so. The first claim is well established by his thought experiment, but trying to also save utilitarianism is a bridge too far, made of one ad hoc claim; utilitarianism does not require us to modify our psychologies.

It is clear that Agar is aware of his claim being questionable. He anticipates criticism to his claim by allowing some psychological modification to be permissible on utilitarian grounds: "This is not to say that utilitarians should be completely indifferent to the states of moral psychologies. They can allow that forms of empathy training that lead to improved responsiveness to human suffering are good in utilitarian terms. But this awareness of the effects of moral psychological improvement can and should be informed by human limits. It need not extend to the endorsement of artificial means" (Agar 2014, 10). This seems completely inconsistent with what Agar has said in the article up until this point. First he claims the utilitarian theory should not concern itself with changing moral psychology, but now it should but only up until the point which it starts becoming problematic for the theory in leading to abhorrent consequences. This again is ad hoc. He does try to resist this attack with some consistency by appealing to utilitarianism as a specifically human morality which lends support to the claim that it should only be concerned with utility maximising moral psychologies which remain within human limits. The problem is that, as has been previously stated, utilitarianism does not demand we only maximised utility within human limits. It demands we maximise utility. At best it demands we maximise utility as far as possible, however if it becomes possible to go beyond human limitation in maximising utility then we are called to do so. As Agar rightly points out; the defence of "ought implies can" will then no longer apply. The strict Utilitarian then, should find no problem with us all taking the moral beta-blocker. The rest of us, including Agar, recognise the problem such an obligation presents, though we should find the fault to lay with utilitarianism and not the project of moral enhancement.

Of course not all modifications of a theory in the light of new evidence should be dismissed as ad hoc. The fact the Agar's enhancement thought experiment more keenly revealed problems with his preferred theory, and perhaps more importantly, the necessity to solve

those problems, is valuable. However, in trying to solve those problems he offers an ad hoc solution with very little explanation. He needs to give us an accountant of why the utilitarian manipulation of moral psychology is permissible within human limits but impermissible outside those limits? His argument as stands appears to be: morality and therefore utilitarianism should be informed by human limitations. When pressed on the point and asked why this should be the case the reply appears to be: because it is problematic for utilitarianism if it is not.

3.4 Conclusion

What can be seen here is that technological idealisation is working as a kind of '*reductio ad absurdum*', however rather than the common method of reducing a theory to its absurd elements; the technological idealisation here builds up the theory to its ridiculous conclusions. A *reductio* style argument attempts to demonstrate that a statement is true by showing an absurd result follows from its denial or attempting to show a statement is false by showing an absurd result follows from its acceptance (Rescher 2015). What is meant by "absurd" could be interpreted in a number of ways, but can be placed in three broad categories:

(1) Self-Contradiction: A bird is a vertebrate animal that flies. A kiwi is a flightless bird. Therefore a bird is not a vertebrate animal that flies.

(2) Falsehood: Young earth creationism claims the earth is between 5700-10000 years old. The earth is approximately 4.54 billion years old. Therefore, young earth creationism is false.

(3) Implausibility/Anomaly: If abortion is wrong because it prevents life, then all forms of contraception are also wrong. It is implausible that all forms of contraception are wrong, therefore it is implausible that abortion is wrong because it prevents life.

Form (1) and (2) are straight-forward examples of a statement simply being false and so do not require any reinterpretation. Form (3) is more interesting as it doesn't rule out the statement or argument as false, but simply highlights a potential problem with the statement or argument and simply requires reconsideration or revision of the statement or argument in question. Technological idealisation in this case most closely resembles form

(3) as it only aims to show the absurd results that can occur from technologically idealising (through thought experiments) different moral theories, or theories of value.

Like the *reductio*, even if correct the theory is not disproven by such an argument; one could bravely (or foolhardily depending on your perspective) bite the bullet and accept the absurdities of one's theories. However, also like the *reductio*, technological idealisation is evidence which should serve as a warning to those who are not already lost in devotion to a particular theory. In this case the warning is aimed at the theory of utilitarianism, acutely presenting the horror of what can be done in the name of maximising utility. In the case of the moral beta-blocker we see a complete rejection that there might be special obligations derived from relationships between individuals, between mother and daughter, and we see the treatment of people as mere means. Again, this does not disprove the theory but it does provide strong evidence against it, in a more understandable way than just saying "utilitarianism treats people as mere means" for example. This process of technological idealisation really brings the theoretical issues of the theory to the forefront, in a way that I expect to even have utilitarians worried.

The next couple of chapters will follow this same process of *reductio* style technological idealisations, though I want to stress that this is not sole use of technological idealisations, it is simply the negative aspect (examples of the positive aspect will follow). However, I believe that it is valuable to give multiple examples of how technological idealisation work in this fashion; I think these characters are interesting in their own self-contained way, though multiple examples will also help clarify any confusion that may still exist and will serve as a proof of the vast utility of technological idealisations, not being just a criticism of utilitarianism.

4. Hedonism and the Experience Machine

With the experience machine the work is mostly done for me; Nozick has exactly technologically idealised the hedonistic theory of value or well-being. Here I feel my job is to present Nozick's account and defend it from criticism. I will conceive Nozick as arguing against ethical hedonism[16] as he is concerned with the question of whether we 'should' enter the experience machine. Ethical hedonism claims: "All and only pleasure has positive importance and all and only pain or displeasure has negative importance" (Moore 2013). The importance referred to here can be translated as non-instrumental value; pleasure is the only good in and of itself and pain is the only bad in and of itself. I will not go into the issue of whether hedonism here is universalist, restricted, or egoist, as it only change how many people should get into how many machines, but in all cases the decision as to whether the individual should get into the machine stands. I along with Nozick claim that the technological idealisation of the experience machine shows that hedonism is problematic as it would have us abandon reality in the pursuit of pleasure.

4.1 The Experience Machine

Robert Nozick's "experience machine" is an example of taking a theory, in this case simple hedonism, to its logical conclusion. The thought experiment is quite short but to the point and has garnered a lot of discussion. Nozick presents it in the form of a question to his readers:

"Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprograming your life's experiences?" (Nozick 1974, 42).

What is presented here is the opportunity to have the perfect hedonistic life; you can programme the machine to give you the experiences which will give you the most pleasure.

---

[16] As opposed to psychological egoism, the thesis that pleasure and pain determine how we act, as this to me is more a scientific claim.

To use Mill's famous dichotomy of pleasures (Mill 1907, p.10-16), the description does not limit this to lower physical pleasures but explicitly references higher intellectual pleasures of writing high literature and a fulfilling social life. This anticipates the criticism that Nozick might be mischaracterising hedonism as just brute physical pleasure. He also anticipates the criticism that entering the experience machine would be a painful and difficult action to take:

"Nor should you refrain because of the few moments of distress between the moment you've decided and the moment you're plugged. What's a few moments of distress compared to a lifetime of bliss (if that's what you choose), and why feel any distress at all if your decision is the best one?" (Nozick 1974, 43).

Of course the last line here is a sarcastic quip before Nozick launches into his criticism of the thought experiment, however, other than that we should see that overall pleasure is served by getting into the experience machine despite the passing pain of the initial action. To be absolutely clear I will go further than Nozick specifically does and offer the experience machine as the perfect possible life in hedonistic terms. Certainly it is presented as better than our current lives, but, so as to halt any possibility of questioning the quality of life and giving the machine the ultimate benefit of the doubt, I will say it is perfect[17].

If we are good hedonists then we should surely jump at the chance to get into the experience machine. Nozick gives three main reasons why we should not enter the experience machine, in other words he gives three reasons why we should not endorse a hedonistic theory of value. "First we want to do certain things, and not just have the experience of doing them" (Nozick 1974, 43). The experience machine and hedonism can only offer us the phenomenological experience of writing a great book but there is more value to the experience than that; there is the value of actually having written the book. If the phenomenological experience is the same in the machine and in reality than hedonism offer no preference, and in the case that it is more pleasurable inside the machine the

---

[17] Presumably the perfect possible life in hedonistic terms would be infinite in duration but those who enter the experience machine will evitable die but let's leave that aside. All Nozick needs for his thought experiment to work is for life in the programmed machine to be predictably superior in hedonistic terms than the life lived outside the machine.

machine is the superior option. Articulating exactly why this is a problem is difficult but we might say that there is non-phenomenological value of being connected to reality (as we should allow the experience machine to give the phenomenological illusion of being connected to reality) or as Nozick suggests we may have some innate desire to experience reality[18]. So the first criticism of hedonism is that it gives no value to reality.

The other two criticisms Nozick offers are simply variations of the theme that Hedonism places no value on reality. They are as follows:

"A second reason for not plugging in is that we want to be a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank… Plugging into the machine is a kind of suicide" (Nozick 1974, 43).

"Thirdly, plugging into an experience machine limits us to a man-made reality, to a world no deeper or more important than that which people can construct. There is no actual contact with any deeper reality, though the experience of it can be simulated" (Nozick 1974, 43).

The third criticism is simply more clearly restating the overall criticism of the experience machine and hedonism being unconcerned with reality. The second criticism adds a little more substance to the overall critic, claiming that we are not only concerned with experiencing reality but of being a real person. The person in the experience machine is not a real person in anything other than the physical sense; they have no personality or even any real history other than just being in the machine. In the same way the criticism applies to hedonism which sees persons as the sum of their pleasurable experiences. There is more to (some) people than that and hedonism misses this important aspect of life.

What Nozick has done with this thought experiment is use (unwittingly) the exact methodology of technological idealisation; he has taken the hedonistic theory of value and technologically idealised it to its logical conclusion, and revealed the absurdities underlying the theory in the process. The absurdity here happened to be that hedonism does not value

---

[18] Though, to return to the question of psychological hedonism, some research seems to suggest this is simply a status quo bias; when people are told they are already in the experience machine and are given the decision to "unplug", most decide to stay in their current condition. (De Brigard 2010)

reality at all and that exiting reality in the pursuit of phenomenological pleasurable experience is a perfectly acceptable, and in some cases recommended, course of action.

4.2 Conclusion

Clearly there is a great deal of the debate which I have not been able to cover here. As I previously mentioned this particular thought experiment has generated a lot of discussion which I would take me far off course if I was to attempt to cover all of it. Here I restrict my interest in the experience machine as an instance of technological idealisation. The conclusion here following from Nozick is that the technological idealisation of the experience machine pinpoints a problem in hedonism; its lack of regard for reality. Hedonism only values pleasure (again, sidestepping the debate as to what that is) and as pleasure is a phenomenological process in the brain, it can be stimulated by simulated experiences (referring to experiences not connect with reality). Such a thing is easily said, but through technological idealisation thought experiment it is more richly understood.

5. Autonomy and the God Machine

This chapter explores the idea of sacrificing autonomy for safety, perhaps more of a concept than a fully realised theory though I believe that the results are interesting all the same. Though it is just a concept, that classic tension between the value (or right) of letting rational people live their life as they choose and the inevitable conflict (often violent) of those choices between people and mediation by a third party, often the state, for the benefit of all, its expression can be found throughout the history of philosophy. Perhaps the two most famous examples would be Thomas Hobbes on the authoritarian, valuing safety over autonomy, side of the debate and John Stuart Mill on the libertarian, valuing autonomy of safety, side. Mill's quote "That the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant" (Mill 2011), sums up his view on the matter quite nicely, whereas Hobbes' axiom "The end of obedience is protection" (Hobbes 1946, 145), essentially saying that safety is always more important than autonomy, bluntly gives the opposite view. Of course each of the views that these men have articulated to us have more nuisance and scope the just the conflict between autonomy and safety, but as far as they endorse that principle the argument I present here will affect them as much. In this chapter I examine the more Hobbesian approach by technologically idealising the sacrifice of autonomy for safety. The idealisation comes in the form of Savulescu and Persson's "God Machine", which represents the greatest extreme of sacrificing all meaningful autonomy for perfect safety. I argue that, after examination, such a concept is not fit to guide of future ethical thinking.

5.1 The God Machine

The God Machine thought experiment appears in the writings of Savulescu and Persson as a challenge to the idea that freedom or autonomy is more valuable than a morally better society; if we had to give up autonomy to make ourselves morally better then should we not? Savulescu and Persson offer a short summary of their thought experiment:

"The Great Moral Project was completed in 2045. This involved construction of the most powerful, self-learning, self-developing bioquantum computer ever constructed called the

God Machine. The God Machine would monitor the thoughts, beliefs, desires and intentions of every human being. It was capable of modifying these within nanoseconds, without the conscious recognition by any human subjects" (Savulescu and Persson 2012, 10).

The important aspects of the description are that the God machine monitors everyone's thoughts, beliefs, desires and intentions; is capable of changing them; and this change occurs without the conscious recognition of the subjects. How exactly this is supposed to take place is not important as I am not concerned with the practical possibility but the theoretical implications. They go on to explain that the God Machine only intervenes in the cases of serious crimes such as murder or rape, presumably allowing petty crimes and misdemeanours to be solved the old fashioned way. The important difference from our current state of affairs, that Savulescu and Persson mention, is that the God machine changes things from begin prohibited, to being impossible:

"Human beings can still autonomously choose to be moral, since if they choose the moral action, the God Machine will not intervene. Indeed, they are free to be moral. They are only unfree to do grossly immoral acts, like killing or raping. This is seen as preferable to physical incarceration, which physically restricts the freedom of the immoral. While people weren't free to act immorally in the 'old days,' since the law prohibited it on pain of punishment, the instalment of the God Machine means that it has become literally impossible to do these things" (Savulescu and Persson 2012, 11).

I propose that we ignore this needless restriction that is places on the God Machine, to only intervene in seriously immoral (or illegal) acts. The God Machine should intervene in all immoral acts for the sake of consistency as they vary only in degrees; if the God Machine is justified to intervene in the case of murder, it is justified to intervene in the case of petty theft. Like the governmental law enforcement it is seemingly replacing, it would be very strange (and worthy of compliant) if it only intervened in serious crimes and let all minor crimes (that it has the power to stop) continue.

To return to Savulescu and Persson's writing, they explain why the God Machine would be justified and desirable, anticipating concern over the infringement of autonomy:

"Freedom, however, is only one value. In the world of the God Machine, there would be no serious crime. There would be great benefits to other people. In the absence of perfectly effective moral enhancement, the loss of freedom in one domain of our lives - to commit evil deeds - would be worth the benefits. We would be otherwise free. Even in those cases in which the God Machine does undermine autonomy, the value of human well-being and respect for the most basic rights outweighs the value of autonomy" (Savulescu and Persson 2012, 13).

The sacrifice of autonomy, freedom and rights is made for well-being and safety. In the World of the God Machine the people have decided that the loss of autonomy is more than compensated for by the safety they receive from the potential injury by other people. However, as we have seen, Savulescu and Persson claim that autonomy still exists in some from in the world of the God Machine; human beings are "free" to choose to act in accordance with the will of the God Machine, they are merely not free to act against it and commit gravely immoral acts.

Despite this attempt at a defence of some minor autonomy the God Machine has still received much criticism for its sacrifice of autonomy for safety; I'll be focusing here on John Harris and Robert Sparrow.

As we have seen, there are substantial issues of liberty which would also need to be resolved and which could conceivably be threatened by any measures that make the freedom to do immoral things impossible, rather than simply making the doing of them wrong and giving us moral, legal and prudential reasons to refrain. (Harris 2011, 105)

"Even though the God Machine only acts to alter an individual's motivations when he or she intends to commit a seriously immoral action, the techniques it uses to do so could also be used to control individuals'' motivations more generally. The God machine 'dominates' its subjects. Thus it is not solely the case that people subject to the power of the God machine are not 'free to fall'. In removing the freedom to fall, the God machine removes their freedom altogether" (Sparrow 2014, 27).

Here we see a complete rejection of Savulescu and Persson's defence that you can still exercise some freedom in the world of the God Machine, for them, to take away the possibility of immoral action makes any moral action meaningless. I think it is a difficult question as to whether I agree with Sparrow and Harris that this is enough to show that the God Machine is overall a bad technology, as it is still possible to fall back on the point that "autonomy" is just one value amongst many and must be weighed against the fact that there are no seriously immoral acts in the world; this tension resembles the oft debated point in introductory classes to the philosophy of religion of whether God should have given us free will despite the evil it would inevitably bring? It is often said in such a class that better not to have freewill and be rid of the suffering we have on earth. However, I think I must uncharacteristically agree with the apologists on this point and say that God (whether in machine form or not) ought to give us freewill. I believe Harris and Sparrow are right that the God Machine renders individual moral actions meaningless; however their criticism goes further than that we simply can't praise people who are connected to the God Machine, but that we cannot acknowledge that they have any agency whatsoever. Their actions are the actions of the God Machine, as the original front piece of the leviathan depicts a giant King made out of the body of other men, the God Machine is the agent of all of society's actions and the people become part of the Machine.

It is interesting to consider whether the original presentation of the God machine which only intervenes in the case of serious moral wrong-doing would invalidate our freedom in the same way as the all controlling God machine which stops all wrong-doing as I have presented it. Savulescu and Persson themselves point out that: "We are not free to commit serious crime even now - the laws prohibits it on pain of punishment" (Savulescu and Persson 2012, 13). The God machine simply turns this prohibition into impossibility, and it might be argued that this is in keeping with the intent of laws of prohibition as well; it seems reasonable to claim that the intent of law prohibiting murder is to make murder impossible. Despite this, I believe the stronger argument is to claim that even this weaker God machine greatly infringes on autonomy. The key aspect is that it removes choice from action. Laws of prohibition can also be interpreted as not intending to infringe on the free choice of its citizens, respecting them as rational agents, but informing them that if they chose to take actions which are harmful to society and do not respect the free choice of others then they

will be rightfully punished. On this interpretation the God machine is a completely different entity to current justice systems, one which like the previous interpretation still dominates its subjects. I conclude form this that even the original interpretation of the God machine fares poorly when it comes to the sacrifice of autonomy.

5.2 The God Machine and the Sovereign

I claim that the God Machine bears many similarities to Hobbes' Sovereign[19]; it is created in the same way and for the same reason, by sacrificing autonomy for safety; and wields the same absolute authority over its subjects. I want to show here that by comparing the two thought experiments, and the conclusions I have drawn regarding the God machine, we might also draw similar conclusions regarding Hobbes' Sovereign (insofar as the similarity holds). I believe this will serve as further evidence of this particular technological idealisations value to existing ethical (and political) debates, and by extension the value of the method of technological idealisation in general.

To first address the point regarding the Sovereign and the God Machine being created in the same way for the same reason, compare these two statements by Hobbes, and Savulescu and Persson, respectively:

"The second law of nature: that a man be willing, when others are so too, as far-forth, as for peace, and defence of himself he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himself" (Hobbes 1946, 85).

"Freedom, however, is only one value. In the world of the God Machine, there would be no serious crime. There would be great benefits to other people. In the absence of perfectly effective moral enhancement, the loss of freedom in one domain of our lives - to commit evil deeds - would be worth the benefits. We would be otherwise free. Even in those cases in which the God Machine does undermine autonomy, the value of human well-being and respect for the most basic rights outweighs the value of autonomy" (Savulescu and Persson 2012, 13).

---

[19] Whether this is intentional or not I cannot tell, I can only say that there was no mention of it in Savulescu and Persson's paper.

Here, wording is the only difference between these two statements; "liberty" or "freedom" is sacrificed for "peace" or "well-being and respect for the most basic rights". This clearly demonstrates that the Sovereign and God Machine are created in the same way by sacrificing autonomy, called liberty or freedom, for safety, called peace or well-being and respect for rights.

The further claim, that the Sovereign and the God Machine wield the same absolute authority over their subjects, is not quite so easy to establish. At first glance there appears to be a vast difference between the authority the two have, the Sovereign being restricted to after the fact punishment and the God Machine intervening in the mind of its subjects and preventing the crime, which would require punishment, before it happens. This is uninterestingly true in the descriptive sense; this is clearly how the two are described to us. The interesting question is whether or not this difference is simply a contingent one due to Hobbes' (I refer to Hobbes here, and in this entire section, as the Hobbes of the Leviathan, not the historical man himself. The man may have had many personal misgivings regarding the God machine and here I am just discussing what he wrote in just one of his books) excusable lack of imagination given the historical context he lived in? I claim that it is contingent, and that these two thought experiments, the Sovereign and the God machine, are indeed very similar.

One might appeal to the text of the Leviathan in order to push back against my claim. It is true that there are places in the text which appear to contradict what I am claiming. For example, examine the following quote:

"But as men, for the attaining of peace, and conservation of themselves thereby, have made an Artificial Man, which we call a Common-wealth; so also have they made Artificial Chains, called Civil Laws, which they themselves, by mutual covenants, have fastened at one end, to the lips of that Man, or Assembly, to whom they have given the Sovereign Power; and at the other end to their own Ears. These Bonds in their own nature but weak, may nevertheless be made to hold, by the danger, though not by the difficulty of breaking them" (Hobbes 1946, 138).

The last sentence in particular seems to explicitly state that the Sovereign acts differently to the God Machine; the bonds are held by the danger of breaking them in the case of the Sovereign which is to say the risk of severe punishment for transgressors, and are not held by the difficulty of breaking them as they are in the case of the God machine. However, on closer examination it becomes clear from the language that Hobbes uses that this is simply a descriptive claim. He recognises that the bonds that hold people to keep to their covenants are "in their own nature weak" as it relies on the character of human beings which are selfishly motivated. He follows this by stating the bonds "may nevertheless be made to hold, by danger", which is just an appeal to practicality. The "danger" of punishment is just a practical way to motivate subjects to remain lawfully, there is no implication that the danger of punishment must be used or even should be used if there is a more efficient method, it is just a tool; and like all tool it can be disposed of once a superior one is discovered.

In order to properly compare the Sovereign and the God Machine, One must look to the normative statements of the Leviathan concerning the Sovereign's authority. Here is a famous quote regarding the limits of the Sovereign's power; "The obligation of subjects to the Sovereign is understood to last as long, and no longer, than the power lasteth, by which he is able to protect them … The end of obedience is protection" (Hobbes 1946, 144-145). The only limit to the Sovereign power is the protection which it provides. So the Sovereign as it is described by Hobbes, which can only attempt to prevent transgression through instilling fear of punishment, is horribly imperfect by his own standards. The fact that punishment exists and has to be used, evidenced by the incalculable amount of crimes committed everyday (many which occur without punishment), means that the Sovereign as Hobbes presents it (being similar to our current governments at least in the aspect of deterrence and punishment) is constantly failing to protect its subjects. To me, this motivates the idea that Hobbes' Sovereign is simply a guide, open to future improvement and which was limited by the times in which it was created. The Sovereign is justified in all it does to protect its subjects, but the Leviathan was a work of its own age, with all the restrictions that entails.

Of course, it is possible to accept that Hobbes intended the Sovereign to be improved, or that were he still alive would approve of modification to his original theory, while still being

sceptical of the idea that the God machine embodies those improvements and is consistent with the work in the Leviathan. It is reasonable to be sceptical of the idea that Hobbes would approve of an artificial intelligence which would invade and manipulate the minds of its subjects. In order to alleviate that scepticism I return to the text of the Leviathan:

"it is annexed to the Sovereignty, to be judge of what opinions and doctrines are averse, and what conducing to peace; and consequently, on what occasions, how far, and what, men are to be trusted withal, in speaking to multitudes of people; and who shall examine the doctrines of all books before they be published. For the actions of men proceed from their opinions; and in the well-governing of opinions, consisteth the well governing of men's actions, in order to their peace, and concord… It belongeth therefore to him that hath the Sovereign power, to be judge, or constitute all judges of opinions and doctrines, as a thing necessary to peace, thereby to prevent discord and civil war" (Hobbes 1946, 116-17).

This is a clear endorsement of indoctrination, in the only way in which it was feasible during the 17th century. The Sovereign has control over all information, all books that are published and what is spoken of in public. This line in particular, "For the actions of men proceed from their opinions; and in the well-governing of opinions, consisteth the well-governing of men's actions" is very similar to the way which the God machine operates. The God machine controls (or "governs") the "opinions" (including beliefs, desire, and intentions) of its subjects in order to control the actions of its subjects. The motivation is, again, also the same; "their peace and concord". My aim here is only to show the similarity between these two thought experiments in this aspect of mental control. Clearly, Hobbes could not have intended this exact extension of his idea in the form of a machine that can literally govern the opinions of its subjects, however in reading this passage it is also clear that the God machine in this aspect is relevantly similar to what Hobbes' wrote here.

Fittingly, there are also similarities to be found in the text in the Leviathan and the God Machine when Hobbes discussed God and liberty. He attempts to resolve the apparent inconsistency between human free will and God's omnipotence:

"Liberty and necessity are consistent: As in the water, that hath not only liberty, but a Necessity of descending by the channel: so likewise in the actions which men voluntarily do;

which (because they proceed from their will) proceed from liberty; and yet because every act of mans will, and every desire, and inclination proceeds from some cause, which causes in a continual chain (whose first link is the hand of God the first of all causes) proceed from necessity. So that to him that could see the connexion of those causes, the necessity of all men's voluntary actions, would appear manifest. And therefore God, that see, and dispose all things, see also that the liberty of man in doing what he will, is accompanied with the necessity of doing that which God will, & no more, nor less. For though men may do many things, which God does not command, nor is therefore author of them; yet they can have no passion, nor appetite to anything, of which appetite God's will is not the cause. And did not his will assure the necessity of mans will, and consequently of all that on mans will dependents, the liberty of men would be a contradiction, and impediment to the omnipotence and liberty of God. And this shall suffice, (as to the matter in hand) of that natural liberty, which only is properly called liberty" (Hobbes 1946, 37-138).

This is an example of Hobbes compatibilism. God, being the first and omniscient cause, has caused all things to happen in accordance with his will. People are free to act in accordance with God's will but no one can act against it. This is strikingly similar to the way Savulescu and Persson talk about the God machine and freedom; people are free to act in accordance with the God machine's will but are prevented from ever acting against it. As Hobbes believes we can still be free under God, Savulescu and Persson believe we can still be free under the God machine.

5.3 Conclusion

What can be concluded from this is that the technological idealisation of sacrificing autonomy for safety fails the test. Unlike the previous examples which focused on a particular theory, here we have an example which is merely a concept, though one with wide scope within the ethical and political literature. Given that it is not a theory the most that can be taken from this is that any theory which endorses the sacrifice of autonomy for liberty, as I have suggested that the Hobbes of the Leviathan does, will be negatively affected by this. There is much more that could be said about the God machine and I particularly think a lot more about the comparison with the God Machine, but here I am not attempting to debate the methods of moral enhancement, or exercise Hobbes scholarship

(for which I am not qualified). Again, I am concerned with this thought experiment as a technological idealisation for the ideal of autonomy and its sacrifice for safety, with the comparison with Hobbes' sovereign further showing the contribution of this idealisation to ethical and political debate.

6. Sex Equality and Ectogenesis

My examples up to this point, regarding technological idealisations, have been negative; this is to say that they have acted in a similar fashion to a *reductio ad absurdum* by revealing absurdities in theories and using them as an argument against including them in our future ethical thinking. This chapter will contain my first positive example (with another to follow) as to how technological idealisation can pick out theories or principles which pass the test and can be included in our ethical thinking regarding future problems. My example here is Sex equality, which I claim pasts the tests despite the fact that it may commit us to endorsing what some might consider radical technology.

Sex equality is almost universally accepted as moral good worth striving towards, amongst those whose opinions are worth considering[20]. Most would likely say, if asked, that attaining true sex equality is an obligation. Towards this goal much work and progress has been made (in the developed world at least) through social and political actions; changing attitudes and laws. Despite this, what I want to claim in this section is that not all the wrongs of sex inequality can be righted through purely social and political actions. The natural inequalities of biology will still remain, even if society and law is changed to completely remove any sex bias. What I am not claiming is that we have reached the stage of a perfect gender neutral society and politic, or trying to discourage the continued efforts of those working towards attaining it. What I am claiming is that gender inequality is not entirely socially constructed and therefore cannot be entirely solved socially. One such non-social, biological, inequality is child gestation and birth. In this chapter I want to focus of this particular technological idealisation and show that while the unequal burden of child gestation and birth is currently an unsolvable biological inequality, through technological idealisation there are potential biomedical solutions which obligate research if we are committed to the principle of sex equality[21].

---

[20] I use the term "sex equality" in this chapter in place of the more common "gender equality" as a reflection of my focus on biology. I am concerned here with the sexual dimorphism and how that impacts equality of those individuals who have different biology. I leave the question of gender open and separate, accepting that individuals may not gender identify with their assigned sex at birth but that they might still be affected by it.

[21] I assume in this chapter that it is desirable to retain sexual dimorphism, as it is otherwise obvious that the only true path to equality is to remove sexual dimorphism all together and have there be only one sex. Wanting to avoid the added issues of the debate as to whether sexual dimorphism is preferable I will assume it

6.1 The Unresolved Inequalities of Biology

The technology which could right the inequity of child gestation and birth, as being solely the burden of women, has come to be known as ectogenesis; artificial wombs. Anna Smajdor gives an extended treatment of the issue:

"Rather than putting the onus on women to have children at times that suit societal rather than women's individual interests, we could provide technical alternatives to gestation and childbirth so that women are no longer unjustly obliged to be the sole risk takers in reproductive enterprises. In short, what is required is ectogenesis: the development of artificial wombs that can sustain foetuses to term without the need for women's bodies. Only by thus remedying the natural or physical injustices involved in the unequal gender roles of reproduction can we alleviate the social injustices that arise from them" (Smajdor 2007, 336-337).

It is important to note here that Smajdor also recognises the injustices that come from the biological differences concerning sex which can only be fully equalised by the biomedical technology of ectogenesis (or perhaps some similar technology). Smajdor is also more specific regarding the inequality itself:

"When a man wishes to have offspring, he is able to do so without risking his bodily integrity, his health, or his privacy. Thus, in terms of personal resource holdings, women are systematically at a disadvantage. In the context of reproduction… the choices and risks open to them [men and women] are vastly unequal" (Smajdor 2007, 341).

This is only a small list of the burdens faced by women in order to reproduce over men, and importantly ones which cannot be fully resolve by social actions.

However, one might contest the idea that there are unequal burdens between men and women for child birth, or the perspective that they should be considered negative for the women rather than a positive experience which men are disadvantaged not to have access

---

is as it's more familiar and accessible ground for debate. Additionally, the principle of sex equality would seem to lose all meaning if there were no longer two (or more) sexes which could either be unequal or equal.

to[22]. To address the first response, let me look more closely at Smajdor work in which she more thoroughly outlines the burdens women face in order to reproduce. First, let's consider the purely physical burdens:

"Pregnant women are likely to suffer health problems including back pain, exhaustion, bowel problems, and urinary incontinence extending for 6 months after delivery and beyond. The prevalence in particular of fecal incontinence following childbirth is something that has only just begun to be recognized, and it has been suggested that for this reason alone, "natural" birth should be something for which women give informed consent based on a full understanding of these risks. Morbidity associated with childbirth has been systematically neglected" (Smajdor 2007, 340).

These are clear all purely negative physical burdens. It is hard to imagine how some might interrupt these as positive experience in any way. Importantly, they are also exclusive to women, meaning that a man is not exposed to these burdens if he decides to reproduce. We should also not ignore the psychological effects of child birth: "Postpartum depression (which affects 13% of women who have given birth) may cause the mother to reject her child or to refuse to nurture it" (Smajdor 2007, 342). This is harmful to the mother, as well as to the child but here given that I am concerned about gender inequality the harm to the child is irrelevant, the harm to the mother being the important differentiator between men and women in childbirth. Additionally, "A difficult labor increases the chances of posttraumatic stress syndrome" (Smajdor 2007, 340), which I could only see as a negative possibility. Given all of this I think it is clear that the physical and psychological burdens of child gestation and child birth are overwhelmingly negative. I think that Shulamith Firestone (also referenced by Smajdor) describes it most poignantly:

"Pregnancy is barbaric. I do not believe, as many women are now saying, that the reason pregnancy is viewed as not beautiful is due strictly to cultural perversion. The child's first response, 'what's wrong with that fat lady?'; the husband's guilty waning of sexual desire; the women's tears in front of the mirror at eight months - are all gut reactions, not to be dismissed as cultural habits. Pregnancy is the temporary deformation of the body of the individual for the sake of the species. Moreover, childbirth hurts. And it isn't good for you.

---

[22] A point I will also explore the other side of in the following chapter.

Three thousand years ago, women giving birth 'naturally' had no need to pretend that pregnancy was a real trip, some mystical orgasm (that far-away look). The Bible said it: pain and travail" (Firestone 1972, 188-189).

I think we can say that Firestone was right in her belief, and that the only reason that one does not often here such beliefs expressed from women and mothers is how society would respond. That they would be betraying their gender, betraying the role they have been told is right and a beautiful thing. Something they should be happy about it and that to complain about it would be rude, unappreciative of the gift of child gestation and childbirth. I think on any close examination we can say that women have much to complain about for being given this "gift". Interestingly, despite the time when she was writing, Firestone also talked about inventing artificial means of reproduction in order to achieve the liberation of women "the tyranny of reproduction", saying "there are the more distant solutions based on the potentials of modern embryology, that is, artificial reproduction, possibilities still so frightening that they are seldom discussed seriously. We have seen that the fear is to some extent justified: in the hands of our current society and under the direction of current scientists (few of whom are female or even feminist), any attempted use of technology to 'free' anybody is suspect. But we are speculating about post-revolutionary systems, and for the purposes of our discussion we shall assume flexibility and good intentions in those working out the change" (Firestone 1972, 193-194).

Continuing with the theme of societal attitudes and society in general, consider the economic burdens of female reproduction. Smajdor rightly notes that the incontinence and back pain that was previously mentioned restrict the employment opportunity of mothers. Of course all the other previously mentioned burdens will affect the employment opportunity of mothers. The psychological burdens will also make it more difficult to work.

Finally, the mere fact that women must gestate and give birth in order to reproduce means that they are innately less valuable employees that men; let us imagine that we have a man and a women who apply for a job. They have the exact same level of competency, work experience, etc. They are identical in every single way other than their biology. However, the fact that they have different biology when it comes to reproduction means that the rational choice, from a capitalistic perspective, for the prospective employer will always be

the male candidate. He will not have to take time off work in order to reproduce, he will not be at risk of the negative burdens of child gestation and birth, and he will not suffer the physical and psychology problems which could affect his productivity. Even worse, this gender imbalance applies even if a woman has no intention of having children. The mere possibility of this happening or her changing her mind regarding reproduction at some future stage of her employment means that the employer must still prefer the male candidate. Evidence of the fact that this is something which companies consider is some of the recent publicised attempts by some tech companies, such as Facebook and Apple, to pay for women to have their eggs frozen so to encourage women to reproduce later, if at all, during their time with their career with the company (BBC 2014) A more cynical assessment might claim that this is merely an attempt to make hiring women a less financial risky in a sector that has been criticised for its lack of female representation. One could attempt to solve this problem by having specific contracts available to women to sign in order to guarantee that they will not reproduce during the period of their employment, making them no less attractive then a male candidate. Although this might be a good stopgap until ectogenesis technology is developed, it further reinforces the inequality in biology which I am discussing. A woman must essentially give us her reproductive autonomy to be economically competitive, but a man need not.

This leads nicely to the final point regarding the negatives of female reproduction; the loss of autonomy. Smajdor is keenly aware of this point also:

"For expectant mothers, the fact of encompassing another life in their bodies often takes a serious toll on their autonomy. Pregnant women are routinely expected to subsume their appetites and desires into those that would be in keeping with the well-being of the fetus. Not only this, but their abilities and rights to make decisions about their medical care are at risk of being overridden in favor of the interests of the unborn child. Respect for one's bodily integrity, something that most men may take for granted at least in a medical setting, is by no means assured for women even in societies that pride themselves on concern for ethics and autonomy. Women are still sterilized against their will and undergo forced abortions and forced caesareans" (Smajdor 2007, 340-341).

I believe this is the most important issue regarding women biology and reproduction. The fact that women must gestate the foetus in their body raises a whole host of autonomy issues that no man need ever face. The issues can be somewhat trivial, such as disapproval of pregnant women smoking or drinking alcohol to complete subjugation of the woman's life to that of the foetus. These issues come up in the context in which society cares for the life of the foetus to an equal or greater extent than the mother's. Smajdor also mentions the reverse scenario is which the death of the foetus is desired. In this case, the foetus being located inside the women's body means that again the autonomy of the mother will be violated. Having ones autonomy potentially violated to the point of having your life taken for the well-being of another being in the most extreme cases, is a terrible burden and importantly once again one that men do not have to suffer in order to reproduce.

Having firmly established the negative burdens of child gestation and birth, there is still one strategy to dismiss them that I must discuss before going on to talk about the solution in the form of ectogenesis. The strategy is to claim that pregnancy is something which women voluntarily undertake and therefore, even though there are recognised inequalities between men in women in reproduction they do not require remedying as they are not forced upon women but undertaken through personal choice. Let me leave aside the obvious reply that not all pregnancies are voluntary, and I will generously assume that any advocate of this position would allow for termination in the case of accidental pregnancy and pregnancy resulting from rape in order to make their argument tenable. First, it is easy to question how voluntary female reproduction is when it is currently the only form of reproduction and therefore integral to any sustainable society. Secondly, the inequality still exists in the risks undertaken to voluntarily reproduce, something which many people consider a fundamental life interest, between the burdens faced by women over men. Smajdor compares the situation to bungee jumping:

In my argument, pregnancy is the bungee jump. Granted, women do often voluntarily make the choice to jump and thus assume the risks involved. However, the prima facie injustice involved lies in the fact that when men decide to jump-or to have a genetically related child- they are able to do so without assuming any of these risks that affect women in similar situations. (Smajdor 2007, 341)

To take the analogy further, we would think it wildly unjust for bungee jump operators to offer full protective gear and additional safety ropes to male customers but not female customers, even though it is the case that they undertake the activity voluntarily. Given that such a case would be labelled as unjust and sexist in a case of a trivial voluntary activity such as bungee jumping, such assessment must clearly be extended to the important and sometimes fundamental so, voluntary activity of pregnancy and childbirth.

Before moving on to the proposed solution to this problem in the form of ectogenesis, let me quickly address one final point of contention. Individual woman do have the option to reproduce without gestation and giving birth, through the use of surrogacy. One might argue, therefore, that no radical technological solution need to create to right the unjust biology distribution of the burdens of reproduction. Smajdor only touches on this point briefly in her article as the flaws of this idea are obvious. It is not the case that surrogacy is a realistic option for all women[23]. There is the financial burden which would instantly exclude most women. However, there is also the fact that such a plan would commit the fallacy of composition. It is true that each individual woman could hypothetically escape the burdens of reproduction by outsourcing, but it is not true of all women or women taken as a whole. I imagine an absurd world in which all women are carrying the babies of other women so that all women are hypothetical free of the burden of their own reproduction. Of course this reveals the mistake in such an idea; surrogacy only redistributes the burden from one woman to another, the unequal burden and injustice in its existence still remains.

Now I can more closely detail the solution to this problem; ectogenesis. In simple terms, ectogenesis refers to artificial womb technology; the gestation of the foetus outside of a woman's body. The specifics of the technology are not of any significant normative interest. However, it is does make the issue more pressing to know that such a Technology is no longer exclusively in the realm of science fiction, with successful gestations and births for rodents via ectogenesis (Randall and Randall 2008, 298). There are a number of advantages which ectogenesis could bring to gestation and childbirth such as a more ideal environment

---

[23] There is an interesting point to be made in the case of surrogacy that this is the burden which men have to live with in order to reproduce; there donor banks and artificial insemination for women for desire to get pregnant without a man but men need a surrogate to reproduce. My next chapter on male pregnancy will cover problematic issues for males regarding reproduction.

for foetus gestation (no drugs, alcohol, tobacco, etc.), less damaging birth, easier alternative to fertility treatment, easier access to the foetus for medical operations, and so on. There may also be disadvantages such as the oft lamented (but seldom explained) loss of the natural, and perhaps alienation and lack of connection to the foetus though bodily gestation. These aspects of the technology are interesting and deserve their own attention but they are not my focus here. My focus is on the one big advantage ectogenesis would bring in the form of the removing of the unjust burdens of child gestation and birth on the female population.

The burdens and inequalities that ectogenesis would remove for women have been discussed above, as are the failed attempts to dispute the existence of those burdens or to give alternatives to ectogenesis. For women not to suffer unjust burdens due to biology in the case of child birth it is clear that the only completely equalising solution is ectogenesis. However, male inequalities in reproduction (though they be fewer in magnitude) also exist and can also remedied through ectogenesis. One such case that is discussed by Randall and Randall (Randall and Randall 2008, 303-3010) is that of a father's reproductive interest in cases which the female does not desire to carry the baby to term and decides on abortion. Given the current state of affairs, the primacy of women's reproductive interests by virtue of their biological burdens and accompanying autonomy concerns is non-ideally just. Of course in a vacuum the concept that individuals should have unequal reproductive rights based on the chance circumstance of their biology rightly appears unjustly discriminatory. In ideal circumstances men and women would have equal consideration given to their reproductive rights. In removing the biological burdens on women, the restrictions on men's reproductive rights are also removed: In cases which the female desires to not gestate the foetus in her body, but in which the male desires to keep the foetus, ectogenesis could allow the male to exercise his reproductive rights; as long as it is stipulated that the procedure to extract and transfer the foetus to the artificial womb is no more dangerous to a woman's health than abortion. Ectogenesis can be the equaliser of rights, most significantly for women, but also for men.

Returning to the concept of "ought implies can", Smajdor makes a very thought provoking remark:

"If the only argument against someone's having a need met is that no one can meet that need, perhaps resources should be diverted toward being able to meet that need. That is to say, in some circumstances, perhaps ought implies ought to be able to. Kant's point, of course, holds when to fulfil a duty would be a logical impossibility. However, many things that are currently impossible (e.g., ectogenesis) are merely contingently so. We have either chosen not to focus research in these areas or are still struggling to find answers to the problem. As long as there is no logical impossibility, we are not exempted from our moral duties simply by the fact that we do not yet have a way of solving the problem" (Smajdor 2007, 338).

I think this paragraph can be summed up in the idea of "ought implies ought to be able to". In the case of ectogenesis, the motivating "ought to be able to" is we ought to be able to have complete sex equality. If one endorse sex equality, then to the extent which they endorse it they should also endorse ectogenesis; being that it is the only way to fully achieve the end goal of a principle they claim to endorse. As I have argued previously; "ought implies can" does not demarcate the moral but simply constrains our obligations to what is possible. It is clear that ectogenesis is in the category of moral goods which are currently outside the boundary of "ought implies can" and also (as Smajdor points out) only contingently so. Therefore, if one endorses the principle of sex equality they should endorse ectogenesis and if one endorses that sex equality and ectogenesis are things we 'ought to be able to' have then they endorse the means to achieving it through research; to bring about this morally good contingent possibility of ectogenesis into existence through the devotion of research time, effort and resources.

6.2 Conclusion

Finally I have a positive conclusion to present; here I am glad to say that sex equality passes the test of technological idealisation. If one admits, as I think one must, that not all inequalities between the sexes are socially constructed, then it is clear that only through some technological idealisation can true sex equality be achieved. What technological idealisation allows us to do here is more easily recognise the injustice between the situation of women and men in reproduction. Once the situation is made contingent there the use of hypothetical technology the appeal of "that's the way it is for women", or "It's unfair but

that's nature", no longer hold the same power. This contributes to the debate regarding reproduction immediately as although we don't have the technology of ectogenesis now we can work towards creating, as well as potential giving other forms of recompense to women who decide to reproduce in our societies today.

7. Reproductive Rights and Male Pregnancy

The other positive example I have to present is reproductive rights, to be more specific, positive reproductive rights of assistance. A particular interesting aspect of the example which I have chosen here is that Robert Sparrow presents the idea of male pregnancy as a negative example, and although he doesn't talk about technological idealisation he does go as far as to say that male pregnancy serves as a reductio ad absurdum for positive reproductive rights. I argue that Sparrow is wrong about this and that although male pregnancy might appear a little absurd, there is nothing morally problematic about it and Sparrow's arguments that there is are weak.

7.1 The Absurdity of Male Pregnancy

In my preceding chapter my presentation of pregnancy was overwhelming negative, so here I would like to redress the balance if only slightly with a positive interpretation of pregnancy and how researching a different technology could still help with sex equality. When shifting pregnancy from a negative to a positive interpretation one also shifts the burden, or in this case deficiency, from females to males. If pregnancy is a wonderful and fulfilling experience as some women claim it to be, the lives of men are necessarily improvised by being biologically unable to experience it. When I say "biological unable", I should add the qualifier "currently". In researching his paper "Male Pregnancy and the Limits of Reproductive Liberty" Sparrow not only came to the conclusion that male pregnancy is possible[24] but additionally that:

"Obviously, it will be imperative to research and minimize the risks to children who might be born to men, before male pregnancy should be attempted. However, there is no reason at this stage to believe that it will be impossible to reduce these risks to levels comparable to those accepted for pregnancies in women" (Sparrow 2008, 278).

This is important as it allows us to avoid wasting time debating the practical semantics of male pregnancy and instead get to the topic I want to explore in this section which is; redressing the imbalance in sex equality through male pregnancy.

---

[24] One reason to believe this is that babies can develop outside of the womb in women.

It is important to note here that the equality issue at stake here is fundamentally different from the one at stake in the case of ectogenesis; unjust burdens on one sex in order to reproduce. In the case of male pregnancy the issue is to be given access to an experience which is currently only available to females. It more closely mirrors debates around reproductive technologies to assist women who are unable to conceive such as "in vitro fertilization" (IVF). The important thing to note here is that this is not a negative reproductive right to non-interference in reproduction but a positive right to give women, who are unable to gestate and conceive without technological interference, the experience of pregnancy and birth. Sparrow notes this is his article (Sparrow 2008, 280-281) that many reproductive rights which we take for granted are positive in this way; state funded medical infrastructure and research for care before during and after pregnancy, IVF as previously mentioned, and the more controversial (though argued for none the less) technologies such as preimplantation genetic diagnosis and human cloning via "somatic cell nuclear transfer" (SCNT). Male pregnancy fits perfectly in amongst these positive reproductive rights[25].

One might stop me from going further down this rabbit hole by claiming that if no males desire pregnancy then they are not harmed by being incapable of experience it. I believe such a criticism is fundamentally mistaken in assuming you must desire something in order to be impoverished by its deficiency, if that were true than we could not call the ignorant individual's life improvised, but such a claim is easy defeated on its own grounds. Sparrow lists a number of people who would or have expressed desire in male pregnancy technology (Sparrow 2008, 283): Men who desire the intimate bonding experience of gestating and/or giving birth to their child. Male homosexual couples for wish to reproduce without a surrogate. Chromosomal males who are transgender females who wish to fulfil the biological role of women in reproduction[26]. Leaving aside the semantical minefield of whether pregnancy in transgender females should be considered male pregnancy, we have clear interest from males in the development of male pregnancy.

---

[25] One might try to frame these as negative rights against those who would interfere in attempting to prevent access to these technologies but the likelihood that one could do such a thing with only their own personal capital and infrastructure without state involvement at any level seems unlikely for what has come to be known as "the 99%".

[26] This is the most likely reason for the technology to be developed according to Sparrow's research.

Uninterestingly, Sparrow sides against male pregnancy by invoking retrograde gender roles:

"However, it is simply not the case that pregnancy is a normal part of men's lives. It is not a tragedy when a man cannot become pregnant-no matter how much he wishes to be pregnant. Barriers to men becoming pregnant do not constitute restrictions of reproductive liberty in the same way as do barriers to women becoming pregnant" (Sparrow 2008, 286).

"Because pregnancy is not a reasonable expectation in men, men who wish to become pregnant are not capable of establishing that this desire should be granted the same moral weight as women's desires to become pregnant" (Sparrow 2008, 288).

Personally I believe such notions as a "normal man's life", or "reasonable expectation in men", have little grounding in reality and even less argumentative force. However, debating Sparrow on this point would take me away from my argument. What is relevant though is that Sparrow takes male pregnancy to be an example of "*reductio ad absurdum*":

"Despite all of this, to talk of a male right to pregnancy is to parody both the language of rights and the desires of infertile women to give birth. Any argument for a right to reproductive liberty that extends as far as a man's right to become pregnant has lost contact with the facts about the biology of reproduction and its significance in a normal human life that made it plausible to defend the existence of such a right in the first place. I take it that there is a fairly obvious sense in which this conclusion constitutes a *reductio ad absurdum* of the notion of reproductive liberty" (Sparrow 2008, 286).

This is another example of technological idealisation. What Sparrow has done is taken the theory of positive reproductive rights as they currently exist, only applying to woman, and with this guiding principle, through an enhancement thought experiment, has arrived at what he sees to be an absurdity. Whether this is a positive or negative example of idealisation is a question which even Sparrow somewhat admits is an open question:

"There is also the contrary danger that I have succeeded only in proving that men do in fact possess such right. One author's *reductio* is, after all, another critic's brave and revolutionary conclusion" (Sparrow 2008, 292).

This perfectly encapsulates the goal of technological idealisations; it is a methodology which takes a theory (in this case positive reproductive right) to its extreme and in doing so reveals interesting aspects of that theory. The debate can be had, and I have started it here on my particular examples, as to whether the results are positive or negative. The results of these debates are important, but more important overall is the methodology of technological idealisations.

7.2 Conclusion

I find Sparrow's interpretation of the idea of male pregnancy being a *reductio* interesting. Although I believe there is much more left to be said on the topic, I also believe the case I have presented here in particular shows the utility of technological idealisations to both sides of a debate; show can see male pregnancy as an absurdity and others can see it as a worthy technology to pursue. Personally, I must put myself into the latter category[27]. However, I propose this only while hypothetically accepting the account of pregnancy which claims it is a wonderful, sometimes even life fulfilling experience. If such an account holds, it seems clear that males wanting to experience it would be missing out greatly; in the same way that females who cannot naturally conceive can feel. Given that I hold no objection to positive reproductive rights for women; it would seem I have little reason to object to positive reproductive rights for all people regardless of sex. Male pregnancy is certain a bizarre idea (though not so bizarre that it cannot be found in nature) but I cannot find anything immoral about it unless I was to give credence to Sparrow's talk about gender roles or the all too common appeal to nature.

---

[27] Though I am arguing for the method and not for my interpretation of the results, I put this for the sake of honesty and completeness.

8. The Value of Technological Idealisations to Future Ethical Thinking

I began this thesis by looking at idealisation in the form of Rawlsian ideal theory, which was then expanded on by Roduit's in the field of human enhancement technology. I sought to show how such idealisations can be applied to ethical theories and the hypothetical technologies which are consistent with their theoretical aims. Rawlsian ideal theory gives interesting information and guidance to current political non-ideal theory, and Roduit's application of ideal theory to the concept of humanity itself gives interesting insight into what it means to be human, although it ultimately collapsed into an ideal "theory of humanity", rather than having anything to do with humanity in the biological sense. This newfound focus on theory as the ideal to guide enhancement, and technology in general, lead to what I've labelled "technological idealisation".

My goal was to show that this idealisation of theory through technology is a valuable concept which can contribute to the ethical debate. Towards this end I looked at a significant number of examples from the philosophical literature. I started with utilitarianism and showed that Agar's hypothetical technology of the moral beta-blocker keenly reveals the problematic demands which utilitarianism makes of its adherents. The case of Dr Angela showed that utilitarianisms adherent demands, to infect her daughter with HIV in order to test an experimental drug on her, are not so easily dismissed when we imagine the technological idealisation of the moral beta-blocker. When faced with truly acknowledging what it would mean to act out the demands of this ethical theory, through vivid thought experiment by technological idealisation, I can easily imagine one's opinion on it changing.

I then looked at the well-known example of Nozick's experience machine and its critique of hedonism, claiming that it wrongly disregards the value of reality. I believe less need be said about this example due to its notoriety. However, its inclusion here served as a paradigm example those understandably sceptical of the process. Having such a well-known thought experiment serve as an exact example of the technological idealisation methodology which I am presenting really helped to legitimise my chosen research.

The God machine thought experiment by Savulescu and Persson provides a thought provoking discussion of the debate surrounding the sacrifice of autonomy for safety, with an additionally interesting parallel with Hobbes' work in the Leviathan. I presented both the original version of the God machine which only intervenes in cases of serious moral wrongdoing, and my own modified God machine which acts in all cases of wrongdoing (out of consistency) and showed that both pose serious problems for autonomy. The technological idealisation here acutely presents the debate of the sacrifice of autonomy for safety in a way that really allows one to focus on the competing values at play and narrow one's stance on them.

Smajdor's discussion of ectogenesis has demanding conclusions for those who accept sex equality, showing that biology should not be a barrier to justice. Through technological idealisation blocking off any appeal to nature and the status quo to prevent us from righting injustice, the contingency and unjust burdens of reproduction on the females becomes undeniably clear. Whether this should be righted by ectogenesis itself or by some other method can be debated elsewhere, but what is clear is that ectogenesis as a technological idealisation provides strong evidence for some kind action to be taken to reduce the burden of reproduction on women.

And finally, the possibility of male pregnancy as written about by Sparrow has interesting implications for advocates of positive reproductive rights; these implications being undesirable to Sparrow but potentially very desirable to others shows the utility of technological idealisations to both sides of a debate. If we accept a positive interpretation of the experience of pregnancy then males (at least those who desire to experience pregnancy) now become the disadvantaged in the mater of reproduction; an interesting comparison with the ectogenesis case. On the one hand, like Sparrow, one might see that as an absurdity which questions our interpretation of positive reproductive rights, or on the other you might conclude that positive reproductive rights demands that we also give assistance to men who desire pregnancy. This different interpretation of the same result is a clear case of technological idealisation adding to a debate, and even creating its own.

After presenting all these examples, I feel secure in concluding two things; that there are many examples of technological idealisation in the philosophical literature and those

technological idealisations are valuable to our ethical thinking. Even outside the field of ethics I believe there is interesting applications of the methodology to be found such as the "brain in a vat" thought experiment following on from the work of Descartes, technologically idealising a particular sceptical scenario. My focus here was on ethics however, and I believe I have provided adequate evidence that there is a valuable evidence and contribution to debate by technological idealisation in this field.

The value of these examples, and the methodology from which they are derived, comes from the ability of ignore practicality concerns and acutely focus on theoretical issues. To question the feasibility of the God machine, the cost of creating the experience machine, or whether a drug could affect our moral psychologies in the exact way the moral beta-blocker does, is to miss the point of their discussions. These technological idealisations and technological idealisations in general, provide interesting theoretical discussions to which practical concerns is irrelevant but which have practical theoretical significance to ethical theory: What does utilitarianism demand of us when we are in a position to adhere perfectly to the theory? If pleasure is the only non-instrumental value, how far can we go in pursuit of it? Does biology circumscribe the extent of the demands of justice? And so on. The acute focus which aids understand and debate of the theoretical concepts at play in these, and other, discussions would be lost if we more to ignore technological idealisation in our ethical thinking.

It is my hope that having established the existence and value of technological idealisation in past and relatively recent philosophical debate, that readers of this thesis will be inspired, or at least open, to including them in their future ethical thinking. This could come in the form of testing their own theories against the methodology, or they might also apply the method of technological idealisation to longstanding ethical debates which are yet to benefit from such an analysis. If the examples here are as valuable as I have argued that they are, I should expect that much interesting and fruitful thought experiments and debate will come from my recommendation.

9. Bibliography

Agar, Nicholas. *Liberal Eugenics: In defence of human enhancement*. Oxford: Blackwell Publishing Ltd., 2004.

Agar, Nicholas. *Humanity's End: Why we should reject radical enhancement*. Cambridge: MIT Press, 2010.

Agar, Nicholas. "Enhancing genetic virtue?" *Politics and the Life Sciences* 29/1 (2010): 73-75.

Agar, Nicholas. *Truly Human Enhancement*. Cambridge: MIT press, 2014.

Agar, Nicholas. "Moral Bioenhancement and the Utilitarian Catastrophe." *Cambridge Quarterly of Healthcare Ethics* 24 (2014): 1-11.

BBC. *"*Tech giants to freeze eggs for their female employees." Accessed February 20, 2016. http://www.bbc.com/news/business-29626291. Last Modified October 15, 2014.

Bostrom, Nick. "A History of Transhumanist Thought." *Journal of Evolution and Technology* 14/1 (2005): 1-25.

Bostrom, Nick, and Rebecca Roache. "Ethical Issues in Human Enhancement." In *New Waves in Applied Ethics*, edited by Jesper Ryberg, Thomas Petersen & Clark Wolf, 120-152. London: Palgrave Macmillan, 2008.

Buchanan, Allen, and Dan W. Brock, Norman Daniels, Daniel Wikler. *From Chance to Choice: Genetics & Justice*. Cambridge University Press: Cambridge. 2000.

Buchanan, Allen. "Moral Status of Human Enhancement." *Philosophy & Public Affairs* 37/4 (2009): 346-381.

Bunnin, Nicholas, and Jiyuan Yu, eds. *The Blackwell Dictionary of Western Philosophy*. Oxford: Blackwell publishing Ltd., 2004.

Chadwick, Ruth. "Therapy, enhancement and improvement." In *Medical enhancement and posthumanity*, edited by B. Gordijn, and R. Chadwick, 25-37. New York: Springer, 2008.

Cong, Le, and F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, Xuebing Wu, Wenyan Jiang, Luciano A. Marraffini, Feng Zhang. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (2013): 819-823.

De Brigard, Felipe. "If you like it, does it matter if it's real?" *Philosophical Psychology* 23/1 (2010): 43-57.

Douglas, Thomas. "Moral Enhancement." *Journal of Applied Philosophy* 25/3 (2008): 228-245.

Douglas, Thomas. "Enhancement, Biomedical." in *The International Encyclopaedia of Ethics* edited by Hugh LaFollette, 1633-1643. Oxford: Blackwell Publishing Ltd., 2013.

Ereshefsky, Marc. "Species Pluralism and Anti-Realism." in *Philosophy of Science* 65/1 (1998): 103-120.

Firestone, Shulamith. *The Dialectic of Sex: The Case for Feminist Revolution.* St Albans: Granada Publishing Ltd., 1972.

Fukuyama, Francis. *Our Posthuman Future: Consequences of the Biotechnology Revolution.* New York: Farrar, Straus, and Giroux, 2002.

Griffin, James. *Value Judgement: Improving Our Ethical Beliefs*. Oxford: Oxford University Press, 1996.

Harris, John. Moral Enhancement and Freedom. *Bioethics* 25/2 (2011): 102-111.

Hobbes, Thomas. *Leviathan*. Edited by Michael Oakeshott. Oxford: Basil Blackwell, 1946.

Hull, D. L. "The ideal species concept - and why we can't get it." in *Species: The Units of Biodiversity*, edited by M. F. Claridge, H.A Dawah and MR. Wilson. Netherlands: Springer 1997.

Mill, John Stuart. *Utilitarianism*. London: Longmans Green and Co. 1907.

Mill, John Stuart. *On Liberty.* The Project Gutenberg EBook, 2011. Accessed 25/02/2016. http://www.gutenberg.org/files/34901/34901-h/34901-h.htm

Moore, Andrew. Stanford Encyclopaedia of Philosophy, "Hedonism" Last Modified October 17, 2013. http://plato.stanford.edu/entries/hedonism/

Nozick, Robert. *Anarchy, State, and Utopia*. Oxford: Basil Blackwell Ltd., 1974.

Plato. *Parmenides*. Translated by Sylvana Chrysakopoulou & Arnold Hermann. Las Vegas: Parmenides Publishing, 2010.

Randall, Vernellia R., and Tshaka C. Randall. "Built in Obsolescence: The Coming End to the Abortion Debate." *Journal of Health & Biomedical Law* 4 (2008): 291-310.

Rawls, John. *A Theory of Justice*. Oxford: Oxford University Press, 1999.

Rawls, John. *The Law of Peoples with The Idea of Public Reason Revisited*. Cambridge: Harvard University Press, 2003.

Rescher, Nicholas. "Reductio ad Absurdum" http://www.iep.utm.edu/reductio/ Accessed September 12, 2015.

Roduit, J. A. R., and H. Baumann, J.-C. Heilinger. "Ideas of perfection and the ethics of human enhancement." *Bioethics* 29/9 (2015): 622-630.

Roduit, J. A. R., and H. Baumann, J.-C. Heilinger. "Evaluating human enhancements: the importance of ideals." *Monash Bioethics Review* 32 (2014):205-216.

Sandel, Michael J. *The Case against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge: Harvard University Press, 2007.

Savulescu, Julian, and Ingmar Persson. "The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity." *Journal of Applied Philosophy* 25/3 (2008): 162-177.

Savulescu, Julian, and Ingmar Persson. "Moral Transhumanism." *Journal of Medicine and Philosophy* 35 (2010): 656-669

Savulescu, Julian, and Ingmar Persson. "Moral Enhancement, Freedom and the God Machine." *Monist* 95/3 (2012): 399-421.

Sterelny, Kim, and Paul E. Griffiths. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press, 1999.

Smajdor, Anna. "The Moral Imperative for Ectogenesis." *Cambridge Quarterly of Healthcare Ethics* 16 (2007): 336-345.

Sparrow, Robert. "Is it "Every Man's Right to Have Babies If He Wants Them"? Male Pregnancy and the Limits of Reproductive Liberty." *Kennedy Institute of Ethics Journal* 18/3 (2008): 275-299.

Sparrow, Robert. "Better living through chemistry? A reply to Savulescu and Persson on Moral Enhancement." *Journal of Applied Philosophy* 31/1 (2014): 23-32.

Stern, Robert. "Does "Ought" Imply "Can"? And did Kant think it does?" *Utilitas* 16/1 (2004): 42-61.

Valentini, Laura. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7/9 (2012): 654-664.