Simulating the Substance View

A Lightweight Case for Human Moral Equality

Sivert Thomas Ellingsen

Thesis presented for the degree of
MASTER OF PHILOSOPHY

Supervised by Morten Magelssen (Researcher, Centre for Medical Ethics, University of Oslo) and Professor Bjørn Torgrim Ramberg

Department of Philosophy, Classics, History of Art and Ideas
Faculty of Humanities
UNIVERSITY OF OSLO

Spring 2018
Simulating the Substance View

A Lightweight Case for Human Moral Equality

Sivert Thomas Ellingsen
Simulating the Substance View: A Lightweight Case for Human Moral Equality

Sivert Thomas Ellingsen

http://www.duo.uio.no

Trykk: Reprocentralen, Universitetet i Oslo
Abstract

A key premise of one standard argument for the view that abortion is immoral (and for many other “conservative” views in bioethics) is that all human beings are equal in moral status, so that it is morally permissible to kill a given unborn human being under a given set of circumstances only if it would be morally permissible to kill a normal, adult human being under equivalent circumstances. This, many critics charge, seems unmotivated: Why should the mere fact that I am a member of the species *homo sapiens* carry any moral significance? One relatively recent strand of thought, which has been dubbed “the substance view of human persons” (or simply “the substance view”), seeks to answer this objection by arguing, very roughly, that all human organisms have rational natures, and that this is what endows them with their moral status. However, the substance view comes prepackaged with a controversial, Aristotelian metaphysics, and will therefore not do much to convince the many who do not already accept such a metaphysics. This work outlines a case for human moral equality which is analogous to that given by proponents of the substance view, but premised on more modest, less controversial metaphysical assumptions. Instead of grounding human moral equality in the claim that all human organisms have rational *natures*, it argues, we may ground it in the less controversial claim that all human organisms have what I call the *genetic basis for rationality*. First, I explain and defend the claim that all or nearly all human organisms have the genetic basis for rationality. Then, I argue that if, as seems very plausible, rationality itself is a sufficient condition of the sort of moral status had by fully-developed human persons like you or me, then having the genetic basis for rationality is sufficient for that same sort of moral status. Finally, I investigate what does – and doesn't – follow from the conclusion that all human organisms are moral equals. Along the way, I also address questions about dispositions, personal identity, and the nature of rationality.
Acknowledgments

First and foremost, I want to thank my supervisors, Morten Magelssen and Bjørn Ramberg. Their thorough and insightful feedback has spurred many of the best ideas that are found in the coming pages, and gently steered me away from countless bad ones. Thanks also to Øyvind Johannes Vardenær Evenstad and Patrick Winther-Larsen, who both took the time to read and comment on an earlier draft of this thesis, and to Dag August Schmedling Dramer and Øystein Linnebo for helpful comments on a presentation of what would eventually become its third chapter. And, last but not least, thanks to my family for their constant encouragement and support.
# Table of Contents

1 Introduction 1

2 Surveying the Landscape 12
   2.1 Marquis 13
   2.2 McMahan 14
   2.3 Boonin 19
   2.4 The substance view 24
   2.5 Liao 27
   2.6 Looking ahead 30

3 Rationality and Its Genetic Basis 31
   3.1 Rationality 31
      3.11 Rationality as reasoning 31
      3.12 Actualism 33
      3.13 Dispositional rationality 36
   3.2 The base and basis of human rationality 46
      3.21 The proposal 46
      3.22 Objections 47
      3.23 Who has the basis? 51
   3.3 Looking ahead 54

4 The Genetic Criterion 55
   4.1 Why not the Genetic Criterion? 55
      4.11 Why not human moral equality? 55
      4.12 The Mix-Up 60
      4.13 Interests 62
   4.2 Why the Genetic Criterion? 64
      4.21 How does first-order rationality generate rights? 64
         4.211 Rationally specific goods 65
         4.212 How rationally specific goods generate rights 67
      4.22 (How) does second-order rationality generate rights? 71
4.221 The consequentialist account again  71
4.222 General morals  80

5 Coda  84

Bibliography  87
Introduction

Here is one formulation of what is often called the standard argument against abortion\(^1\) (henceforth, the Standard Argument):

1. A fetus is a human being
2. It is morally wrong to kill a human being
3. Abortion constitutes the killing of a fetus
4. So, abortion is morally wrong

A couple of clarifications are already in order. The term “human being” is sometimes used synonymously with the term “human organism” – i.e., to denote all members of the species *homo sapiens* – and sometimes only to denote those members of the human species that are *persons* – i.e., roughly, self-conscious and rational. Here, I will use it in the first, wider sense. The adjective “human” will be used in the same way. The term “human person,” however, will denote the second, smaller group. The term “fetus,” meanwhile, is sometimes used to refer to an unborn baby at any stage of development, and sometimes only to unborn babies after the end of the embryonic stage of development (i.e., roughly, after the eight week of pregnancy). Again, I will use this term in the first, wider sense.

The pivotal premise of the Standard Argument – the one around which reasonable disagreement about its soundness ought to center – is very arguably the second one. Premise (3) will presumably be granted by most everyone, though perhaps there could be quibbles about edge cases. Perhaps, e.g., performing a lifesaving hysterectomy on a pregnant woman in the knowledge that the death of the fetus will result counts as an abortion, but also as merely letting die rather than as killing. Even if such quibbles are granted, however, their cash value is limited, since one can perfectly well accept them while also remaining staunch in one’s opposition to those abortions that do uncontroversially count as killing, which make up the vast majority of abortions. For example, it could be held that we are sometimes permitted to bring about the death of a fetus as a *foreseen* consequence of our actions (such as in that hysterectomy I

---

\(^1\) See e.g. Oderberg 2000, p. 3
mentioned a moment ago) even if we may never bring it about as an intended consequence. And again, it is not clear whether such “indirect abortions” actually count as abortions at all. It is not even out of the question simply to stipulatively define abortion as the intentional killing of a fetus and dispense altogether with the now-tautological third premise.

In popular debates, premise (1) is often taken by both sides to be the bone of contention. In such debates, slogans like “Life begins at conception” and “A fetus isn’t human; it’s just a clump of cells” are often presented as standalone arguments, suggesting that if only we settled whether the fetus is a living human organism, we would also settle whether (or when) we are permitted to kill it. But this is very arguably wrong, simply because it already is settled, as a matter of empirical fact, that the fetus is a living human organism from the moment of conception. Crack open an introductory textbook in embryology, and you will usually find a claim to this effect near the start.2

So, (2) is the pivotal premise. Its prima facie plausibility should not be underestimated. Most ordinary people, including many or most of those who consider themselves pro-choice, would probably assent to it (in many cases, one assumes, without realizing its full consequences). As S. Matthew Liao notes, the idea that membership in the human species confers moral status also seems to be the motivation for much of our current preoccupation and concern with human rights.3

Now, against both of these points, it could be objected that the terms “human” and “human being” are ambiguous, as already mentioned. When the man in the street talks about the rights of human beings, perhaps he really means the rights of human persons, on the tacit understanding that these rights are a result of their personhood rather than their humanity. Well, perhaps. But in that case, why include the qualifier “human” at all? If we really believed in our heart of hearts that it is only the personhood of human persons that has any moral relevance, focusing so much energy on the rights of human persons instead of the rights of persons plain and simple would be as strange as focusing a similar amount of energy on the rights of some other subset of persons who also happen to share some other attribute which we all understand to be morally irrelevant. A declaration of the rights of human persons would be as bizarre as a declaration of the rights of green-eyed persons or persons who live on the 33rd parallel north.

Now, while the man in the street (including, in many cases, the pro-chooser in the street) may accept (2) but think that (1) is up for debate, pro-choice philosophers standardly accept (1) but dispute (2). These disputations take many forms, but one argument one often hears is that

2 See https://www.princeton.edu/~prolife/articles/embryoquotes2.html for a selection of examples.
3 Liao 2010, p. 160
it is *speciesist* to give any special moral status to human beings simply because they are human beings.

I think this charge of speciesism can be dispensed with pretty quickly. Even if true, it gives no independent and non-circular reason to reject (2). After all, speciesism is supposed to be analogous to racism and sexism, and, in particular, to be immoral and irrational for the same reasons that racism and sexism are immoral and irrational. But racism and sexism are surely immoral and irrational because they claim that belonging to a certain race or sex confers a certain moral status when it actually does not. Thus, if speciesism is objectionable, then it is objectionable because it claims that belonging to the human species confers a certain moral status when it actually does not. But in this context, where the moral status of humans *qua* humans is exactly what’s at issue, it is obviously circular to reject (2) on the grounds of an undefended conviction that it is objectionable in this way. Nor does (2) entail what is usually taken to be the most repugnant consequence of speciesism, namely that that non-human animals have fewer and/or weaker moral rights than humans (and quite possibly no moral rights at all). Moral status is not a scarce resource like gold or water, so endowing human beings with a lot of it does not leave less for everyone else. It does not rule out endowing non-human animals with significant moral status, or even the same moral status as humans.

Now, if the critic argued, rather than merely assumed, that there are no good reasons to believe humanism, then he would have a good case that accepting (2) is at least irrational, and therefore potentially speciesist. The charge of speciesism now becomes non-circular, but only at the cost of losing all independent dialectical force. For the real objection is now no longer that (2) is *speciesist*, but that (2) is *unmotivated*. So at best, the speciesism objection is parasitic on what I call the unmotivatedness objection, to which we now turn.

In its simplest form, the unmotivatedness objection takes the form of an incredulous stare at (2). In a more elaborate form, it proceeds from the complaint that it is, in general, utterly mysterious why species membership *as such* should carry any moral value. But the humanist is not committed to the claim that membership in the human species (or any other species) carries any moral value as such. Instead, he can claim that humans have their moral status *qua* humans, not directly in virtue of their humanity, but in virtue of certain essentially human properties – i.e., in virtue of properties that are necessarily had by every human. In other words, the humanist can argue for his position with a syllogism of the following form:

(5) All humans are F
(6) All Fs have a right to life
So, all humans have a right to life

Note that here (and for the rest of this work, except where otherwise noted) I will use the terms “right to life” and “human moral equality” in a somewhat non-standard (and largely interchangeable) way. This is simply because they are convenient labels for a notion that is central to this work. The latter of these terms gives the clearest idea of their real meaning in the present context. Human moral equality, as I now use the term, is roughly the assertion that all humans are fundamentally the same with respect to when we may and may not kill them. In other words, human moral equality is the assertion that, for any two humans, call them Alice and Bob, it is morally permissible to kill Alice under a given set of circumstances if and only if it is morally permissible to kill Bob under the same (or equivalent) circumstances. So, in particular, causing the death of a fetus is morally permissible only if it would be permissible to cause your death or mine under the same (or equivalent) circumstances. Meanwhile, to have the right to life, in its present, non-standard meaning, is simply to be such that one may be killed only under those circumstances (or relevantly the same circumstances) in which you or I may be killed. So, if human moral equality is true, it follows that all humans have a right to life, but not that all entities that have a right to life are humans. (There is, therefore, a sense in which this work presumes that all normal human adults, at least, do have the same fundamental moral status. We will return to this point later.)

The question is whether there is a syllogism of this form whose middle term renders both premises true. In other words, is there a property that is had by all human organisms, and only by entities with a right to life?

One popular candidate for such a property is rationality, yielding the following syllogism:

(8) All humans are rational
(9) All rational entities have a right to life
(10) So, all humans have a right to life

Here, the appeal of the minor premise (“All rational entities have a right to life”) is fairly easy to see. However, the major premise (“All human organisms are rational”) clearly stands in much greater need of defense. Intuitively, many human organisms are not rational. And more to the point, most or all fetuses seem to fit into this category. After all, fetuses don’t engage in the sorts of activities we tend to think of as constitutive and definitive of rationality: arguing
discursively, fixing one’s will on certain ends, deliberating about which means will best accomplish these ends, and so on. In short, they don’t reason. Admittedly, there does seem to be a sense in which human beings qua human beings are rational animals, just as there seems to be a sense in which dogs qua dogs are four-legged animals. But just as dogs being four-legged does not entail that all dogs are four-legged, so the fact that we are rational animals does not entail that each and every one of us is rational. “Humans are rational animals,” in the sense that renders it true, is a statement about the typical or distinctive properties of the species as a whole, not an attribution of rationality to each and every individual specimen. It is a generic, not a universal. And, it is far from clear that belonging to a species that is typically F endows one with any particular moral status, even if F-hood itself does. (For example, if two normal human adults Jack and Jill are alone in a certain room with a certain statue of Napoleon, it is surely typical of inhabitants of that room to have the right to life, since both Jack and Jill, who make up two thirds of the group in question, presumably do have such a right. But clearly, the fact that the statue of Napoleon is also a member of this group does not entail that it, too, has a right to life.)

A possible response to this objection goes as follows: While it’s true that not all human beings actually engage in reasoning, actual reasoning activity is not necessary to count as rational. Rather, to be rational is to have a capacity of the right sort to engage in reasoning. And according to one strain of thought, all human beings do have such a capacity, in virtue of their rational natures. On this view, to be a member of the human species (or any other species, for that matter) just is to have a certain nature or essence. These natures or essences guide the development and maturation of the organism, and do so by endowing them with certain basic (though not always immediately exercisable) capacities, including, in our case, the capacity for rational agency.

This suggests an essentialism. It does not, however, suggest the sort of essentialism that modern analytic metaphysicians have inherited from Saul Kripke. On this modern, “Kripkean” essentialism, Fs are essentially G just in case necessarily, all Fs are G, and the essence of a species or a specimen is simply the set of all its essential properties. So, in particular, if human beings are essentially rational, it will be true (and necessarily true at that) that all human beings are rational, and we’re right back to square one.

---

4 Kripke 1980, p. 39 et passim
No, the essentialism I have in mind is much older. Its best-known exponent is Aristotle. On this essentialism, the essence of a thing is not the set of all its essential properties, but, roughly, a sort of substrate out of which its essential properties flow. A bit more precisely, and following authors like Fabrice Correia⁵, this notion can be cashed out as follows: The essential properties of any given object are those properties that that object has in virtue of what it is (i.e., for the Aristotelian, in virtue of its essence). But from the fact that a certain object has a certain property in virtue of its essence, it does not necessarily follow that all other objects that have the same essence also have that property. Perhaps an essence can be blocked from expressing the essential properties that flow out of it. For instance, while not all dogs are four-legged, it is plausible that dogs are essentially four-legged, four-legged in virtue of being dogs – it’s just that in some cases, injury or genetic defects have blocked them from expressing their four-legged nature. Further, a four-legged dog is to that extent a better – a more “doggy” – dog than a three-legged dog, since it better approximates the ideal of “dogginess.” Further still, the maturation and development of living organisms is, in essence, the striving of that organism towards the ideal given by its form. Aristotelian essentialism, then, comes prepackaged with a teleology.

Thus, what all human beings have in common is not that they realize the rational capacities – exemplify the rational ideal – that lies in their nature. In fact, it’s plausible that no human being exemplifies the ideal perfectly. Rather, what they have in common is that they are ordered towards rationality, and will become rational when all goes well. Those human beings that are rational (which may or may not be all of them) are rational in virtue of being human. It is this ordering towards rationality, which is a great good, that gives them their right to life.

For proponents of what is often called the substance view of the fetus⁶, this essentialism becomes the background theory for a systematic metaphysics of human beings. As the name suggests, the key claim in this metaphysics is that human beings are substances in the full-blooded Aristotelian sense of the word – independently existing composites of matter and form that persist through change and development.

The substance view thus not only brings with it an Aristotelian essentialism, but also a certain theory of personal identity. Since we are substances, our identity- and persistence-conditions are the same as the identity- and persistence-conditions of any other substance. Hence, what makes me the same being as the five-year-old child I once was (or the fetus I was

⁵ Correia 2006, p. 755 et passim
⁶ Also sometimes called the substance view of human beings or the substance view of human persons.
before that) is not, e.g., any sort of psychological continuity, but the fact that I have remained one and the same composite of matter and form through it all.

This sort of Aristotelian substance theory – and especially the essentialism and teleology it brings with it – is often accused of being long past its sell-by date, of being at odds with the scientific view of the world we now have. I myself don’t buy these accusations, but I do think they highlight a practical problem. You won’t get very far with an argument that’s premised on a metaphysics most people reject. You can, of course, try to convince those people of that metaphysics while you’re at it, but that’s not a road we should go down unless we absolutely have to. The abortion debate is intractable enough as it is – I would rather not add a debate about the foundations of metaphysics on top of it.

Those of us who are attracted to the substance view, but who would rather not have to complicate the debate by defending its contentious background metaphysics, therefore have a strong interest in reconstructing it in less metaphysically contentious terms – in simulating the substance view. But is this possible? Can we avoid committing ourselves to the substance view’s contentious background metaphysics while giving an independently plausible argument that reaches similar conclusions through a similar reasoning, and which exploits the same basic intuitions? That is the central question of this work. And while I do not aim to answer it definitively, I do want to suggest what I think is one promising way in which this might, with more work, be done.

As is hopefully clear, my “simulation” is meant as a friendly supplement to the substance view, not a competitor to it – a Coke Zero to its regular Coke, something to serve up to friends who have foolishly gone on a no-Aristotelianism diet. But nor should the simulation only be of interest to those who already find the substance view plausible. Rather, it should be of interest to anyone who takes an interest in the ethics of abortion (or in any of the many other ethical questions for which human moral equality has implications).

None of this, by the way, is meant to suggest that ethical arguments can or should be metaphysically neutral. I’m as reactionary in my metaphilosophy as I am in my essentialism – I think metaphysics is first philosophy, that when faced with a theory that claims to have no metaphysical commitments, we should only ask how well it has managed to obscure its metaphysical commitments. It’s just that I want to avoid premising my argument on highly controversial metaphysics. I say “highly controversial,” not “controversial.” Almost every substantive philosophical theory is bound to be controversial, in the sense that there is no universal consensus among philosophers about whether or not it is true. But the greater the consensus, the better. My ambitions toward lightweightness therefore don’t entail that I won’t
help myself to some philosophically contentious premises, and even a good number of them. It should also be noted that I will not have the room to defend most of these theories in a way that would (or should!) fully satisfy a skeptic. But as far as I can tell, these premises are all respectable, in the sense that a significant number of people with relevant expertise accept them, and have good reasons for doing so. (Respectability thus understood is, I take it, quite compatible with controversy, and even unpopularity. In most contexts, at least, a large or largish minority of experts can constitute a “significant number.”) If I succeed in deriving my conclusion from respectable premises in this way, I will, I take it, have shown that that conclusion itself is (or ought to be) respectable, and has therefore passed the first test any philosophical hypothesis must pass. The job this leaves for future works is to investigate whether the hypothesis can also pass the other tests that come after that – whether it is not merely respectable, but also true.

The intuition at the heart of the substance view, I take it, is that there is some sense in which all human beings – even those that do not or cannot engage in reasoning or rational agency – have certain basic rational capacities, and that these capacities endow them with significant moral standing. Going back to a notion of essence introduced earlier, we can reformulate this intuition in the following way: Those members of the human species that are rational are rational in virtue of the fact that they are members of the human species. The natural followup question is what makes us members of the human species, which is an instance of the more general question of how (or whether) we can distinguish different biological species at all. As we have seen, there is a view that a species is defined by a certain sort of nature, so that this notion naturally yields the thought that those of us who are rational are rational in virtue of our natures or essences. From this point of view, the source of the substance view’s metaphysical contentiousness is its account of species.

What if we replaced this notion of a species with a less contentious one? The notion I have in mind is the following: To be a member of the human species just is to have a human genome. Now, this is not a satisfactory account of species membership, for rather than resolving the problem, it simply shifts it from one place to another – from the question of what makes one species different from another to the perhaps equally difficult question of what makes the genome of one species different from that of another. (Indeed, defining a human as a living organism with a human genome seems downright circular. It may very well be impossible to give a good answer to the latter question if one does not, at some level, already have a good answer to the former.) But it seems accurate as far as it goes, even if that isn’t very far at all. And more importantly, it suggests an interesting gloss on the intuition at the heart of the
substance view: that those human beings who are rational are rational in virtue of their *genomes*, and that it is these genomes that endow *all* human beings with their moral standing.

Note the similarities between genomes and Aristotelian natures. Both are “substrates” that define a certain species, and out of which the defining characteristics of that species typically flow. Like Aristotelian natures, genomes can also be prevented from expressing these defining characteristics in a particular specimen. To return to a previous point, the genomes of dogs code for four-leggedness, but that does not mean that all dogs are four-legged. However, they also differ in at least one big way: The existence of the human genome, unlike the existence of human nature, is uncontroversial, easily verified, and unlikely to ruffle the feathers of philosophical naturalists. If we want to simulate the substance view, it is therefore well worth looking into whether the human genome can play the same role in our theory as does human nature in its standard, “heavyweight” form, letting us argue for human moral equality on the basis of our rational genomes instead of our rational natures. The purpose of this work is to outline one way in which this might be done.

The central argument of this work (or just the Central Argument for short) looks like this7:

1. All human organisms have the genetic basis for rationality (*P*)
2. All entities with the genetic basis for rationality have a right to life (*P*)
3. Therefore, all human organisms have a right to life (*1, 2*)
4. All human fetuses are human organisms (*P*)
5. Therefore, all fetuses have a right to life (*3, 4*)

Again, “human organism” is here simply defined as “living organism with human DNA,” leaving little room for serious doubt about whether (4) is true. This leaves two harder tasks. The first is to explain in greater detail what it is to have the genetic basis for rationality, as well as defending the claim that we do all have such a basis. The second of the hard tasks is, of course, to defend (2). In particular, if we are to stay wholly true to the basic intuition behind the substance view, we should argue for (2) by arguing that the genetic basis for rationality endows all humans with certain basic rational capacities, and that these capacities are sufficient for the right to life.

---

7 Because of the Central Argument’s importance to this work, it seems natural to “reset” the numbering of statements here. In all places where the context does not suggest otherwise, any subsequent references to statements (1) through (5) will be references to the premises and conclusion of the Central Argument.
The project will unfold in the following way. Chapter 2 will survey some particularly relevant parts of the current philosophical debate about the ethics of abortion in a bit more detail. Chapter 3 will be devoted to defending (1) and (perhaps more importantly) getting clearer on what it is to have the genetic basis for rationality. Chapter 4 will be devoted to defending (2). Finally, in the Coda, I will explore what sorts of further conclusions we can and cannot draw from the conclusion. For even if the Central Argument is successful, I will argue, the pro-life philosopher’s work is far from done. As a general statement about moral status, it leaves many concrete questions about the moral – and legal – status of abortion unanswered.

In closing, another note on terminology may be in order. It is hard to know exactly what to call the different camps in the abortion debate, for at least three reasons. First, most of the standard ways of drawing the line seem to have an undesirableslant in one direction or the other – that is, they seem to be either uncharitable or euphemistic. For example, to divide the debate into those who are “anti-abortion” and those who are “pro-abortion” is surely unfair on those who are “pro-abortion,” since it implies that they think abortion is a morally good or obligatory thing to do. But of course, most of them do not think that – they merely claim that abortion is morally permissible. Similarly, the terms “pro-choice” and “pro-life” seem at once euphemistic and uncharitable. For they suggest that being for or against abortion simply amount to being in favor of choice and being in favor of life, respectively. Consequently, each term also implies that the other side is against choice and against life, respectively.

Second, claiming that there is a line to be drawn at all – a sharp dichotomy between thinking that abortion is permissible and thinking that it is impermissible – can easily obscure the fact that, in reality, there is a smooth and gradual spectrum of views on the ethics of abortion, going all the way from those who think that all forms of abortion are morally abhorrent to those who not only defend the prima facie permissibility of all forms of abortion, but also of many forms of infanticide. Even assuming both that the abortion debate is primarily a debate about when the right to life kicks in and that the right to life does kick in at some point, the right to life could kick in between any one of an infinite (or at least very, very large) number of points between conception and infancy. (Further, and as we will see much later, there is room for a great deal of disagreement about the ethics of abortion even among those who think the right to life kicks in early.)

Third, all of the standard terms are sometimes used to denote views on the moral status of abortion, and sometimes to denote views on its legal status. (Sometimes they can also be used to denote a “package” of views on both matters, or in a way that leaves it ambiguous which of the previous three meanings is intended.) A pro-lifer, for example, could be someone who
thinks that abortion is immoral, or he could be someone who thinks that it ought to be illegal. But as we will see near the end of this work, the logical relation between the ethics of abortion and the politics of abortion is more loose and separate than we might think. For example, a person could be (and some people are) pro-life in the moral sense, but not in the legal sense.

Still, in deference to convention, and with full knowledge of their flaws, I have here used the terms “pro-choice” and “pro-life,” and will go on doing so. As I use them, these terms will denote only views on the ethics of abortion. More specifically, they will denote, respectively, the view that most abortions that are actually performed are morally permissible and the view that most abortions that are actually performed are morally impermissible.
Before we get into my defense of the Central Argument, it will be useful to go over some of the literature that that defense will draw on. This way, we won’t need to make awkward expository pauses later on, and we will get a clearer picture of the conversation this work is meant to contribute to and what sort of contribution it is meant to make to that conversation.

While personal identity is not the main focus of this work, it will be useful to frame the survey around it, as many of the works I deal with here are unified by a concern for the relevance of questions of personal identity to abortion (and other bioethical issues). And to that end, it will be useful to begin with a few brief general observations about personal identity.

The term “personal identity” is traditional, and I will defer to tradition and use it freely. However, the term can be somewhat misleading, as Jeff McMahan points out. For, it suggests that the question of personal identity is a question about the identity-conditions of persons, and by doing so, it begs an important question. When we ask the question of personal identity, what we are really after is, roughly, the identity- and persistence-conditions of the sort of entity I refer to when I use the first-person pronoun – that is, to beings like us. Thus, to frame the question as a question about the identity-conditions of persons is to presume in advance that beings like us are essentially persons. But as we will see shortly, this claim is open to dispute.

As McMahan also points out, and as the previous paragraph suggests, questions about our identity- and persistence-conditions are closely tied to questions about what kinds of things we essentially are. In general, theories of personal identity will usually take the form “we are essentially Fs, and therefore have the identity- and persistence-conditions of Fs.” But a claim of this form does not settle all the salient questions, for even after it has been shown that we are essentially Fs, an open question might very well remain as to what the identity- and persistence-conditions of Fs are. Thus, a theory of personal identity has two main tasks: to state what we are essentially, and to state the identity-conditions of what we are essentially.

---

8 McMahan 2001, p. 5
9 McMahan 2001, p. 7 et passim
2.1 Marquis

A natural starting point for our survey is what is almost certainly the most influential anti-abortion argument given in the recent literature, namely Don Marquis’ “future-like-ours” argument, which he first put forward in the article ‘Why Abortion is Immoral’ (1989). Marquis’ argument starts with a general account of the wrongness of killing. As Marquis points out, such an account must ground the wrongness primarily in the killing’s effects on the victim, not on the perpetrator or on third parties like the friends and family of the victim. For while there is no doubt that killing often has bad psychological effects on the killer and on third parties, the correlation is too weak and contingent to do justice to how wrong killing really is. Such an account must also do justice to the fact that it is, at least usually, wrong to kill those who have no current desire to go on living, be it because of temporary unconsciousness or temporary, irrational suicidal urges.

But what sort of “wrongmaking” property might an act of killing have even in those circumstances where it does not violate any of the victim’s present desires? Marquis proposes the following: *It takes away his future*. If I was killed right now, all my unfinished projects and unrealized ambitions would thereby be thwarted. Any goods I might have produced or enjoyed in the future would be taken from me. And that, for Marquis, is the reason why killing is wrong (or at rather, one reason – Marquis does not claim, and his argument does not require him to claim, that taking away a valuable future is necessary in order for an act of killing to count as seriously wrong, only that it is sufficient).

Marquis’ account, then, does not make the wrongness of killing contingent on whether the victim currently values his future. The question is rather whether the victim’s future will be (or would be) valuable to the victim once it has arrived, regardless of whether he currently sees its value. And with respect to the ethics of abortion, the account makes the much-discussed question whether the fetus is a *person* largely irrelevant to the question whether (or when) we may kill it. Rather, the question is now whether the fetus, regardless of its *present* properties or capacities, has a valuable future. And it seems clear that a standard fetus does have a valuable future. If it is not aborted, it will, under normal circumstances, be born, grow up, and go on to lead a life broadly similar to yours or mine. In a slogan, the fetus has a *future like ours*. So, if it is wrong to kill you or me because it deprives us of *our* futures, it is just as wrong to kill a fetus.
since that will normally deprive the fetus of a future that is relevantly the same as ours. Under normal circumstances, at least, abortion is morally equivalent to murder.\textsuperscript{13}

2.2 McMahan

Criticisms of the future-like-ours argument often take aim at two substantial assumptions it seems to make about personal identity. First, it seems to assume that the fetus is (at least from an early stage, and possibly from the moment of conception) identical to its adult self – that “I was once a fetus,” as Alexander Pruss puts it.\textsuperscript{14} (Let us give the name “the I-was-once-a-fetus view” to any view to the effect that we begin to exist no later than at conception, regardless of which systematic theory of personal identity, if any, is used to underwrite that view.) If this does not hold, abortion does not deprive the fetus of a future like ours, and therefore does not, after all, have the same “wrongmaking” properties as does killing you or me. Secondly, it seems to assume that identity is what matters. Let’s take these in turn.

At first blush, the first assumption may seem so obvious as to be almost beyond dispute. Pruss gets this point across nicely:

\begin{quote}
That I was once a fetus seems innocuous and obvious. After all, is it not biologically evident that first I was an embryo, then I was a fetus, then a neonate, then a baby, then a toddler, then a child, then an adolescent, and then an adult? Does not my mother talk of the time when she was “pregnant with me,” thereby implying that it was I who was in her womb when she was pregnant? Is not the sonogram of my daughter the sonogram of that same daughter of mine who was eventually born?\textsuperscript{15}
\end{quote}

But there are respectable theories of personal identity on which the I-was-once-a-fetus view turns out false, notably the psychological theories – i.e., very roughly, those theories which say that two beings like us are the same just in case there is a sufficient degree of psychological sameness or continuity between them. Proponents of psychological theories disagree about the type and extent of psychological continuity needed for personal identity, but these disagreements need not concern us very much here, for it is plausible that for most of its gestation, the fetus has no psychological continuity with any future entity. It presumably has no mental states at all for part of its development, and \textit{a fortiori} no psychological connections with anything at all. And, even long after it has acquired the capacity for simple mental states, such as pains and pleasures, it surely lacks the psychological sophistication (and, in particular,

\textsuperscript{13} Marquis 1989, p. 192 \textit{et passim}
\textsuperscript{14} Pruss 2011
\textsuperscript{15} Pruss 2011, pp. 19 & 20.
the self-consciousness) to be psychologically connected in any significant way with any future entity.

Thus, psychological theories of personal identity very strongly suggest that I did not come to be at conception, but at a much later point – perhaps even after birth. To be sure, a certain organism did come to be at my conception, and that organism would eventually be animated by me, but I was not yet present. Before a certain point the fetus is not a being like us at all, but a sort of empty vessel. Nor, therefore, does it have a future like ours. Before the fetus is animated by a being like us, abortion does not really amount to depriving a certain already-existing being like us of a valuable future, but rather to bringing it about that a certain being like us never comes into existence at all. And this obviously seems far less morally objectionable than killing that same entity. If it wasn’t, turning down an offer of potentially fecund sexual intercourse would be tantamount to murder. (There are, of course, other respectable theories of personal identity that fit far better with the I-was-once-a-fetus view. In particular, there is animalism – i.e., roughly, the view that we are essentially organisms, and therefore have the identity- and persistence-conditions of organisms.)

The second assumption may seem as obvious as the first, but has famously been challenged by Derek Parfit and others. The future-like-ours argument seems to presume that to deprive the fetus of its future automatically constitutes a harm to the fetus simply because that future is its future – i.e., that the mere numerical identity between the fetus and its adult self is sufficient to endow the fetus with an interest in what happens to its adult self. In short, the future-like-ours argument seems to presume that identity is the relation that grounds rational egoistic concern. But on closer inspection, it seems that identity is simply correlated with the relations that ground rational egoistic concern – often called the prudential unity relations – and can, at least in principle, come apart from them.

One standard way to show this is cases of branching. Suppose I use a teleportation machine that works by disintegrating me and then reconstituting an exact, qualitatively identical copy of me at my destination. The logic of identity tells us that I and the clone cannot be one and the same. For it is possible that I and the clone not be identical. (Consider, e.g., the case where, because of a glitch, the machine fails to disintegrate me after the copy has been made. Then I and the clone will exist simultaneously, as obviously distinct entities.) But identities are

---

16 Note that pro-choice arguments based on these sorts of considerations predate the future-like-ours argument, though they are today often framed as responses to it. Tooley 1972 contains a notable example.
17 Parfit 1984, pt. 3
necessary\textsuperscript{18}, so if I \textit{could} have been distinct from my clone, it follows that I actually am. And yet, it seems eminently rational for me to be as concerned for my clone as I would be for myself, on the exact same grounds. I have the exact same sort of interest in what happens to the clone as I would normally have in what happens to \textit{me}. And surely the reason for this is that, although I am not identical to the clone, I have the right sort of \textit{continuity} with it – chiefly, it seems, psychological continuity.\textsuperscript{19} Surely, that is, my egoistic concern for the clone is warranted largely by the fact that we have (or will have) so many of the same memories, beliefs, desires, and so on, in some appropriate sense of “same.” (Of course, the branching cases also suggest that this sort of continuity cannot be sufficient for personal identity, since they seem to show that it is possible for me to be continuous in this way with another, distinct individual.)

Thus, it seems that whether I have an interest in what happens to a certain future person is not determined (at least not exclusively) by whether I am numerically identical to him, but (at least to a large extent) by whether I have this sort of continuity with him. But if identity is not what grounds our interests, it is not given that the fetus has any particular interest in its future. And if psychological continuity is one of the most important among the relations that \textit{do} ground our interests, it seems that the fetus’s interests in its own future will be weak to nonexistent, for reasons already covered. So, even supposing, for the sake of argument, that the fetus is identical to the adult, it is not given that to deprive it of its future is to harm it, at least not to the same extent or in the same way as taking away \textit{my} future is to harm \textit{me}.

These sorts of concerns lie at the heart of the next work I want to discuss – Jeff McMahan’s \textit{The Ethics of Killing} (2001). This is a long and rich work, and I will not even begin to try giving a full summary of it here. Rather, I will simply note a few highlights that will be particularly relevant for what lies ahead.

\textit{The Ethics of Killing} begins with a long chapter on personal identity, culminating in McMahan’s explanation and defense of what he calls the \textit{embodied mind account} of personal identity. The embodied mind account can be thought of as a sort of compromise between psychological accounts and animalism, both of which McMahan criticizes at length.\textsuperscript{20} On the embodied mind account, we are essentially embodied minds, and therefore have the identity- and persistence-conditions of embodied minds. In a nutshell, these conditions go as follows: Two embodied minds are the same embodied mind just in case they have the same basic

\textsuperscript{18} This result was famously proved by Saul Kripke. For a more thorough introduction, see e.g. Burgess 2014.

\textsuperscript{19} Note that I here use the term ‘psychological continuity’ in its loose, colloquial sense, not in the special, restricted sense in which Parfit (1984, p. 206) and others have used it.

\textsuperscript{20} McMahan 2001, ch. 1, scts. 3 & 4
psychological capacities, realized by the same physical base (the brain, in our case).\textsuperscript{21} On the embodied mind account, the relation between me and my organism is not one of numerical identity, as animalism would have it, but nor is it one of total, non-overlapping distinctness, as Cartesian dualism (and perhaps certain other psychological theories) would have it. Rather, the relation is one of parthood and constitution: I am a proper part of my organism, and constituted by it, bearing much the same relation to it as a bronze statue is often said to be to bear to the lump of bronze out of which it is fashioned.\textsuperscript{22} I am neither an organism nor a mind, but a minded phase in the history of an organism. As should be clear, the embodied mind account vindicates our first objection to the future-like-ours argument: It entails that we come into existence only after our organisms’ brains have developed so much that they can support a sufficiently advanced set of psychological capacities, which plausibly happens quite late in the game – sometime between twenty and twenty-eight weeks after fertilization, McMahan argues.\textsuperscript{23}

However, McMahan also thinks that identity is not what matters, leading him to formulate a notion of what he calls time-relative interests. On the simplest and most natural way to think of interests, the extent to which I have an (egoistic) interest in a certain future is a function of two things: (a) whether it is my future, which is very plausibly an “either-or” measure, and (b) how good said future will be for the entity whose future it is, which is very plausibly a degreed measure. But as already alluded to, (a) seems to presume that identity is what matters – that mere identity between me and my future self is both necessary and sufficient for my rational egoistic concern for that future self. Given that identity is not what matters, a new notion of interests is therefore called for. On this new notion, the extent to which I have an (egoistic) interest in a certain future is a function of two degreed measures: (a’) the strength of the prudential unity relations between my current self and the future entity in question, and (b) how good said future will be for the entity whose future it is.\textsuperscript{24} And this vindicates the second objection to the future-like-ours argument. For it is plausible that, even after we have come to animate our organisms, our time-relative interests remain weak for quite some time. (This is certainly the case on McMahan’s account of the prudential unity relations, on which “the basis for an individual’s egoistic concern about the future—that which is both necessary and sufficient for rational egoistic concern—is the physical and functional continuity of enough of those areas of the individual’s brain in which consciousness is realized to preserve the capacity

\begin{thebibliography}{9}
\bibitem{McMahan2001} McMahan 2001, \textit{et passim}
\bibitem{McMahan2001ch1sect55} McMahan 2001, ch. 1, sect. 5.5
\bibitem{McMahan2001pp267268} McMahan 2001, pp. 267 & 268
\bibitem{McMahan2001p80} McMahan 2001, p. 80 \textit{et passim}
\end{thebibliography}
to support consciousness or mental activity."\(^{25}\) Thus, even what McMahan calls “late abortion” (that is, abortion performed after the fetus has come to constitute a being like us) is not morally equivalent to killing a normal adult.\(^{26}\)

McMahan argues that the badness of death and the wrongness of certain kinds of killing is explained by the fact that they thwart our time-relative interests. However, he does not think that time-relative interests alone can furnish us with a full ethics of killing. More specifically, he does not think that time-relative interests alone can account for the wrongness of killing persons – i.e., roughly, self-conscious and rational entities. This is, in particular, because such an account is hard to square with what he calls the Equal Wrongness Thesis – the claim that, in some fundamental sense, all persons have an equal right to life.\(^{27}\) Persons, McMahan argues, have an intrinsic and inalienable value, and killing persons is wrong because it destroys – and is an affront to – the goodness inherent in the person himself, rather than because it destroys or prevents the coming-into-existence of the goods that his life would otherwise have contained:

\[\text{[A] person, a being of incalculable worth, demands the highest respect. To kill a person, in contravention of that person's own will, is an egregious failure of respect for the person and his worth. It is to annihilate that which is irreplaceable, to show contempt for that which demands reverence, to assert a spurious authority over one who alone has proper authority over his own life, and to assume a superior position vis-à-vis one who is in reality one's moral equal. Killing is, in short, an offence against what might be called a requirement of respect for persons and their worth. Indeed, because killing inflicts the ultimate loss—the obliteration of the person himself—and is both irreversible and uncompensable, it is no exaggeration to say that it constitutes the ultimate violation of the requirement of respect.}^{28}\]

Further, if the value of persons is to be intrinsic and universal in this way, it must be had in virtue of “those properties, or some subset of those properties, that relevantly differentiate persons from animals.” And these properties seem to be certain psychological capacities, notably the capacities for self-consciousness and rational and moral agency.\(^{29}\) Thus, McMahan’s ethics of killing is “two-tiered,” as he puts it: Killing persons is wrong because it violates the requirement of respect, while killing non-persons is wrong because (or to the extent that) it thwarts their time-relative interests in continuing to live.\(^{30}\)

\(^{25}\) McMahan 2001, pp. 67 & 68
\(^{26}\) McMahan 2001, ch. 4, sect 2
\(^{27}\) McMahan 2001, ch. 3, sect. 3.1
\(^{28}\) McMahan 2001, p. 242
\(^{29}\) Ibid.
\(^{30}\) McMahan 2001, p. 245
2.3 Boonin

Our next stop on the tour is David Boonin’s *A Defense of Abortion* (2002). The book is structured as a series of responses to anti-abortion arguments, but while giving these responses, Boonin also presents and defends several positive claims.

The most substantial of these positive claims is Boonin’s account of moral status and the wrongness of killing, which he begins to develop in response to the future-like-ours argument. Unlike McMahan, Boonin does not take aim at the argument’s presumptions about personal identity. Indeed, one of the first things he does is grant the truth of the I-was-once-a-fetus view, in this memorable passage:

> On the desk in my office where most of this book was written and revised, there are several pictures of my son, Eli. In one, he is gleefully dancing on the sand along the Gulf of Mexico, the cool ocean breeze wreaking havoc with his wispy hair. In a second, he is tentatively seated in the grass in his grandparents’ backyard, still working to master the feat of sitting up on his own. In a third, he is only a few weeks old, clinging firmly to the arms that are holding him and still wearing the tiny hat for preserving body heat that he wore home from the hospital. Through all of the remarkable changes that these pictures preserve, he remains unmistakably the same little boy.

> In the top drawer of my desk, I keep another picture of Eli. This picture was taken on September 7, 1993, 24 weeks before he was born. The sonogram image is murky, but it reveals clearly enough a small head tilted back slightly, and an arm raised up and bent, with the hand pointing back toward the face and the thumb extended out toward the mouth. There is no doubt in my mind that this picture, too, shows the same little boy at a very early stage in his physical development. And there is no question that the position I defend in this book entails that it would have been morally permissible to end his life at this point.31

Instead, Boonin takes aim at Marquis’ general account of the wrongness of killing, arguing that we are better served by a desire-based account – i.e., roughly, an account on which killing beings like us is wrong because it thwarts certain desires of ours. (This is a bit of a simplification, as Boonin also reads Marquis’ account as a desire-based one, albeit more complicated than his own. More specifically, Boonin reads Marquis’ account as boiling down to the claim that killing is wrong because it thwarts either present or future desires. This reading of Marquis is also a big part of Boonin’s reason for regarding his own account as superior. For as we will see, that account grounds the wrongness of killing *only* in its thwarting of present desires, and is therefore, Boonin argues, more parsimonious than the alternative.32) Chief among these, of course, is the desire to go on living, but since very many of our other desires

---

31 Boonin 2002, pp. xiii & xiv
32 Boonin 2002, pp. 73 & 74
are contingent on continued life, killing thwarts many other desires as well. Killing me now, for instance, would not just thwart my desire to go on living, but also my desire to finish writing this chapter, among many others.

As already emphasized, an obvious problem for any account that seeks to ground the wrongness of killing in its thwarting of present desires is that it seem to give us no reason not to kill people who have no present desire to go on living, even though it is clear that killing such people can often be just as wrong as killing anyone else. The two standard examples here are the temporarily unconscious person and the suicidal, lovelorn teenager.\(^{33}\) The temporarily unconscious person, being unconscious, seems to have no present mental states at all, and therefore no desires. The teenager does have desires, but not only do these not include the desire to go on living – they actually include the desire not to go on living.

Boonin, however, argues that his account has a way out of this problem. For on that account, killing is wrong because it thwarts our ideal (as opposed to actual) and dispositional (as opposed to occurrent) desires.\(^{34}\) I will take these notions in turn.

The distinction between occurrent and dispositional desires is a special case of a broader distinction – very many types of mental states can apparently be either dispositional or occurrent, with belief being the most discussed in the literature. In any case, the distinction is easier to illustrate than it is to spell out precisely, so let’s start with an illustration. You very likely believe that Paris is the capital of France (call this proposition P), but it is almost as likely that you were not consciously entertaining this belief before you read this sentence. But now, you are consciously entertaining it. Where you previously merely dispositionally believed that P, you now believe it occurrently. In other words, I have a belief (or other mental state) occurrently when I am consciously entertaining it, and dispositionally when I have it but am not consciously entertaining it. As Robert Audi points out, we can shed further light on the distinction through an analogy between the mind and a computer. A belief (“file”) is then dispositional when it is stored in the computer’s memory and occurrent when it is up on the screen.\(^{35}\)

Now, much as it is possible to believe dispositionally, it seems possible to desire dispositionally. An hour ago you probably were not consciously entertaining your desire to go on living, but surely we would still say that you did have that desire at that time. Further, and as Boonin points out, it seems clear that the general prima facie prohibition on violating the

\(^{33}\) Boonin 2002, p. 64  
\(^{34}\) Boonin 2002, sects. 2.8.3 & 2.8.4  
\(^{35}\) Audi 1994, p. 420
desires of others includes a prohibition on violating dispositional desires. For example, the infidelity of an unfaithful husband is plausibly wrong at least partly because it violates his wife’s desire that he stay faithful to her. But if it should turn out that the wife had not been consciously entertaining this desire at any time at which the husband was actually cheating on her, that would do nothing to excuse his infidelity. Her desires become no less morally weighty simply because they are dispositional. And this seems to save the temporarily unconscious person. For while the temporarily unconscious person has no occurrent desires, he seems to have plenty of dispositional desires, much as he still seems to have plenty of dispositional beliefs. (For example, surely I do not lose the belief that Paris is the capital of France every time I fall into a deep sleep, only to form that belief again when I wake up the next morning.) Surely, in fact, the temporarily unconscious person has all the same dispositional desires he had when he fell unconscious; and included among these is presumably a desire to go on living.

What about the suicidal teenager? Here, Boonin argues, the distinction between actual and ideal desires comes to the rescue. This distinction, too, is best introduced with an illustration. Suppose I come home parched one day, strongly desiring to get a drink of water from the tap. What I don’t know is that terrorists have just poisoned the city’s water supply, and that I will die if I get that drink of water. If I knew that, I would, of course, no longer desire to drink from the tap. Here, Boonin would say, I have an actual desire to get drink of water, but not an ideal desire for it. For while I do desire to get a drink of water, I only do so because I find myself in non-ideal epistemic circumstances.

In short, then, my actual desires are simply those desires I actually have (either dispositionally or occurrently) at any given moment, while my ideal desires are those actual desires idealized. They are what I would desire at that moment if the circumstances were ideal, given the desires I actually have and my reasons for having them. Boonin gives no systematic account of what the ideal circumstances look like, but argues that they include knowledge of all the relevant facts and a more nebulous attribute that involves having the proper general outlook on things. Thus, e.g., my desire for a drink of water is non-ideal because it is contingent on ignorance of one of the relevant facts, namely the fact that the water supply has been poisoned. And once again, it seems plausible that the general prima facie prohibition on thwarting desires is stronger when (or even, perhaps, present only when) those desires are ideal.

---

36 Boonin 2002, pp. 67 & 68
37 Note that the distinction between actual and ideal desires, unlike that between occurrent and dispositional desires, is not an exclusive one. Some of my desires may well be both actual and ideal, since I may well have actual desires that would “survive” the idealization process.
38 Boonin 2002, pp. 71 & 72
For example, under normal circumstances, it would be wrong to thwart my desire for a drink of water, e.g. by knocking a glass of water out of my hand. But in the case where the water is poisoned and my desire for water is non-ideal, such an act would clearly be unproblematic, and even laudable or obligatory (at least if done with the intention of preventing me from getting poisoned). And this seems to save the suicidal teenager. For it is plausible that if the circumstances were ideal – if he was more cool-headed and better informed – the teenager would realize that his present troubles are not quite as dire as he thinks, and certainly not worth killing himself over. He would no longer desire to die, and would, in fact, want to go on living.

Now, unlike Marquis’ account, Boonin’s account places no prohibition on killing beings that have no actual desires. For as already suggested, dispositional desires seem to be a subspecies of actual desires, and only beings with actual desires can have ideal desires in Boonin’s sense, since ideal desires are nothing but actual desires idealized. Consequently, for all the desire account says, the fetus does not acquire the same sort of moral status as you or me until it has formed some actual desires. So, a fortiori, the fetus does not acquire the same sort of moral status as you or me before it has formed the capacity to form desires. But, Boonin argues, this happens no earlier than the start of organized cortical brain activity, around 25 weeks after fertilization at the earliest. And since the vast majority of abortions are performed before this point, drawing the line here amounts to embracing a pro-choice position.\(^{39}\)

I will briefly touch on two other points from Boonin that are useful for our purposes here. One is his discussion of methodology. Boonin uses – and defends his use of – a broadly Rawlsian method of reflective equilibrium (RE).\(^{40}\) Roughly speaking, this method involves examining our intuitions about concrete cases, often with the aid of thought experiments, and “[attempting] to develop a credible moral theory that would serve to unify and underwrite these various judgments. We ask: What sort of more basic principle or set of principles would have to be true in order for these sorts of more particular judgments to prove to be correct?”\(^{41}\) Boonin argues (and I tend to agree) that RE is the best starting point of moral inquiry, if only because it is not clear where else we could start. But as he also points out, even the RE skeptic has a powerful reason to use that method when arguing for pro-choice views, simply because it is such a widespread method among those who argue for pro-life views. Given this, even if RE generally has no particular tendency to deliver truth or justification, an RE-based case against

---

\(^{39}\) Boonin 2002, sct. 3.6

\(^{40}\) Boonin 2002, sct. 1.2

\(^{41}\) Boonin 2002, p. 10
the pro-life position will show that the arguments of the pro-lifers fail on their own terms.\textsuperscript{42} It might be added that this point goes the other way too: Pro-choicers seem no less dependent on RE than pro-lifers, and so, an RE-based case against abortion will, if nothing else, show that the case for the pro-choice position fails on its own terms.

Finally, it is worth briefly going over Boonin’s discussion of what he calls “the species essence argument,”\textsuperscript{43} for among all the arguments or families of arguments that Boonin deals with, this is the one that most closely resembles the argument underlying the substance view. (More on that very shortly.) Boonin’s treatment of the argument is quite perfunctory, but it does raise several challenges that any proponent of it must meet. His main spokesman for the species essence argument is Stephen Schwarz, who argues that “the basic inherent capacity to function as a person” is an essential property of human organisms.\textsuperscript{44} This capacity may be blocked, and therefore not immediately exercisable, but it will nevertheless be there.

Boonin has two objections to the species essence argument. Firstly, he questions whether all human organisms really do have the capacity for personhood. Irreparably brain-damaged people, he points out, certainly seem to lack it. The proponent of the species essence argument could say that this is one of the cases where the capacity is present but blocked, but on what grounds? It won’t do simply to say that such brain-damaged humans would have been persons if only their brains were undamaged, for surely that is just to say that they would have been persons if only they had the brains for it. But this seems to be true of pretty much any organism:

\textit{But then it is difficult to see why we should not also call the spider crawling up my window a person. If he were able to develop a big enough brain, he too would be able to function as a person, so he is simply a person whose capacity is blocked by the fact that he will never have a large enough brain.}\textsuperscript{45}

Secondly, even if it should turn out that all human organisms do have the capacity for personhood, Boonin questions the moral relevance of this fact. For the claim that (say) the zygote and the brain-damaged person have a right to life in virtue of their capacity for personhood sounds suspiciously like the claim that they have the same moral status an \textit{actual} person simply because they have the \textit{potential} for personhood. But as should be immediately clear (and as Boonin later argues at greater length\textsuperscript{46}), this is an implausible claim, for in general,

\textsuperscript{42} Boonin 2002, sct. 1.2.2
\textsuperscript{43} Boonin 2002, sct. 2.2
\textsuperscript{44} Schwarz 1990, p. 101
\textsuperscript{45} Boonin 2002, p. 24
\textsuperscript{46} Boonin 2002, sct. 2.6
potential F-hood does not automatically bestow the rights and privileges associated with actual F-hood. (For example, the fact that the Vice President of the United States is a potential President of the United States does not endow him with the privileges of the actual American President.)

2.4 The substance view

This brings us, at last, to the substance view. As already noted, the substance view is based on some very old ideas, but in its current form, it is a relative newcomer to the debate. As far as I can tell, it was first formulated and defended in the 1990s by Patrick Lee and Stephen Schwartz. Other proponents include Christopher Kaczor, Christopher Tollefsen, J. P. Moreland, and Francis Beckwith. At the heart of this view lies an anti-abortion argument that Lee has dubbed “the argument from substantial identity” (henceforth, the ASI), and which he schematizes in the following way:

(1) You and I are intrinsically valuable (in the sense that makes us subjects of rights)
(2) We are intrinsically valuable because of what we are (what we are essentially)
(3) What we are, is each a human, physical organism
(4) Human physical organisms come to be at conception
(5) Therefore, what is intrinsically valuable (as a subject of rights) comes to be at conception

The basic strategy of the ASI, then, is much the same as that of the future-like-ours argument: identify the attribute that gives you and me our right to life, then argue that this attribute is also present in the fetus.

All of the ASI’s premises seem plausible in isolation, and several of them have already been brought up. Premise (1), I take it, simply states that it is very seriously wrong to kill beings like you or me, and that it is wrong because beings like you or me are, in some important sense, intrinsically valuable and worthy of respect. We earlier saw McMahan endorse something like it, in the form of his ethics of respect (though McMahan would, of course, disagree with Lee about exactly which beings are intrinsically valuable and worthy of respect). Premise (2) has plausibility on at least some conceptions of what we are essentially. For example, if we are

---

47 See Lee 1996/2010 and Schwarz 1990
48 See e.g. Kaczor 2015, Tollefsen 2004, Moreland 1998, and Beckwith 2003
49 Lee 2004, p. 250
essentially persons, there is clearly a strong case to be made that we have our moral status in virtue of what we are essentially. Premise (3) is a restatement of animalism. Premise (4), I take it, simply says that our organisms come to be at conception, without claiming that we come to be at conception, and should therefore have plausibility for anyone, regardless of which theory of personal identity he adopts.

Observant readers may already have spotted the problem here. While each of these premises may be defensible in isolation, they clash when taken together. More specifically, while both (2) and (3) can be held with some plausibility, it is hard to see how they can plausibly be held together. For there seems to be no metaphysics of beings like us – no theory of personal identity – that renders both of them tenable. If we are essentially persons or minds or embodied minds, then (2) is plausible but (3) is not. If, on the other hand, we are essentially organisms, then (3) follows trivially, but (2) loses its plausibility. For why should the mere fact that we are organisms of a certain kind endow us with any intrinsic value to speak of? (This is simply a rehearsal of the speciesism and unmotivatedness objections discussed in the previous chapter.)

The main challenge for the ASI proponent is therefore to formulate and defend a metaphysics of beings like us that renders both (2) and (3) plausible. This is where the substance view comes in. Since much of the purpose of this work is precisely not to defend or presuppose the substance view, the following won’t dwell on how its proponents do defend it. Rather, the point will simply be to describe what they believe. For the sake of completeness, though, it may be worth noting that arguments for the substance view (like arguments for other theories of personal identity) often take the form of inferences to the best explanation. The substance view, its proponents claim, can account for our intuitions about the nature of human persons in a more satisfactory way than any other theory on offer. In particular, Beckwith has argued that the substance view has the best explanation for the attitude we instinctively take (and ought to take) towards a relative with full retrograde amnesia, affectionately dubbed “Uncle Jed.”

On the substance view, we are essentially human organisms, and to be a human organism is to be a certain type of substance, in a broadly Aristotelian sense of the word. In other words, it is to be matter (hyle) unified and informed by a certain form, nature, or essence (morphe). This is not the place to give a full-fledged description of hylomorphism (let alone a defense of it), but a few words of explanation will be helpful. Among philosophers today,

50 Beckwith 2003, pp. 135 & 136
51 For that, see e.g. Oderberg 2007 or Feser 2014, ch. 3
52 Note, by the way, that I take hylomorphism and hylomorphic dualism to be distinct notions. I define hylomorphism as the view that substances in general are unions of hyle and morphe, and hylomorphic dualism as
one sometimes finds a certain deflationary construal of hylomorphism (encouraged, to be fair, by a diet of bad examples and by the difficulty of extricating oneself from the all-consuming influence of the mechanistic worldview we have inherited from the early moderns). On this construal, matter is something like fundamental particles, and form is simply the way in which these fundamental particles happen to be put together. This sort of view is a basic presumption of McMahan’s case against hylomorphic dualism, and a motivation for Bernard Williams’ view (cited by McMahan) that hylomorphism is nothing but “a polite form of materialism.”

But on the traditional view, at least, form is something much more than that. On that view, the form of a substance is the principle that accounts for the fact that it has any actual properties at all, and therefore, in the last analysis, for its very existence. One way to illustrate the cleavage between the deflationary and traditional views is through their profoundly different conceptions of matter. On the traditional view, matter without form (or “prime matter”) is not to be equated with those fundamental particles that are not proper parts of any larger object. Rather, it is a mere abstraction that could not possibly exist in reality, since it is an undifferentiated “gunk” that has every attribute potentially and no attribute actually. Thus, even fundamental particles (if such there be) are unions of matter and form rather than formless matter.

Further, on the traditional view, the forms of living organisms not only imbue them with their current properties – they also guide their development. It is the fact that the development of the substance is continuously guided towards the same goal by the same form that explains how it can remain one and the same thing even as it undergoes dramatic changes in attributes and composition. Hence, the form can be thought of as the “blueprint” for an ideal specimen, and the development of the organism, when all goes well, as the realization of that blueprint.

Proponents of the substance view often cash out this thought in terms of a doctrine of basic capacities. On this view, our forms endow us with capacities to do all the things that typify a human organism at peak performance. In particular, then, any human organism will have the capacity for self-consciousness, moral and rational agency, and so on, since these are all features of a mature human organism that is working as it should. While human organisms obviously differ radically in the extent to which they are currently able to exercise these capacities, the capacities themselves are still present, and equally present, in all of them. Here,

the view that beings like us are unions of hyle and morphe. Thus, it is in principle possible (though rarely or never done in practice, as far as I know) to be a hylormophist without being a hylomorphic dualist, so long as one denies that beings like us are substances. In particular, it might be claimed that beings like us are mere attributes of and/or phases in the histories of organism-substances.

53 McMahan 2001, ch. 1., sct. 2.1; Williams 1986, p. 197
54 Metaphysics, bk. IX, pt. 7
55 See e.g. Beckwith 2003, p. 50
temporary unconsciousness can again serve as a good illustration. If you or I fell temporarily unconscious, we would lose our ability to immediately exercise our capacities for, e.g., moral and rational agency, but we would not lose those capacities themselves. The general explanation for how this can be seems clear enough: When we fall temporarily unconscious, the ground or basis of those capacities – the thing about us that enables us to act morally and rationally – has not been lost or damaged, but simply temporarily put out of commission, and as long as we still have the basis itself, we still have the capacities.

Where the controversy lies is in what is identified as the basis, which also has implications for who has the capacities in question. Many bioethical “liberals” identify the basis with the brain. (As we have seen, both McMahan and Boonin certainly take views to this effect.) And since some human organisms (fetuses in particular) have underdeveloped or undeveloped brains, it follows that they lack the relevant capacities. But on the substance view, the ultimate basis of our rational capacities is not our brains, but our forms, natures, or souls. (On an Aristotelian anthropology, these terms are equivalent and interchangeable.) And unlike the brain, these forms, natures, or souls are by definition had – and had equally – by all human organisms, since it is precisely in virtue of having these exact forms that they count as human organisms at all. The brain simply determines the extent to which these capacities are currently exercisable. And as we have seen, capacities do not seem to lose their moral weight simply because they are not easily or immediately exercisable. If they were, the temporarily unconscious person would have no moral standing, even given that he still has all the right capacities. Thus, to the extent that our moral status is grounded in our capacities for certain sorts of rational agency – and again, I think it’s pretty clear that both Boonin and McMahan think something like this – it follows that all human beings have the same moral status. And indeed, since being rational plausibly just is to have the right sort of capacity for rational agency (a point I will elaborate on in the next chapter), it also follows that all human beings are rational.

2.5 Liao

The final point on the agenda is S. Matthew Liao’s genetic account of rightholding. As far as I know, Liao first proposed this account in his doctoral dissertation, and then developed it in his 2010 paper ‘The Basis of Human Moral Status’. I only learned of Liao’s genetic account after having independently come to a similar view, but am greatly indebted to him for how my views on the matter have developed since then.

---

56 Later reworked and published as Liao 2015.
The starting point of Liao’s account is the widespread intuition (previously discussed in Chapter 1 of this work) that any member of the human species has a certain moral standing—and, in particular, certain moral rights—in virtue of being a member of the human species. However, it has proven difficult to justify this intuition, and, in particular, to defend it in a non-speciesist way. That is, it is hard to find a plausible candidate for a sufficient criterion of moral status which is had by all human organisms, but which is also distinct from mere membership in the human species. Liao proposes one such criterion, namely having what he calls the genetic basis for moral agency.

Notably, Liao offers no positive arguments for the genetic account, and questions whether such arguments are even necessary or possible. Rather, what he does is explain what the genetic account actually says, as well as defending it against certain objections.

Liao’s account says that we have our moral status in virtue of having the genetic basis for moral agency, while I will suggest that we have our moral status in virtue of having the genetic basis for rationality. However, I am not convinced that this is a substantive difference. As we will see in the next chapter, I conceive of rationality (in the sense that will interest us here) as consisting of a sufficient sensitivity to reasons in general, while Liao seems to conceive of moral agency as a sufficient sensitivity to moral reasons specifically, defining it as “the capacity to act in light of moral reasons.” Thus, our accounts would differ in their verdicts about concrete cases only if there were (or could be) beings who had the genetic basis for rationality, but not the genetic basis for moral agency. But it is not clear whether such a being is possible. Moral agency seems to be so tightly interwoven with other types of rational agency that it is hard to see how one could have the basis for the latter while completely lacking the basis for the former. That said, even if Liao’s account and mine do not differ in their concrete verdicts, I am open to the possibility that Liao’s account may be more theoretically virtuous because more salient. In other words, it may be that the basis for rationality is not even in principle separable from the basis for moral agency, but that it imbues us with our moral status only in virtue of its tendency to imbue us with moral agency. In any case, though, since Liao’s conception of moral agency is so similar to my conception of rationality, it should be easy enough to reformulate my defense of my rationality-based account into a defense of an analogous moral-agency-based account, if that turns out to be the way to go.

57 Liao 2010, p. 160
58 Liao 2010, p. 159 et passim
59 Liao 2010, p. 163
60 Liao 2010, p. 164
It is also important to note that Liao takes the ultimate moral relevance of the genetic basis for moral agency to lie in the fact that it gives us a functional basis for moral agency, i.e. in the fact that it is a “set of physical codes that generate moral agency.”\textsuperscript{61} Thus, his account does not have the undesirable implication that only entities with genomes (let alone entities with \textit{human} genomes) can have the sort of moral status that we do. Non-carbon-based life forms, advanced artificial intelligences, and even immaterial agents like angels or demons, could all (actually or hypothetically) have the same moral status as we do, on the same grounds, so long as they had something functionally equivalent to our genomes, i.e., some set of codes or other that generated the same sort of moral agency our genome generates in us. (As we just saw, Liao does say some things which suggest that the functional basis must be \textit{physical}. But if the existence of non-physical moral agents is metaphysically possible – a claim which is, of course, open to dispute – I see no reason to include this restriction.) However, since the main goal of Liao’s account is to explain how membership in the human species endows moral status, and since all the rational agents we know of are entities with genomes, he focuses his discussion on the genetic case.\textsuperscript{62} I agree with Liao that any entity that has something functionally equivalent to the human genome thereby also gains the same moral status as any human organism, and I believe that my account makes ample room for this belief. We will see how in the next two chapters.

It should be clear that Liao’s account is not a speciesist account. As we just saw, it does not restrict our moral status or its grounds to carbon-based life forms, let alone to members of the human species. Relatedly, the genetic account is not a species-norm account – that is, it does not make the fallacious inference from the claim that humans \textit{generically} have a certain rightsmaking property to the claim that they \textit{universally} have the corresponding right.\textsuperscript{63} For as Liao argues, all (or nearly all) human organisms have the genetic basis for moral agency.\textsuperscript{64} (Moral agency itself, of course, seems to only be had generically, but the genetic account does not ground our moral status in our moral agency itself, but in its genetic basis.)

Nor, Liao argues, is the genetic account a potentiality account. That is, it does not make the fallacious inference from the claim that all humans \textit{potentially} have a certain rightsmaking property to the claim that they all \textit{actually} have the corresponding right. Rather, it grounds human moral status in a property that all humans actually have, namely the genetic basis for

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{61} \textit{Ibid.}
\item \textsuperscript{62} Liao 2010, p. 169
\item \textsuperscript{63} Liao 2010, p. 162
\item \textsuperscript{64} Liao 2010, p. 165-168
\end{itemize}
\end{footnotesize}
moral agency.\textsuperscript{65} (In some cases, of course, the genetic basis is expressed imperfectly or not at all. But on Liao’s account, it is the genetic basis, not its expression, that is the ultimate ground of our moral status.)

Nor, finally, does the genetic account ground our moral status in an occult or ontologically contentious property, such as having an immortal soul or being made in the image of God. Rather, it grounds our moral status in a perfectly ordinary biological property which is in principle detectable through perfectly ordinary empirical means – i.e., it satisfies one conjunct of what Liao calls the Species Neutrality Requirement.\textsuperscript{66} (The other conjunct being that it must be possible for the property in question to be had by non-human entities. As we have seen, the genetic account satisfies this requirement too.)

\textbf{2.6 Looking ahead}

We should now have a better idea of what proponents of the substance view and their pro-choice interlocutors believe. At the end of the section 2.4, I suggested a certain view of where the basic disagreement between them lies. On this view, the disagreement does not directly concern personal identity, but rather the ultimate basis of our rightsmaking, rational capacities. Therefore, and as already suggested in the previous chapter, the question whether we can simulate the substance view boils down to the question whether we can identify a basis for our rational capacities other than our brains or our Aristotelian forms – one whose existence is as uncontroversial and naturalistically acceptable as the former, but which yields the same sort of strong human moral equality as does the latter. If something like Liao’s genetic account is elaborated and positively defended in a certain way, it can, I think, be seen as proposing one such alternative basis, namely the human genome. The next two chapters will be devoted to giving such an elaboration and positive defense.

\textsuperscript{65} Liao 2010, sect. 3
\textsuperscript{66} Liao 2010, p. 160
Rationality and Its Genetic Basis

The first premise of our argument asserts that all human organisms have the genetic basis for rationality. But what is it to have the genetic basis for rationality? And what, for that matter, is rationality?

This chapter will be devoted to answering these two questions. In the process, the first premise of our argument will also be defended.

3.1 Rationality

3.1.1 Rationality as reasoning

To be rational, I take it, is roughly to reason – i.e., to be motivated to do things by reasons. I will take the notion of a reason as a primitive, but a few things can be said by way of characterization. In general, a reason is an entity (I won’t commit myself to any particular account of what they are – perhaps reasons are ordinary concrete objects like tables and chairs, perhaps they are states of affairs, perhaps they are abstract objects like propositions, or perhaps they are mental objects like beliefs or desires) that makes it more or less sensible for an agent to perform a certain action than it otherwise would have been. Hence, a reason is always a reason to do something. Beyond this, all I want to assume about reasons is that they exist and, again, that they have some sort of normative force – that the existence of a given reason can make it more or less rational for me to act in a certain way than it would otherwise have been. In other words, all I want to assume about reasons is the falsehood of a thoroughgoing metanormative error theory – i.e., roughly, the view that there is no normativity of any kind. Since almost everyone (including, anecdotally, many or most error-theorists about moral normativity) seems to find such a theory wildly implausible, this assumption should not strike anyone as controversial.67

For normal, adult humans like you and me, reasons and rationality pervade our dealings with the world. Indeed, it can easily be argued that, by definition, everything we do is rational, as rationality is precisely what separates the things we do from the things that merely happen to us. Wittgenstein famously asked what is left after we subtract the fact that my arm went up

67 Though in fairness, some authors have recently defended metanormative error theory. See e.g. Streumer 2017.
from the fact that I raised my arm\textsuperscript{68}, and the answer seems to be precisely my \textit{reason} for raising my arm.

I have already described rationality in terms of acting and doing, and for expository clarity, I will continue to do so. It’s worth noting, though, that reasoning need not involve actions or intentions to act. In line with the traditional distinction between practical and theoretical reason, a reason can be a reason to do something in the normal sense (i.e., to perform a certain action), but it can also be a reason to \textit{believe} something. But belief formation, I take it, is rarely or never a matter of performing a certain action. It’s not, e.g., as if I resolve to form the belief that there is a tree in front of me upon seeing a tree.\textsuperscript{69} Rather, under normal circumstances, my seeing the tree just spontaneously forms my belief that there is a tree in front of me, without the intermediation of any intention to form such a belief. From here on out, my talk of “reasons to act” and the like should, unless otherwise noted, be taken as a shorthand for reasons to act \textit{or} reasons to believe.

The notion of rationality is closely tied to the debate about personhood which is central to the abortion debate. As we saw in the previous chapter, a common gambit on the pro-choice side is to argue that our personhood is what gives us our right to life, and that fetuses (and other undeveloped or damaged human organisms) lack personhood. For personhood, on their conception of it, requires certain mental processes and properties – or the capacity for certain such processes or properties – that these organisms do not have. And as we have also seen, the property of personhood is usually defined in such a way that it either entails or is equivalent to the possession of certain rational capacities.

Our axiomatic slogan – “rationality is reasoning” – can be elaborated a bit further. To reason, I take it, is to see and be moved by reasons. But this explication is really just another slogan, and raises two further questions. Firstly, what is it to see a reason? Secondly, what is it to be moved by a reason?

Here, I use “to see” in a broad, metaphorical way – I don’t want to enter into the debate about whether we literally perceive reasons (moral or otherwise) in the same way we perceive colors, smells, and so on.\textsuperscript{70} If R is a reason for me to do X, I see R when I know that (a) R exists or obtains and (b) R is a reason to do X. I am moved by R when I revise my beliefs, preferences,

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{68} Philosophical Investigations §621
\item \textsuperscript{69} It may also very well be that we can have reasons for mental states other than beliefs. For example, it certainly seems that \textit{anger} is more reasonable in some circumstances than in others. If this is the case, I take it that what I have said about reasons for belief also goes for reasons for other kinds of mental states. Just as I rarely or never resolve to form a certain belief before I actually form it, I rarely or never resolve to get angry before I actually get angry.
\item \textsuperscript{70} But for more on that debate, see e.g. Audi 2013.
\end{itemize}
\end{footnotesize}
credences, intentions, or plans in the way R demands, because R demands it. In other words, if R is a reason to believe a proposition p, I update my credence in p; if R is a reason to perform an action A, I upgrade my intention to or preference for doing A; and so on. Note that I speak of updating one’s credences or intentions, not, e.g., of forming the belief that p or forming an intention to do A. This is because reasons are not always sufficient – that is, they are not always such that they, in and of themselves, give us all the motivation we need to actually believe or do a given thing, untrumpable by any other reason to the contrary. R may genuinely give me a reason to do A, but there may be other reasons that give me a stronger reason not to do A. But this clearly does not detract from the fact that I have been moved by R. Me being moved by R does not consist in my forming the intention to do A, let alone actually doing A, but in my taking R into consideration when deciding whether or not to do A.

Reasons, I take it, can also be stronger or weaker than one another. For instance, the fact that I have to go to the store to cash in a million-dollar lottery ticket is a stronger reason for me to go to the store than is the fact that I’m all out of milk, provided that my desire to be a millionaire is stronger than my desire for milk. However, this is another point I will not pursue here, beyond noting that the strength of a reason seems to put further constraints on what it is to be moved by it – the stronger a reason R is to do A, the more heavily I should weigh it when deciding whether or not to do A.

Being moved by a reason must also involve updating one’s beliefs or intentions because that reason demands it. Without this proviso, an entity that didn’t see reasons – or worse, saw them but was not moved by them – would count as rational if it just accidentally happened to always act in the way reason dictated.71

3.12 Actualism

So much for reasoning. But what sort of relation does reasoning bear to rationality itself? When does a person instantiate reasoning in such a way as to qualify as rational?

This may seem like a puzzling question – for after all, isn’t the answer obvious? Isn’t there but one way to instantiate anything? Doesn’t the above entail, or at least strongly suggest, that an agent is rational when it is actually, currently reasoning – when it is actually seeing and being moved by reasons – and non-rational in all other cases?72

---

71 This point, I take it, is closely related to Kant’s distinction between acting in accordance with duty and acting from duty (Groundwork, sc. 1), which is arguably the application of this point to the special case of moral reasons.
72 Note that I take non-rationality and irrationality to be distinct notions. Unlike ascriptions of non-rationality (or arationality, which I take to be synonymous with non-rationality), ascriptions of irrationality usually seem to carry
This theory, call it actualism, is appealing in its simplicity. But if it is taken the way my formulation of it suggests that it should be taken – i.e., as giving the necessary and sufficient conditions for when we can make any ascription of rationality whatsoever – it falls apart quickly. Actualism is too restrictive about which entities do and do not qualify as rational – it makes rationality too fleeting.

For example, I presumably cease to act rationally whenever I fall asleep, so actualism seems to entail that I cease to be rational whenever I fall asleep. But that is clearly false. It might, perhaps, be argued that even when I am asleep, there are some subconscious rational processes going on that count as rational agency. But even if this is the case, we can imagine another scenario in which I have fallen into a deep but reversible coma which has done no damage to my brain, but which has ensured, for its duration, the total cessation of all the higher brain processes that might plausibly be called rational. Even in this case, we would surely still say that I was rational.

The actualist might now try to make his account more inclusive by adding various historical conditions to it – i.e., by claiming that present actual rational agency is not always necessary for rationality, but that having acted rationally in the past or being such that one will act rationally in the future may also be sufficient in some cases. He might, for example, say that an agent is rational at any moment when that agent is either (1) acting rationally, (2) in possession of a past history (possibly a past history of the right sort) of acting rationally, or (3) will (either with certainty or with a sufficiently high degree of probability) act rationally in the future. Note that this account is only an example. The actualist may go for an account that only includes some of these disjuncts (e.g., one which includes only the first two disjuncts, defining rationality as having either a past history of rational agency or present rational agency). But since the above account includes all the historical dimensions – past, present, and future – it is

---

73 ‘Actualism’ is also a term for a theory in the metaphysics of modality. This sort of actualism may be defined as the negation of possibilism – possibilism being, roughly, the view that there are possibilia. (See e.g. Williamson 2015, p. 22 et passim.) As far as I can tell, actualism in this sense bears no interesting relation to actualism in the present sense.
the most inclusive of the historical actualist accounts: Any agent that is deemed non-rational by this account will also be deemed non-rational by any “historical” actualist account of the same type. Hence, if it turns out that even this account is too restrictive – that it fails to deem certain rational agents rational – then any variation on it will have the same problem.

One problem with this account is that past or future rational agency are not sufficient for rationality on their own (and therefore not sufficient for rationality in disjunction with present rational agency either). This is because it is apparently possible to lose and acquire rationality. Intuitively, a past history of rational agency is no infallible indicator of present rationality, since the agent may have lost the rationality that once enabled him to act rationally. Similarly, the fact that an agent will (either certainly or with a sufficiently high degree of probability) act rationally in the future is not an infallible indicator that that agent is rational at the present moment, since, intuitively, it may well be that the agent will only come to act rationally at a future time because he will become rational at some point after the present but before that future time. This is not to say, of course, that a past or future history of rationality can’t constitute (defeasible) evidence of rationality – it obviously can. If I know that Bob has acted rationally at some point in the past, that is *prima facie* (i.e., absent any reasons to think that Bob lost his rationality at some point between that point in the past and the present moment) a very good reason to think that Bob is rational right now. The point is just that such a history is not that *in virtue of which* an entity is rational. Bob is not rational because he has acted rationally in the past. Rather, Bob acted rationally in the past because he was (and may still be) rational.

A case like McMahan’s “Superchimp” can serve to illustrate this point.74 Suppose an ordinary adult chimpanzee is somehow enhanced – perhaps through gene therapy or a brain transplant – in a way that renders it about as intelligent as a 10-year-old human child. The chimp would then have had rational actions in its future even before the gene therapy was administered. But that obviously does not mean that the chimp was rational at that time. Similarly, we can imagine the enhancement being reversed so that the chimp goes back to being an ordinary chimp, *Flowers for Algernon*-style. The post-reversal chimp has a past history of rationality, but is obviously non-rational all the same.

The Superchimp case is a case in which “historical” actualism is overly inclusive, deeming a non-rational agent rational. It can therefore be taken to show the following: Though historical actualism was supposed to liberalize the actualist account, perhaps it went too far –

74 McMahan 2001, p. 163
perhaps we should seek a golden mean, a theory that is more liberal than the original actualist proposal, but more restrictive than the historical actualism outlined above.

Even such an account, however, would not be the end of the actualist’s troubles. Again, the Superchimp case was a case where historical actualism was too inclusive. It therefore arose from the fact that past or future rationality is not sufficient for rationality. However, past or future rationality is not necessary for rationality either. Just as the presence of a past history of rationality is no infallible indicator of the presence of rationality, so the absence of such a history is no infallible indicator of the absence of rationality. Instead, it may simply mean that the agent in question has only just become rational, and hasn’t yet had the opportunity to act rationally. Similarly, just as a rational future is no infallible indicator of a rational present, so a non-rational future is no infallible indicator of a non-rational present. Instead, it may simply mean that the agent is currently rational, but will lose his rationality before he gets the chance to exercise it again. Therefore, there are also cases in which the original historical account – and therefore, a fortiori, a more restrictive version of it – is too restrictive, deeming rational entities non-rational.

In fact, there seem to be cases where rationality is present even though it is not, has never been, and never will be exercised. Suppose I build a Calvin and Hobbes-style Duplicator with which to make a perfect, qualitatively identical clone of myself. The procedure is almost completely successful, and in particular, my clone enters the world with all my memories and all my innate and acquired rational capacities, such as the ability to speak English or do simple mental mathematics. However, the procedure also goes wrong in one big way: The Duplicator does a bad job of copying my kidneys, so the clone comes into existence in a deep coma due to kidney failure – a coma from which it never wakes up.

In this case, my clone has never actually acted rationally, and it never will. Thus, it clearly fails to qualify as rational on any actualist account, historical or not. But it seems clear that the clone is rational – just as rational as me, in fact.

3.13 Dispositional rationality

In all its appealing simplicity, actualism turns out to be too simple. More specifically, it goes wrong by failing to distinguish between occurrent and dispositional rationality. This is another special case of the broader occurrent/dispositional distinction which we earlier saw Boonin apply to desire. But as was also noted when we last looked at the distinction, it can be applied to many types of mental states, and among these are plausibly the mental states that are constitutive of rationality. For much like ascriptions of belief or desire, ascriptions of rationality
can, depending on the circumstances, seemingly ascribe one of two distinct but related properties. There’s the sort of rationality exhibited by agents when (or to the degree that) they are actually acting rationally, but there’s also a deeper, more abiding rationality which consists in being disposed towards or having a capacity for rational agency in some appropriately strong sense. At this point, it should be mentioned that I take the rational capacities I have mentioned in previous chapters and am now mentioning again to be equivalent to the rational dispositions I will now discuss. I also take it, however, that this does not necessarily mean that there is, in general, no interesting distinction between what capacities are and what dispositions are, let alone that there is no interesting distinction between how attributions of capacities and attributions of dispositions work in our language.

One way to get a better grip on the distinction between occurrent and dispositional rationality is to consider the distinction between the utterance “He’s not rational” and the utterance “He’s not being rational.” The former seems to ascribe a lack of dispositional rationality, while the latter seems to ascribe a lack of occurrent rationality. And, it seems quite possible for the first to be false and the second to be true of the same person at the same time. Rationality, in the dispositional sense, is not “use-it-or-lose-it.” And while some form of actualism is probably the right account of occurrent rationality, no form of actualism can adequately account for dispositional rationality – for the sort of rationality that can be attributed even to the temporarily unconscious.

Further, dispositional rationality, rather than occurrent rationality, is what we should be interested in here. This is because, of the two, dispositional rationality is a far, far better candidate for being the basis of our right to life. Occurrent rationality, as we have seen, is too fleeting – just as the temporarily comatose person remains rational in one sense of the word, he also clearly retains whatever moral status he had before the coma.

So what is dispositional rationality? Intuitively, and as already noted, it is a disposition towards rational agency. But what sort of disposition? To answer this question, we must first make some general observations about dispositions.

In the most general terms, a disposition is an object’s tendency to behave in a certain way if exposed to a certain stimulus. Stock examples include fragility (i.e., the disposition to shatter if exposed to a sufficient amount of force) and water solubility (i.e., the disposition to dissolve if put into water).

75 I owe this example to Bjørn Ramberg.
The above suggests (and most of the literature seems to agree) that dispositions haveour core features. These are their bearers, realization conditions, manifestations, and bases.
Roughly and intuitively, these can be characterized in the following way. The bearer is the thing
that has the disposition, so that, e.g., the bearer of a certain instance of fragility may be a certain
Ming vase. The realization conditions are the events or states of affairs that will or would lead
to the disposition’s manifestation, such as the vase being hit with a hammer or being dropped
on to a concrete floor from a great height. The manifestation is the way in which the bearer will
or would behave upon being exposed to the realization conditions, such as the vase shattering.
Finally, the base is those properties of the bearer in virtue of which it has the disposition in
question, such as the microstructural properties of the vase.\footnote{Note that I don’t necessarily mean to claim that all dispositions have all of these features. In particular, it has been argued (see e.g. Mumford 2006) that there are baseless (or “bare”) dispositions. This is of little practical interest here, however, since the dispositions that are of interest to us here very arguably do have bases, as we will see shortly.}

A basic presumption of this thesis will be what I call the \textit{causal picture of dispositions}.\footnote{See Prior, Pargetter & Jackson 1982 for one of the foundational texts of this approach to dispositions} On the causal picture, the realization conditions and base of a certain disposition jointly make
up (or would have made up) the sufficient cause for a certain event, that event being the
manifestation, and it is in virtue of standing in these relations to each other that these things
jointly make up a disposition.

An immediate worry is that the causal picture makes the delimitation of realization
conditions and bases arbitrary. For suppose we have a disposition with realization conditions
A, base B, and manifestation C. On what basis have we decided to classify A as the realization
conditions and B as the base? What's stopping us from classifying B as the realization
conditions and A as the base? Either way, the core facts stay the same: A and B jointly make
up a sufficient cause for C. We need some principled way of drawing the line.

We can find such a way by reintroducing the one feature of dispositions that has so far
been absent from the causal picture – the bearer. Intuitively, the base is the bearer's contribution
to bringing about the manifestation, while the realization conditions are the external world's
(external, that is, to the bearer) contribution to bringing it about. This suggests the following
way of drawing the line: The base are those parts of the sufficient cause that are (or essentially
involve) intrinsic properties of the bearer, while the realization conditions are all the other parts
of the sufficient cause.

Drawing the distinction in this way raises a further question, namely how we should
draw the intuitive but hard-to-articulate distinction between intrinsic and extrinsic properties.
For our purposes here, the following informal characterization will do: *A property is intrinsic just in case it would be transferred from original to copy by a Star Trek-style matter replicator.* Thus, unproblematic examples of extrinsic (that is, non-intrinsic) properties include historical and locational properties: A copy of our Ming vase will not be in the same place as the original vase, nor will it be the case that the copy once sat on Lord Acton’s mantelpiece simply because the original once sat there. On the other hand, clear examples of intrinsic properties include things like mass and shape.

Given the causal picture, many questions about dispositions turn into questions about causality. This is not necessarily to say that a complete account of causality will bring with it a complete account of dispositions (or vice versa), but at the very least, it will certainly suggest certain accounts of dispositions to us, as well as putting restrictions on which accounts of dispositions we cannot accept. The fact that David Lewis (among many others) felt so attracted to a counterfactual analysis of dispositions surely has a great deal to do with the fact that Lewis (among many others) felt equally attracted to a counterfactual analysis of causality. Similarly, it is no accident that so many Aristotelians and neo-Aristotelians want to explain both phenomena in terms of powers.\(^\text{78}\)

It is also worth noting that the causal picture, just in and of itself, does not seem to commit us to any particular account of causality. As we have seen, it conditions views of dispositions both among those, such as David Lewis and David Hume, who take “deflationary” views of causality and among neo-Aristotelians like Stephen Mumford, who take “inflationary” views of it.\(^\text{79}\) In line with my ambitions towards metaphysical non-presumptuousness, I do not want to even begin to offer a comprehensive account of causality or dispositions here. Rather, what I say from here on out will depend on two heuristics. The first is the same sort of case-based, intuition-driven approach that underlies the rest of this work. That is, I will leverage our pre-theoretical judgments about dispositions and causality while suspending judgment about what sort of comprehensive account of causality (if any) these judgments commit us to. Secondly, in line with the general consensus on the matter, I will assume that dispositions can be analyzed – or at least usefully thought of – in conditional terms.

---

\(^\text{78}\) See Lewis 1997 for Lewis on dispositions. For examples of recent powers-based accounts of causation and dispositions, see Mumford & Anjum 2015 and Molnar & Mumford 2003.

\(^\text{79}\) Deflationary views of causality I take to be those views of causality which say that there is nothing over and above cause and effect – that A being the cause of B consists in nothing but the fact that (very roughly speaking) when A happens, B happens too. Inflationary views of causality I take to be those views of causality which say that there is something over and above cause and effect – e.g., that A and B’s “constant conjunction” is no accident, but rather the result of A and B’s causal powers.
The last point is worth dwelling on. Informally, a conditional analysis of dispositions is one which characterizes the presence of any given disposition as boiling down to the truth of a certain “if-then” statement. More specifically, to say that a certain thing is disposed to give a certain response to a certain stimulus at a certain time just is to say that if that thing is exposed to that stimulus at that time, then it will give that response. For example, to say that a certain vase is fragile just is to say that if that vase is exposed to a sufficient amount of force – by being dropped from a great height, struck with a hammer, or what have you – then it will shatter.

Now, dispositions are standardly divided into “canonical” and “conventional” dispositions – the former being dispositions whose realization conditions are explicitly specified (‘The vase’s disposition to shatter if hit with a hammer’) and the latter being dispositions whose realization conditions are not explicitly specified (‘The vase’s disposition to shatter’). Since the general idea behind conditional analyses is that X’s having a disposition with manifestation M and realization-conditions R boils down to the fact that X will manifest M if R comes about, the conditional analysis tacitly assumes (as I have also done so far, and will go on doing) that conventional dispositions are really canonical dispositions in disguise – that all dispositions have realization conditions, though we sometimes don’t feel the need to spell out what they are.

There’s a question as to what sorts of conditionals dispositions should be analyzed in terms of. One natural proposal – and one that has been tried\(^{80}\) – is to analyze them in terms of the truth-functional material conditional (in symbols, ‘p → q’). However, this analysis has notorious weaknesses. The truth-conditions of the material conditional are very simple: ‘p → q’ is false when p is true and q is false, and true in all other cases. Thus, on this analysis, saying, e.g., that a Ming vase V is fragile at a time t just is to deny that V is exposed to a given amount of force at t without shattering. So, V will be fragile at any time at which it is not exposed to that amount of force, and in general, any given object will have any given disposition at any time at which that object is not exposed to the realization conditions of that disposition. The Oort cloud will be (and will always have been) disposed to shatter if struck with a hammer, since no-one has ever struck the Oort cloud with a hammer; the number five will be (and will always have been) water-soluble, since no-one has ever put the number five into water; and so on. But this is clearly absurd.

Because of these sorts of problems, conditional analyses of dispositions now almost universally endeavor to analyze dispositions in terms of *counterfactual* conditionals – i.e., in

\(^{80}\) Notably in Carnap 1936 & 1937
terms of statements of the form ‘If p had been the case, then q would have been the case’ (in symbols, ‘p □→ q’). Intuitively, counterfactual conditionals differ from material conditionals – which, as we saw, are such that the falsehood of the antecedent guarantees the (trivial) truth of the whole conditional – in that they let us reason non-fatuously about things that didn’t actually happen. This intuition is borne out by what is by far the most popular semantics for counterfactuals conditionals, namely the so-called Lewis-Stalnaker semantics (LSS), the gist of which is this: ‘p □→ q’ is true at a possible world w₁ just in case q is true at every possible world w₂ such that (a) p is true at w₂ and (b) w₂ is no more similar to (no “closer to”) w₁ than any other possible world at which p is true. In snappier form: p □→ q iff all the closest p-worlds are also q-worlds.⁸¹

The simplest counterfactual analysis of dispositions, which has appropriately been termed the Simple Counterfactual Analysis (SCA), goes like this: An object O will be disposed to give a response R in response to a stimulus S at a time T just in case, if O had been exposed to S at T, then O would have given R. Thus, combining SCA with LSS, we get the following: The fact that the Ming vase is disposed to shatter if hit with a hammer amounts, roughly, to the fact that at the closest possible world where the Ming vase is actually hit with a hammer (which may or may not be the actual world), the Ming vase shatters.

However, the SCA is not without problems of its own. More specifically, it faces at least three sorts of counterexamples. The best-known of these⁸² involve finks – i.e., dispositions whose realization conditions are also the conditions for the bearer losing the disposition in question. Along somewhat similar lines, blockers are external objects that prevent dispositions from being manifested. Imagine, e.g., that our Ming vase has been wrapped in bubble wrap. Then it will not shatter if struck with a hammer, though intuitively, it remains fragile all the same. Conversely, mimics are external agents which, given a certain stimulus, will intervene to produce a certain effect in a certain object, though the object itself is not disposed to give that effect in response to that stimulus. An ordinary styrofoam plate, unlike, say, a porcelain plate, is intuitively not disposed to break if struck with a hammer. But suppose that the plate is within earshot of the infamous Hater of Styrofoam, whose mission in life is to destroy things that are made out of styrofoam. If I strike the plate with a hammer, the plate will make a distinctive sound that the Hater of Styrofoam will immediately recognize as the sound of a styrofoam plate

⁸¹ Lewis 1973 and Stalnaker 1968 are the foundational texts of this approach to counterfactuals. Lewis’ and Stalnaker’s analyses differ on some points, notably the way in which the closeness relation between possible worlds works, but these differences need not concern us here.

⁸² First mentioned in Martin 1994, but popularized by Lewis 1997
being struck. Enraged, he will storm over to the plate and tear it in half. Hence, it is, in fact, the case that the plate would break in half if struck with a hammer. But we surely don’t want to call it fragile on that account.83

I think these problems can be at least partly sidestepped through what has been dubbed “the strategy of getting specific.”84 On this strategy, to say, e.g., that a Ming vase has a disposition to shatter if dropped from a great height is really under-specific loose talk. What the vase has is not really a disposition to shatter if dropped, full stop, but a disposition to shatter when it is dropped from a minimum of such-and-such a number of meters and such-and-such a force of gravity and such-and-such a hardness of the landing surface and… In short, in addition to those realization conditions that have been explicitly specified (i.e., the vase being dropped), many background conditions must also be met.

(Note, by the way, that I suspect that the distinction between background and foreground conditions is a completely subjective and pragmatic matter: The background conditions for the realization of a given disposition are just those parts of its realization conditions we don’t feel the need to explicitly mention, e.g. because we don’t anticipate that they will change any time soon or because we find them too obvious to be worth mentioning. Note also the close connection between this view and the view that conventional dispositions are canonical dispositions in disguise: Sometimes, when we describe dispositions “conventionally” instead of “canonically,” we feel no need to explicitly mention any of their realization conditions, and sometimes (not to say always) when we describe them “canonically,” we feel no need to explicitly mention all of their realization conditions.)

Without these sorts of background conditions, strange consequences follow. Suppose I take our Ming vase on a spaceship to a planet where the force of gravity is much weaker than on Earth. If I drop the vase from a great height on that planet, it will not shatter, but will float gently to the ground and land unharmed. Should this lead us to the strange conclusion that I deprived the vase of one of its dispositions simply by changing its location? If all the vase has is a disposition to shatter if dropped, it should, since it is clear that the vase would shatter if I dropped it on Earth, but would not shatter if I dropped it on the other planet. But if the realization conditions also implicitly include various background conditions, including an Earthlike force of gravity, it should not. On this view, the vase has the same disposition to shatter if dropped (loosely speaking) on both planets – it’s just that, thanks to a change in background conditions, the realization conditions, and therefore the manifestation, simply aren’t realized when I drop

83 Lewis 1997, p. 11
84 Manley & Wasserman 2008
it on the other planet. It was true on Earth, and remains true on the other planet, that the vase would shatter if it were dropped on to such and such a surface from such and such a height with such and such a force of gravity, and... (This also highlights the seemingly subjective and pragmatic nature of the foreground/background distinction. The reason why we hardly ever feel the need to explicitly include Earthlike gravity in the realization conditions of the vase’s disposition to shatter if dropped is presumably because we don’t anticipate finding ourselves in a place without Earthlike gravity any time soon. But if we did live in a world where such a change in gravity was a possibility, say because interplanetary travel had become common, I suspect we would promote Earthlike gravity to the foreground.)

How does all this help us ameliorate the SCA’s troubles? By opening up the possibility that the background conditions of any given disposition will include no-finks clauses, no-maskers clauses, no-mimics clauses, and so on. This may seem ad hoc, but the SCA is elegant and appealing enough that a little adhocness is an acceptable price to pay for saving it. Further, the move is made less unpalatable by the fact that, as noted, the claim that dispositions have background conditions should be independently plausibly to anyone who accepts the view that conventional dispositions are canonical dispositions in disguise.

It should also be noted that dispositions seem to come in strengths. Sometimes this is explicitly mentioned, as in a statement like ‘I am strongly disposed to believe that it was Peter who stole the silverware’. But often – probably much more often – the notion of disposition-strengths is more implicit, as in a statement like ‘Be careful with that vase – it shatters very easily’. What is being asserted here, I take it, is that the vase has a very strong disposition to shatter. In this, the vase differs from other objects (such as other, sturdier vases) that might still be disposed to shatter, but which do not shatter quite as easily. Intuitively, the strength of a disposition has to do with how easily it could be realized – how close at hand its actualization is.

How do we fit this notion into what has been said so far? It may seem tempting to define it in terms of probability: The more probable a disposition's realization conditions are, the stronger that disposition is. But it is plausible that many or all probabilities are subjective, and even more plausible that they are least relative to a body of evidence. Thus, if disposition strengths are probabilities, they inherit these properties. But surely there woud be dispositions and disposition strengths even if there were no subjects. And although we can, of course, revise our beliefs about disposition strengths in light of new evidence, surely this does not change the strengths themselves, though it does change our probabilities. If I learn that the vase I am holding in my hand, which I had previously taken to be a very fragile Ming vase, is made of
sturdier stuff than I thought, I will go from believing that the vase is strongly disposed to shatter
to believing that it is only weakly disposed to shatter. But what has happened here is clearly not
that the vase itself has thereby become more weakly disposed to shatter – it did not become
sturdy only when I added this new information to my evidence. Rather, it was always sturdy –
I just didn’t know it.

Instead, I propose to analyze strengths in terms a notion that’s already embedded in
LSS, namely the notion of closeness between possible worlds. On this analysis, the strength of
a disposition $D$ at a time $t$ and possible world $w$ is the distance between $w$ and the world (or
worlds) nearest to $w$ at which $D$ manifests itself at $t$. The farther from actuality a disposition’s
manifestation is, the weaker that disposition is.

With these observations about dispositions in mind, we can say a bit more about
dispositional rationality. Given the two-part characterization of rationality I offered previously
– the one on which rationality involves seeing and being moved by reasons – it is natural to say
that dispositional rationality just is the disposition to be moved by the reasons one sees. In other
words, a rational disposition is a disposition whose bearer is a certain agent $A$, whose
realization-conditions are $A$’s knowing that a certain entity $R$ obtains or exist when $R$ is, under
the circumstances, a reason for $A$ to do a certain action $X$, and whose manifestation is $A$’s
updating his intention to do $X$ in the way $R$ demands, because $R$ demands it. Given SCA, this
amounts to the claim that one is rational when (or to the extent that) for each of the reasons one
could have, one would be moved by it if one saw it.

Thus, rationality is not one disposition, but several – for each possible circumstance
under which $R$ is a reason for an agent $A$, there is $A$’s disposition to be moved by $R$ if he sees
$R$ under those circumstances. This gives us a neat way of accounting for the fact that rationality
seems to come in degrees, that we can speak of one agent being more or less rational than
another. For if our rational dispositions are several, we can roughly equate an agent’s degree of
rationality with his number of rational dispositions. I say “roughly” because it seems that not
all rational dispositions count for the same. More specifically, one’s sensitivity to reasons that
are close at hand seems to count for more than one’s sensitivity to more remote reasons:
Intuitively, the fact that I was unmoved by a reason I actually saw bespeaks a greater lack of
rationality than the fact that I would have been unmoved by a reason I could have seen if the
world were very different from how it actually is. This suggests that when rational dispositions
are counted up to measure degree of rationality, they should be weighted according to strength:
The stronger a rational disposition of mine is – the closer its realization is to actuality – the
more rationality it imbues me with.
But the distinction between rationality and non-rationality is a distinction of kind as well as of degree. We can speak of one entity being more or less rational than another, but we can also, it seems, attribute rationality or non-rationality, full stop. In this way, rationality works in somewhat the same way as does a property like height. We can speak of Bob being more or less tall than Alice, but we can also speak of Bob simply being tall or short, full stop. Now, in the case of tallness, the relative notion seems to be prior to the absolute notion. To say that someone is tall, absolutely speaking, is to say that he exceeds (or exceeds by a certain amount) a certain “ideal” height. (Note that this analogy is not perfect. The “ideal height,” and therefore the truth-values of ascriptions of tallness, seems to be context-relative. For example, it is natural to say that 170 cm is tall for Indonesia, where the mean adult height is about 150 cm, but not tall for the Netherlands, where 170 cm is a bit below the mean adult height. But rationality is presumably not relative in this way.)

It seems reasonable to think that the relation between rationality as a relative measure and rationality as an absolute measure works in a similar way. To be rational, absolutely speaking, is simply to exceed some minimum degree of rationality, that is, to be sensitive to some minimum weighted number of reasons. It is not easy, of course, to say exactly what the minimum is (nor exactly how the weighting should work), just as it is not easy to pinpoint the exact height one has to exceed in order to count as tall (or tall for a man, tall for an Indonesian, or whatever). This vagueness, however, is plausibly a feature rather than a bug. After all, it is hard to identify exactly how mentally sophisticated one has to be in order to count as a rational animal rather than an advanced but non-rational animal. This is demonstrated, for example, by the fact that we often struggle to decide whether rationality is present in “edge cases” such as certain species of great apes or certain mentally retarded human beings. The fact that my analysis of rationality, just by itself, offers little help in adjudicating these edge cases may seem like a deficit in explanatory power. But while explanatory power is a good thing, it is, as with most good things, possible to have too much of it. A theory can explain too much, too easily, and an analysis of rationality that offered us a quick and easy verdict on this matter would arguably have fallen prey to precisely this problem, making it suspiciously easy to answer a question we know to be difficult.

With this, we have answered several questions about our dispositional rationality. Firstly, it is, unsurprisingly, a set of dispositions – a set of dispositions whose bearers are us, whose realization conditions is our seeing a given reason, and whose manifestation is our being rational in a given way. But while explanatory power is a good thing, it is, as with most good things, possible to have too much of it. A theory can explain too much, too easily, and an analysis of rationality that offered us a quick and easy verdict on this matter would arguably have fallen prey to precisely this problem, making it suspiciously easy to answer a question we know to be difficult.
moved by it. In a slogan, rationality is reasons-sensitivity. This, however, still leaves one big question unanswered, namely what the base of our rationality is. What is it about us that causes us to be moved by the reasons we see? Once we have answered this question, we will finally be in a position to learn what it is to have a rational genome.

3.2 The base and basis of human rationality

At the start of the previous section, I roughly characterized a dispositional base as that in virtue of which the bearer of a given disposition has that disposition. Later, I characterized it, less roughly, as that part of the sufficient cause of the disposition’s manifestation that belongs intrinsically to the bearer. The question of the base of our rational dispositions highlights one of the rough elements in that first, rough characterization – one way in which it can come apart from its second, more precise formulation. One the one hand, it is standard – and intuitively appealing – to say that before a certain level of brain development, human organisms lack rational dispositions, and therefore also rationality itself, given this chapter’s account of it. This intuition is vindicated by the causal picture of dispositions, for the interactions between our brains and central nervous systems on the one hand and the external world on the other hand are surely precisely what constitutes our seeing of reasons, and thus causes us to be moved by them. It seems, therefore, that the brain and central nervous system are the base of human rationality. But on the other hand, our brains and central nervous systems come from somewhere. More specifically, they form and mature thanks to a complex interaction between our genes and the external environment. Thus, the ultimate cause of our rationality (or rather, the ultimate cause of our rationality within ourselves) seems to be our genomes. They are ultimately the things in virtue of which we have our rational dispositions.

3.2.1 The proposal

How can these facts be squared with one another? If the genome isn’t the base of our rational dispositions, what is it? I say that it’s the basis. By that, I mean this: The brain and central nervous system make up the base for our first-order rational dispositions – that is, for our dispositional rationality – while the genome makes up the base for our second-order rational dispositions.

What do I mean by “first-order” and “second-order”? Generally, a disposition with a bearer B is higher-order (that is, not first-order) just in case its manifestation consists in B

---

86 Note that these definitions of ‘base’ and ‘basis’ are meant to be stipulative. I don’t claim that the ordinary-language meanings of ‘base’ and ‘basis’ differ in a way that is at all analogous to the way outlined here.
gaining a certain disposition. A bit more precisely, we can define dispositional orderedness inductively: A disposition with a bearer B is first-order just in case its manifestation does not consist in B gaining a disposition, and nth-order (for all n other than 1) just in case its manifestation consists in B gaining n-1th order disposition. B's second-order disposition to do X will then be B's disposition to give itself a first-order disposition to do X (i.e., to give itself the disposition to do X, pure and simple), B's third-order disposition to do X will be B's disposition to give itself a second-order disposition to do X, and so on.

The idea, then, is this. My genome, plus the right input from the environment, jointly make up a sufficient cause for the formation of my brain and central nervous system. Once my brain and central nervous system are in place, they, combined with the right input from the environment, make up a sufficient cause for my reasoning activity. My brain and central nervous system therefore constitute the base for my (first-order) rational dispositions, and hence for my rationality. But the genome makes up the base for my organism’s disposition to develop the brain and central nervous system87, hence gaining my first-order rational dispositions – it constitutes the base for my second-order rational dispositions, for the fact that I was disposed to become rational in the first place. This is what I mean when I say that we have the genetic basis for rationality.

3.22 Objections

It might be objected that this gives a simplistic picture of how rationality and reasoning are generated. After all, it's not really as if my brain plus environmental (and hence, I take it, external to me) input jointly make up a sufficient cause of my reasoning, nor is it the case that my genome plus environmental input jointly make up a sufficient cause for the formation of my brain. A bevy of other intrinsic factors – factors that, in general, involve my organism satisfying certain minimal requirements for normal, healthy functioning – are also necessary for rationality and reasoning to come about. To see this, consider a human corpse. If an undiscovered congenital heart defect were to suddenly kill me now, my corpse would still be in possession of the same brain I had when I was alive. But from this, it obviously does not follow that my corpse would interact rationally with its environment in the same way I would. In short, therefore, our brains and genomes cannot be the sole bases of our first- and second-order rational dispositions, since they are not the only parts of the sufficient causes of the manifestations of those dispositions that are or involve intrinsic properties of ours. To be sure,

87 For expository simplicity, I will henceforth refer to the brain and central nervous system simply as “the brain.”
the brain and genome will be causally necessary for these manifestations, but so are many other intrinsic factors, including, e.g., the fact that I have a beating heart.

One thing this shows us, I think, is that while there is a sense in which we are the bearers of our first- and second-order rational dispositions, there is another, equally good sense in which our brains and genomes are their bearers, while their bases are those specific parts of the brain or genome that account for our rationality. (This sort of ambiguity is quite common. For example, my disposition to reflexively jerk back my knee if my knee is struck with a reflex hammer can, it seems, equally well be described as my knee’s disposition to jerk back if struck with a reflex hammer.) From this point of view, the brain’s rational dispositions amount, roughly speaking, to the fact that if it is integrated into the body of any living human organism that satisfies some minimal requirements for healthy, normal functioning and exposed to the right stimuli (i.e., to stimuli that amount to the seeing of reasons), then it will respond, or cause the organism to respond, to those stimuli in a particular, rational way (i.e., in a way that amounts to being moved by those reasons). The rational dispositions of the genome will amount, roughly speaking, to the fact that if it is integrated into the body of any living human organism that satisfies some minimal requirements for healthy, normal functioning and exposed to certain other stimuli, then it will cause that body to develop a brain that is rationally disposed in the way just described. And our first- and second-order rational dispositions will simply amount to the fact that we have brains and genomes that are disposed in the way just described.

The requirement for an otherwise minimally functional organism can then be thought of as playing a role at the intrinsic, basal level that is equivalent to the role played by background conditions at the extrinsic level of realization conditions. And so, just as it is legitimate to say that a realization-condition of a Ming vase’s disposition to shatter is the vase’s being dropped even though, in reality, dropping the vase will only cause it to shatter if certain background conditions are also met, so it is legitimate to say that our brains and genomes are the bases of our rational dispositions even though the bearer must also meet certain other “background conditions” in order for those dispositions to be present.

Another objection is that the above account presupposes a certain, contentious view of the relation between me and my organism, as it entails that certain features of my organism – viz., its brain and genome – are intrinsic features of me. This is a consequence of three claims that were made earlier:

- I am the bearer of both my first- and my second-order rational dispositions
The bases of my first- and second-order rational dispositions are my brain and my genome, respectively.

The base of any disposition is an intrinsic feature of the bearer.

I take this conclusion to be equivalent to the conclusion that my brain and genome are proper parts of mine. Call this the Parthood Claim.

The Parthood Claim might be thought contentious because it might be thought to have substantive implications for personal identity. In particular, it might be thought to suggest that we should be animalists of some sort. But this impression is, I think, mistaken. Firstly, the Parthood Claim does not entail animalism. Even if my brain and my genome are proper parts of both me and my organism, the fact that X is a proper part of both Y and Z clearly does not entail that Y = Z. My left thumb is a proper part of both my left hand and my organism, but it does not follow from this that my left hand is identical to my organism.

What, if anything, does the Parthood Claim entail about personal identity? In general, a theory of personal identity can posit four sorts of relations between me and my organism: (A) the two may be identical, (B) my organism may be a proper part of me, (C) I may be a proper part of my organism, or (D) none of the above may be true – that is, my organism may be entirely distinct from me, with no mereological overlap whatsoever. (A) is the stance taken by animalists, and is obviously compatible with the Parthood Claim, though not entailed by it. (B) is the stance taken, e.g., by those forms of substance dualism which regard me as the mereological fusion of my soul and my body. Like (A), (B) is not only compatible with the Parthood Claim, but actually entails it, given the truism that my brain and genome are proper parts of my organism and the truism that proper parthood is transitive. (C) is the stance taken by McMahan, and as he points out, the most natural way to cash it out is to say that I am constituted by my organism, much as a bronze statue is constituted by a certain lump of bronze. But as McMahan also points out, when one thing constitutes another, those two things plausibly share all non-historical properties.\(^{88}\) (For example, while a lump of copper constitutes a certain copper statue, the lump will have the exact same shape and location as the statue. On the other hand, from the fact that the lump came into existence billions of years ago, it clearly does not follow that the statue came into existence billions of years ago.) So, if I am constituted by my organism, every non-historical property of my organism is also a property of mine. But having a certain genome as a proper part and having a certain brain as a proper part are non-historical.

---

\(^{88}\) McMahan 2001, p. 89
properties, so if I am constituted by my organism, it follows that I inherit those properties from my organism. (C), too, seems to not only be compatible with the Parthood Claim, but to actually commit us to it. That leaves only (D), which does seem to be incompatible with the Parthood Claim. But what sort of theory could underwrite (D)? Only, it seems, a form of substance dualism on which I am identical to my soul, rather than being the mereological union of my soul and my body. For what could I be, if I am entirely distinct from and non-overlapping with (and therefore presumably not realized or constituted by) my body? The only halfway plausible answer seems to be that I am a disembodied ego. Hence, the only clear implication the Parthood Claim has for personal identity is that it commits us to the falsehood of this extreme sort of substance dualism. But that, I take it, is not a very controversial point of view to take within the present debate, where neither side is too keen on substance dualism of any kind. Rather, as we have seen, the pro-choicers tend to embrace the embodied mind account or physicalistically tinged psychological accounts, while the pro-lifers tend to be animalists.

Finally, it might be objected that my account gives too much credit to “internal” factors like the brain and genome, and not enough to the input from the external environment. But it is hard to see how we could coherently give any more credit to the environment, given that the base of a disposition is characterized precisely by being internal and intrinsic to the bearer of that disposition. Environmental input does play a very important role in the formation and exercise of our rational capacities, but by definition, it plays that role at the level of realization conditions, not at the basal level. Further, there are several reasons why we should not want to say that environmental input alone constitutes a sufficient cause of our rationality, without any “internal” input at all.

Firstly, basing rationality in the genome explains why it develops convergently among almost all human beings. It also explains why, as far as we know, rationality develops only among members of the human species, who are distinguished from other organisms precisely by the fact that they have a human genome. Liao brings up these points while defending his claim that moral agency has a genetic basis:

---

89 Admittedly, it’s open to dispute whether humans really are the only rational animals around. There are those, such as Singer and Cavalieri (1993), who argue for extending basic human rights to certain non-human primates, and such a position goes well with the view that these primates are rational animals. But this does not undermine my point – quite the contrary. If we redrew the border between rational and non-rational animals so as to include great apes, we would still be drawing the border in a way that overlapped with the borders between species, and the organisms we would be including within it would be precisely those extant organisms whose genomes are the most like ours.
The capacity for moral agency is grounded in psychological capacities such as rationality and empathy that uncontroversially have a genetic basis. Indeed, rationality and empathy, two essential components of the capacity for moral agency, develop in all normal human beings according to a fairly predictable schedule. If the capacity for moral agency did not have a genetic basis, the development of its essential components would not be so regular.\(^90\)

In short, if our rationality were down to the (cultural or natural) environment, we should not expect it to be a cross-cultural human universal, for the same reason we should not expect an uncontacted tribe in the Amazon to speak French.

Secondly, crediting too much of our rationality to the environment threatens to create a vicious infinite regress, a sort of chicken-and-egg problem. For if we are made rational by our environment, it is surely, in large part, our cultural and linguistic environment. This is certainly the usual idea among those, such as the late Wittgenstein, who take what might be called an externalist view of human rationality. Thus, on this view, the rationality of any given human being will be credited to the fact that he was raised by other human beings who already had a culture and a language, and who had a culture and a language advanced enough to imbue him with rationality. But such a community must surely be a community of rational beings. Therefore, there cannot have been a first rational human being, since \textit{ex hypothesi}, we can become rational only by being raised in a community that already contains rational beings. Granted, we can break the regress by claiming that the first rational human being was raised in a community of rational non-human beings. Setting aside hypotheses about ancient aliens or the fire of rationality being handed down by the gods, the idea must presumably be that these beings were our immediate evolutionary precursors. But then we can ask how the rationality of these non-human beings was based. If it was environmentally based too, the regress goes on.

And if it was not – if it was internally based – why on Earth should we think that our rationality isn’t internally based too? Given that these beings were our immediate evolutionary precursors, shouldn’t they have passed down any internal – and, in particular, any genetic – basis for rationality to us?

3.23 Who has the basis?

Do all human genomes contain the genetic basis for rationality? Certain parts of the human genome, such as the part that determines that we have two eyes, are common to all or almost

---

\(^{90}\) Liao 2010, p. 164. In a footnote to this passage, Liao goes on to note that “Piaget’s work in human cognitive development, for example, supports the idea that rationality develops according to a fairly predictable schedule.”
all of us. Indeed, some parts of it are plausibly essential — that is, necessarily common to all human genomes. Other parts, such as the parts that determine eye color, vary from genome to genome. In which of these categories does the genetic basis for human rationality belong?

Since the genetic basis for human rationality is not yet well understood (but presumably very complex), there is no direct way of finding out whether a given human being has it. Empirically, however, almost all human beings become rational as they mature, and many more fail to become rational only because the development of their rationality was stopped short by death or severe injury. Thus, the basis of rationality is, at the very least, an extremely common part of the human genome.

In fact, the only plausible candidates for human organisms who lack the genetic basis for rationality seem to be those that suffer from certain severe genetic defects. For example, infants with anencephaly are born without a cerebrum, and therefore lack the base of human rationality. Authors on bioethics often presume that at least some causes of anencephaly are genetic, McMahan (who at one point offhandedly mentions “those genetically destined to develop into anencephalic infants”91) being an example. And if those authors are right, it follows that in at least some cases of anencephaly, the basis of human rationality is also missing.

As Liao points out, however, the actual existence of any defects that might fit this bill is currently uncertain at best and implausible at worst. Most of the birth defects that might at first seem to qualify for the job, such as anencephaly, are now thought to have causes that are mostly or entirely environmental; and even where genetic causes are possible, there is no reason to think that the genetic basis for rationality is missing entirely, rather than simply having been prevented from being expressed by some other part of the genome.92 However, even if there are no actual cases in which the genetic basis for rationality is missing, it is still worth thinking about what we should think if confronted with such a case.

It is worth noting, first, that any (purportedly) human organism that lacked the genetic basis for rationality would likely be much more alien than an anencephalic infant. As mentioned, the genetic basis for human rationality is almost certainly very, very complex, and it is plausible that many parts of it also provide the basis for other traits. Even if we do not find it plausible that the basis for rationality is itself an essential attribute of the human genome, we might well find it plausible that some of these other attributes are, so that any organism whose genome fully lacked the basis for rationality would not count as human. Some of these other

91 McMahan 2001, p. 336
92 Liao 2010, pp. 166 & 167
attributes may well also be necessary for the very life of the organism, so that no living human organism could lack the genetic basis for rationality. As Liao puts it:

[P]resumably, the genes for moral agency are not just some minor genes such as the genes for toe nails. Indeed, the genes for moral agency may be necessary for the development of a brain generally, and/or for the various capacities necessary for moral agency such as rationality and empathy. ... [A (purported) human being without the genetic basis for moral agency], if it could survive at all, might turn out to be one without a brain and/or one without the capacities for rationality and empathy. In such a case, would the resulting entity still be a human being? It is not obvious that it would be. And if such an entity were not a human being, then it would not be necessary to concede that there could be human beings who lack the genetic basis for moral agency.93

In short, it seems, as noted earlier, that some parts of the human genome must be essential, and the parts of it that provide the basis for rationality seem to be as good of a candidate as any. Indeed, this gives us one attractive way of construing the claim that we are essentially rational animals – one that does not reduce that fact to the truth of the generic “Humans are rational animals,” which does not do justice to the intuition that even seemingly non-rational humans remain rational animals in some important sense, nor commit us to the apparent absurdity that each and every human organism is rational.

Here’s a natural followup question: Given that all humans have the genetic basis for rationality, do they all have it to the same degree? Given my account of the base and basis of human rationality, this question can be rephrased and refined in the following way. It seems evident that humans differ in their degree of first-order rationality, plausibly to such an extent that some of them do not have it at all. Even among those of us who are rational, there are great differences in general intelligence (i.e., the trait that IQ tests are meant to measure), and it seems very plausible that in general, higher intelligence means a higher degree of rationality and lower intelligence means a lower degree of rationality. Similarly, some people have mental illnesses or personality disorders that negatively impact the capacity for rational agency in one way or another. Do our second-order rational dispositions differ in the same way? Do we all have the same (or at least the same weighted number of) second-order rational dispositions?

It might be thought that if there is a genetic basis for some of the things that account for differences in degree of first-order rationality, such as intelligence differences or mental illness, that will be clear proof that these differences are the result of differences in the genetic basis for rationality, and therefore of differing degrees of second-order rationality. And, it might be

93 Liao 2012, p. 267
added, there is no doubt that some of these things do have at least partly a genetic basis. For instance, meta-studies suggest that between 50% and 70% of the causes of IQ differences are genetic, while the same is true of between 60% and 80% of the causes of schizophrenia. But while I have no reason to doubt the conclusions of these meta-studies, I don’t think these conclusions show that we differ in degree of second-order rationality. For it could well be that the genetic causes of differences in degree of first-order rationality do not consist in differences in the genetic basis for rationality, but rather in some other genetically based trait coming in and regulating or suppressing the expression of the genetic basis for rationality (i.e., the realization of our second-order rational dispositions). And the latter possibility is quite compatible with the claim that we all have the same second-order rational dispositions – it’s just that we vary in the extent to which these dispositions are realized or realizable.

Further, even assuming, for the sake of argument, that some differences in degree of first-order rationality are caused by differences in degree of second-order rationality, it is quite clear that many of them are not. For while there seems to be a clear consensus that the causes of intelligence differences, mental illness, and so on are partly genetic, the consensus is at least as clear on the fact that they are also partly non-genetic, and due to differences in environmental input, gene expression, or what have you. Thus, it seems clear that our degrees of second-order rationality differ much less than our degrees of first-order rationality, and quite possible that they do not differ at all.

### 3.3 Looking ahead
At this point, it may seem tempting to despair of the prospects for defending human moral equality on the grounds of our status as rational animals. For as we have now learned, the basis of human rationality is probably universal (or at least very nearly so) among humans, but its base is not. Thus, it seems that many humans, including many fetuses, are not rational. They are not yet rationally disposed, though they are disposed to be rationally disposed. According to the second premise of the Central Argument, this second-order disposition is sufficient to endow them with the right to life. But why should we take this second premise to be true? Even if rationality itself is sufficient for the right to life, why think that the mere disposition to develop it should be similarly sufficient? That is the question I will now try to answer.

---

94 Bouchard & McGue 2003; Hunt 2011; McGuffin, Owen & Farmer 1995
4

The Genetic Criterion

In the previous chapter, we saw that having the genetic basis for rationality involves having genetically based second-order rational dispositions. We also saw why it is reasonable to suppose that all – or virtually all – human organisms have such dispositions. What is not yet clear, though, is why having such second-order rational dispositions should entail the right to life, as the second premise of our central argument asserts. Answering that question will be the focus of this chapter.

The second premise asserts a sufficient criterion for the right to life – call it the Genetic Criterion. After considering some possible reasons to reject the Genetic Criterion, I will lay out a reason to accept it.

4.1 Why not the Genetic Criterion?

4.1.1 Why not human moral equality?
Since the Genetic Criterion entails human moral equality, any objection to human moral equality in general is also an objection to the Genetic Criterion in particular. And there are plenty of such general objections, often proceeding from the idea that human moral equality has absurd consequences in concrete cases. In particular, it is often held that such absurd consequences follow from extending the right to life to the unborn at the earliest points in their development – that is, to zygotes or embryos.

One popular case that purports to show this is what Liazo calls the Embryo Rescue Case.⁹⁵ Suppose you are fleeing a burning hospital. As you do so, you come across an IVF lab containing a fertilized egg (a zygote) and a patient who cannot get out of the hospital on his own. You do not have the time or resources to get both the egg and the patient out before they are consumed by the flames – you must choose one or the other. The Genetic Criterion suggests that these two choices are on a moral par – that it is a matter of moral indifference whether to save the fertilized egg or the patient. And to make matters worse, if the choice had been between several eggs and one patient, it seems that you would have been obligated to save the eggs rather than the patient. But this seems absurd.

---

⁹⁵ Liazo 2006
Along similar lines, Toby Ord, among others, has argued that if the unborn are on a moral par with normal human adults (a claim he calls the Claim), then early embryo loss – which is far, far more common than most people realize – is one of the great ills of our time, something we ought to do everything in our power to stop (a conclusion he calls the Conclusion). But again, this seems absurd.\textsuperscript{96}

Both of these cases more or less tacitly assume that human moral equality, which concerns only the \textit{intentional killing} of our fellow humans, also commits us to a certain view of how we ought to respond to their \textit{non-intentional} (\textit{e.g., natural or accidental}) deaths. More specifically, they assume that if human moral equality is true – that is, if any two killings of any two human beings will be equally morally objectionable, all else equal – then, all else equal, the non-intentional deaths of any two human beings will be equally tragic, and we will be equally obligated to prevent both of them. But this, in turn, seems to presuppose that we have a unified account of the wrongness of killing and the badness of death – that what accounts for why it is wrong to kill a given being at a given time is also what accounts for why it would be bad for that being to die non-intentionally at that time, so that the wrongness of the killing will always be equal to the badness of the death.

As alluded to in Chapter 2, McMahan rejects this presupposition. His reason for this is, roughly, that the badness of a person’s death often seems to come apart from the wrongness of killing him. For example, it is natural to think that the death of a child is more tragic than the death of an adult. Further, the death of a seriously ill person who, had he not died of \textit{that} exact cause at \textit{that} exact time, would only have lingered on in agony for a few more hours before dying of some other complication, may be thought not to be a tragedy at all, but rather a welcome release from suffering.\textsuperscript{97} But though these people seem to differ quite a bit in how bad it would be for them to die, they seem to differ much less, if at all, in how morally objectionable it would be to kill them. For example, it’s not as if it would not be perfectly fine to kill the seriously ill person, even without his consent, simply because his death would be a welcome release from suffering.

This, McMahan argues, is because in the case of persons, the badness of death and the wrongness of killing are explained by different features. The badness of any given death, including the deaths of persons, is determined by the decedent’s time-relative interest in continuing to live, and thus by (a) the quality and quantity of life he loses out on by dying and

\textsuperscript{96} Ord 2008
\textsuperscript{97} McMahan 2001, p. 110
(b) the strength of the prudential unity relations he bears to his future self. McMahan thinks that
(b) will not, in general, vary considerably from person to person\textsuperscript{98}, but (a) seems sufficient to
explain why some of the deaths described above are considerably worse than others. The child
presumably has a greater quantity of life ahead of it than the adult, and therefore, all else equal,
more quality of life in total. Meanwhile, the seriously ill person clearly has neither a high quality
nor a great quantity of life to look forward to. However, the wrongness of killing persons is not
determined by the extent to which it thwarts time-relative interests, which vary from person to
person, but by the fact that it violates what we have seen McMahan call “the requirement of
respect,” which he argues is fundamentally the same for all persons.\textsuperscript{99}

A time-relative interests account of the badness of death also entails that the non-
intentional death of an embryo is not a great tragedy. In this case, as opposed to the cases
described above, the reason for this will lie in (b) rather than (a), since the embryo’s prudential
unity relations to any future entities will be low to nil. McMahan himself, of course, thinks that
it would not be particularly wrong to kill the embryos either, since they are non-persons – and
as non-persons, he thinks, both the badness of their deaths and the wrongness of killing them
is determined by the extent to which it thwarts their time-relative interests. But even if we accept
a broadly McMahanian combination of a “two-tiered” ethics of killing and a time-relative
interests account of the badness of death, it is not given that our version of the two-tiered ethics
must draw the boundary between the tiers in quite the same way McMahan draws it. In
particular, it may be that the higher tier, whose members all have the same fundamental right
to life, contains all human beings. Given such a combination of positions, the embryo will be
in much the same position as the seriously ill person (though, again, for different reasons) – its
death would not necessarily be a great tragedy, nor something we are obligated to prevent, at
least not to nearly the same extent we are obligated to prevent the death of a normal, healthy
human person. But to kill it would nevertheless be a moral outrage.

In short, then, I see no particular reason (beyond reasons having to do with the
plausibility of each of these views taken independently) why human moral equality could not
be held in conjunction with a time-relative interests account of the badness of death. But if the
adherent of human moral equality goes for this combination of views, he will no longer be
committed to rejecting our intuitive reactions to early embryo death or the Embryo Rescue
Case. In both of these cases, he can maintain that although it is just as wrong to kill an embryo
as it is to kill you or me, for just the same reasons, its non-intentional death, be it through a

\textsuperscript{98} McMahan 2001, p. 233
\textsuperscript{99} McMahan 2001, ch. 3, sct. 3.2
hospital fire or natural embryo loss, is not nearly as much of a tragedy as your death or mine would be under equivalent circumstances, nor something we are obligated to make an equivalent amount of sacrifices to prevent.

Further, even if the advocate of human moral equality rejects the time-relative interests account of the badness of death and embraces a unified account of the wrongness of killing and the badness of death, the two arguments outlined above still have several other weaknesses. In both of these arguments, the basic conclusion takes the form of a two-part conjunction: Human moral equality commits its adherents to some concrete moral judgment, and that judgment is obviously absurd. In some cases, the counterargument will, as above, be meant to show that the first conjunct is false, and that human moral equality does not, after all, require its adherents to bite the bullet on these cases. In others, the counterargument will be meant to show that the second conjunct is false, and that biting the bullet is more reasonable than it seems.

For instance, one indication that something is amiss with the Embryo Rescue Case is that the apparent absurdity of the conclusion seems to depend crucially on the location of the embryos. The embryos in the Embryo Rescue Case are not located within a womb, and intuitively, this is a big part of why it seems so implausible to ascribe a right to life to them in this case. It may or may not be false that embryos have a right to life when located within a womb, but it is certainly not obviously false in the way that ascribing the same sort of status to the embryos in the IVF lab seems to be. But location is surely not a morally relevant attribute – one does not gain or lose a right to life simply by moving from one place to another. (Incidentally, the moral irrelevance of location changes is a point that is also stressed by many pro-choice philosophers. For instance, such a premise is a key part of Peter Singer’s argument against viability as a criterion of fetal moral status. Elsewhere he has also famously used similar premises to argue that we have considerable obligations towards the global poor.) So, given that the location of the embryo is irrelevant to its moral status, we have two conflicting appearances: Ascribing moral status to an embryo within a womb does not seem obviously false, but ascribing it to an embryo in an IVF lab does seem obviously false. (To be clear, ascribing the right to life to an embryo within a womb may or may not be false, but it certainly doesn’t seem obviously false – not, at any rate, as obviously false as ascribing it to an embryo in an IVF lab seems to be. This is shown by the very fact that the critic feels the need to appeal to the Embryo Rescue Case at all: If claiming that an embryo in a womb has the right to life was as obviously false as claiming the same thing about embryo in an IVF lab, there would be

---

100 Singer 2011, p. 127
101 Singer 1972
no need to go through the obvious falsehood of the second claim in order to show the obvious falsehood of the first. Instead, the obvious falsehood of the first claim could just be pointed out directly.)

Now, it could be that the Embryo Rescue Case simply highlights an absurdity that lies latent in supposing that embryos have a right to life in the womb. But then again, we might just as well reverse this reasoning, arguing that the apparent relative non-absurdity of supposing that embryos have a right to life in the womb shows that the view that we ought to rescue the embryos from the IVF lab is less absurd than it seems. Either way, the effect is the same: In the name of coherency, we sacrifice one appearance to save another. How do we break the standoff? How do we choose which appearance to sacrifice? The only way that suggests itself is to appeal to some theory which can deliver a general verdict on the moral status of embryos. But in that case, it is that theory, not the Embryo Rescue Case, that is doing the real work. The case itself proves nothing. This is consonant with Liao’s conclusion that “if one needs to invoke something like … [McMahan’s] time-relative interest account to show that embryos are not rightholders, then the Embryo Rescue Case is not making its case independently”\(^{102}\).

Let’s move on to Ord. The claim he calls the Conclusion – that early embryo death is a very great evil, and something we ought to do a great deal to prevent – does indeed seem absurd. But I suspect that this appearance has at least one source which, when examined more closely, should lead us to doubt whether Ord’s argument is quite as devastating for human moral equality as he takes it to be. What I have in mind is the simple fact that early embryo death is invisible to almost all of us almost all the time, since it almost always happens before pregnancy is detected or detectable. Suppose that one day, expectant mothers all over the world are struck by a sudden, Scourge-like spike in miscarriages (call it the Second Scourge). Unlike early embryo deaths, the Second Scourge happens long after pregnancy has become noticeable, but before most pro-choice philosophers would be prepared to grant any significant moral status to the fetus – near the end of the first trimester, say. Would it be as apparently absurd to get worked up about the Second Scourge as it would be to get worked up about early embryo death? Clearly not. In fact, most people – including, especially, the mothers of its victims – probably would get worked up about it. But where does the difference lie here? Not, as far as most people are concerned, in any differences in moral status: Most pro-lifers would claim that both the embryos and the victims of the Second Scourge have the right to life, while most pro-choicers would

\(^{102}\) Liao 2006, p. 146
deny that either group has the right to life. No, the key difference seems to be that in the latter case, people would notice.

This suggests the following. The reason why we find it hard to get worked up about early embryo death is that it is almost always entirely invisible to us, and has no noticeable impact on our lives. In other words, we find it hard to get worked up about early embryo death for roughly the same as the reason why we are much more strongly affected by – and would be prepared to do much more to prevent – the death of a close friend or a family member than the death of a stranger on the other side of the globe. In general, this kind of moral favoritism may or may not be legitimate, or at least unavoidable. After all, none us have the time or resources to carry all the ills of the world, so perhaps it’s better that, for the most part, we all limit ourselves to sweeping in front of our own doors. But the point is that, even if acceptable, this sort of favoritism certainly should not be predicated on a difference in moral status. For example, the reason why I care strongly about the death of a friend or family member, and would be prepared to move Heaven and Earth to stop it, is not (and should not be) that he gains additional moral status by dint of being my friend or family member. So, the reason why we instinctively care relatively little about early embryo death is not, as Ord suggests, because “few people really believe that full moral status begins at conception,” just as my relative lack of concern for the deaths of strangers should not be taken to show I secretly believe that full moral status begins only when you get to know me.

4.12 The Mix-Up
The Scourge and the Embryo Rescue Case are counterexamples to human moral equality in general, not counterexamples to my particular, genetic way of justifying it. The latter type of counterexample is in shorter supply, but does exist. For instance, in a critique of Liao’s genetic account, Christopher Grau offers the following case, which I will call The Mix-Up:

[D]ue to your being exposed to a new and dangerous type of radiation, your child (let’s call her Betsy) suffers from a genetic disorder that has left her alive and conscious but with a level of intelligence that falls well below what is necessary for ever exercising moral agency. This disorder is new, very rare and not well understood, but the doctors, having performed many tests, assure you that, though your child’s genetic damage does permanently undermine her capacity for moral agency, the damaged genetic material is not among the material that makes up ‘the genetic basis for moral agency’. As a Liaoian, you breathe a sigh of relief, confident that your child is in fact deserving of the moral status we extend to most humans. However, imagine that several months later

---

103 Ord 2008, p. 19
the doctor calls back: there’s been a terrible mix up, and your child’s test results were swapped with those of another severely disabled child. It turns out that your child does, as a matter of fact, suffer from genetic damage to the relevant portion of the genome responsible for moral agency, for she is missing a single highly relevant gene. (‘Sorry honey, the new results are in, and it turns out Betsy may not be a rightholder after all!’)\textsuperscript{104}

At least two separate arguments against the genetic account can be read into The Mix-Up. Firstly, The Mix-Up could be taken to show that the genetic account is overly narrow, relative to its own ambitions. For as we have seen, one of those ambitions is to explain how it could be that all human beings are rightholders. But surely there is a good \textit{prima facie} case to be made that Betsy, which the genetic account would apparently deem a non-rightholder, is human despite her genetic defects. Secondly, it could be taken to show that the genetic account identifies the source of our rights with something that has no moral relevance. Grau focuses mostly on the second point. “Surely,” he writes, “there is a problem with any account of moral status that could (even just theoretically) put you in this position. How can something which matters so much (your child’s moral status as a rightholder) reasonably hinge on something that seems to matter so little (whether damage to some portion of genetic material should be construed as removing a capacity or instead permanently undermining it)?”\textsuperscript{105}

Against the first point, we must first observe, as Liao does, that the genetic account does \textit{not}, in fact, entail that Betsy is not a rightholder. The account would have this implication if it claimed that having the genetic basis for moral agency (or something functionally equivalent) is a \textit{necessary} condition of rightholding, but all the genetic account actually claims is that it is a \textit{sufficient} condition of rightholding.\textsuperscript{106}

A possible counter-response is that, even if the genetic account does not strictly speaking \textit{entail} that Betsy is not a rightholder, it also gives us no plausible way of maintaining that she is. For it is hard to see what attribute of Betsy’s we could plausibly identify as her “rightsmaking” feature, at least if we don’t want to go full speciesist and identify that property with her membership in the human species. (This is what Grau ends up arguing that Liao ought to do.)\textsuperscript{107}

I have two things to say to this. Firstly, as we will see later on in this chapter, there is a certain theory on which even human beings like Betsy do have something functionally

\textsuperscript{104} Grau 2010, pp. 390 & 391
\textsuperscript{105} Grau 2010, p. 391
\textsuperscript{106} Liao 2012, p. 267
\textsuperscript{107} Grau 2010, p. 395 & 396
equivalent to the genetic basis for moral agency (or rationality), and therefore count as rightholders on the genetic account. But that theory is the substance view in its standard, metaphysically heavyweight form, which I cannot appeal to here, given my official agnosticism about it. But secondly, it is not clear that The Mix-Up (or at least Grau’s philosophical interpretation of it) describes a possible scenario at all, simply because it is not clear that there could exist a living human organism who lacked the genetic basis for moral (or rational) agency. The reasons for this, recall, were outlined in Chapter 3. Firstly, it is not at all clear that a being like Betsy would count as a member of the human species at all (on either a phenotype-based “folk” notion or a genotype-based “scientific” notion of species). Secondly, given the complexity of the genetic basis for moral agency and the fact that it likely also provides the genetic basis for many other important features, it is not clear that Betsy’s genetic defects would be compatible with life in the first place.

Against the second point, Liao points out that, if taken seriously, it proves far too much. Any account of moral status that grounds it in some objective and empirically detectable attribute (be it sentience, species membership, or whatever) seems to allow for the possibility of a situation analogous to The Mix-Up. If the attribute is objective, it is at least in principle possible to be wrong about whether it is present in a certain being; and if the attribute is empirically detectable, this mistakenness could have its origin in the basis of something like a mix-up at the lab.

4.13 Interests
Another objection to the Genetic Criterion is that beings that have second-order rationality do not necessarily have interests, which are necessary for moral status. This objection presumably rests on the idea that interests require a degree of psychological sophistication that some entities with second-order rationality (including, presumably, most or all fetuses) do not have. Call this the psychological view of interests. Responding to this objection, Liao notes that the psychological view is a contentious notion of interests that cannot simply be taken for granted. There are other, equally respectable notions of interests on which animals – and even non-sentient organisms like plants – can have interests.

Let’s try to say a bit more. On one simple and obvious form of the psychological view, interests presuppose desires. Call this the desire view. And in its simplest form, the desire view claims that the things that are in my interest just are the things that I actually desire. But this

---

108 Liao 2012, pp. 271 & 272
form of the desire view is quite evidently false. It is clearly possible to actually desire things that are not in one’s interest, as when I desire to drink a glass of water that has, unbeknownst to me, been poisoned; and it is equally possible to have things in one’s interest that one does not actually desire, as when, after drinking the poisoned water, I don’t know that there is an antidote to the poison in my medicine cabinet, and therefore have no desire to go and get the antidote from my medicine cabinet.

But as the familiarity of these examples suggests, Boonin’s account of desires, which was described in Chapter 2, may give us the resources to formulate a more sophisticated version of the desire view. For while not everything I actually desire is in my interest, there is surely some plausibility to the idea that the things that are in my interest just are those things that I desire ideally – i.e., those things that I would desire if the circumstances were ideal.

This claim may be true in the strictest sense – it may indeed be the case that something is in my interest if and only if I desire it ideally. But to say that something is in my interest because I desire it ideally, and not the other way around, is surely to put the cart before the horse. In the broadest sense, I take it, something is in my interest when it is good for me. But surely the things that are good for me are not good for me because I desire them, even ideally – rather, I desire them because I take them to be good for me, and when I desire them ideally, the desire is ideal because I am (or would be) correct in taking them to be good for me. In other words, the ideal circumstances for desire just are those circumstances in which interests and desires coincide. Thus, it is hard to see how we can even make sense of the notion of desiring under ideal circumstances if we do not have a prior and independent notion of interests.

Secondly, this version of the desire view suggests that beings without actual desires cannot have interests, since ideal desires (at least on Boonin’s view) are only possible where there are actual desires. But again, this is surely false. Surely, for example, there are things that are good or bad – harmful or helpful – for lower animals, and perhaps even for non-sentient nature. As living organisms, they can flourish or falter, and it seems very nearly a tautology to say that it is good for them to flourish rather than falter. Now, the interests of these entities probably carry much less moral weight than the interests of persons (and perhaps even none at all), but that’s beside the point here – the point is just that they do have interests.

McMahan’s notion of time-relative interests, also surveyed in Chapter 2, gives us a more subtle and plausible way of cashing out the psychological view. Time-relative interests do not presuppose desires, but they do require that one bear sufficiently strong prudential unity relations to some future entity, and therefore seem to require a strong degree of psychological continuity with that future entity (either in the form of sameness of psychological states, or, as
McMahan would have it, in the form of sameness of psychological capacities. But like the desire view, this version of the psychological view clearly cannot account for the interests of entities that have no psychological connections to anything, such as non-sentient organisms. And, just as time-relative interests cannot account for the interests of the “lowest” organisms, they also cannot account for our obligation towards the “highest” organisms – i.e., towards those organisms that are persons. As we have seen, McMahan himself admits this quite freely, consigning persons to a higher, separate tier of his ethics of killing and arguing that the wrongness of killing them is not accounted for by the fact that it thwarts their time-relative interests. But as should soon become apparent, the question at issue may be thought of precisely as whether or not the unborne (and other human beings “at the margins of life”) fit within that higher tier. From a McMahanite perspective, at least, dismissing the Genetic Criterion because some beings who have the genetic basis for rationality do not have time-relative interests is therefore circular, since part of the question at issue is precisely whether or not beings who have the genetic basis for rationality are among those entities whose moral status is determined by their time-relative interests.

4.2 Why the Genetic Criterion?
Having considered some objections to the Genetic Criterion, I will now motivate it. In broad strokes, the argument will go as follows. First, I will sketch out one account of why rationality itself (i.e., first-order rational dispositions of the right sort) is sufficient for the right to life. Then I will argue that this account also commits us to the view that second-order rational dispositions generate the same sort of right to life as do the corresponding first-order rational dispositions – that the “rightsmaking” features of first-order rationality are also present in second-order rationality. Finally, I will argue that this conclusion generalizes to many accounts of moral status other than the one I offer here.

4.2.1 How does first-order rationality generate rights?
Is (first-order) rationality sufficient for the right to life? Intuitively, it certainly seems so. There’s a reason why rationality is a very old and very popular criterion of full moral status – one that can arguably be read into thinkers as diverse as Aristotle and Kant – and there’s reason why many other proposed criteria of full moral status seem to presuppose it, as emphasized in previous chapters. Another way to tease out the plausibility of rationality as a sufficient criterion of the right to life is a thought experiment like the following. Suppose you, the president of the Earth federation, get word that a species of aliens is coming to visit your planet. Apart from the
fact that the aliens come in peace, all you know about them is that they’re rational. As president, it is your responsibility to decree how earthlings should treat the visitors. Should you decree that we treat them in the same way we ought to treat each other? That certainly seems like the advisable – and quite possibly even the morally obligatory – course of action.

Now, this does not, strictly speaking, seem to show that the aliens have the same rights as us – it could just be taken to show that we have a duty to treat them as if they did. But in general, the fact that we have a duty treat Fs as if they were G is, I take it, prima facie evidence that Fs really are G. Further, it’s not clear that there really is such a thing as merely treating someone as if he had a certain right – if he ought to be so treated, then ipso facto, he really has that right. For to say that Bob has a right to X over Alice in circumstances C is surely just to say that Alice has a duty to endow Bob with X in C. But to say that Alice has a duty to treat Bob as if he had a right to X over her in C is surely just to say that Alice has a duty to act as if she had a duty to endow Bob with X in C. But the claim that you have a duty to act as if you had a duty to perform a certain action surely entails (and is quite possibly equivalent to) the claim that you really do have a duty to perform that act. Hence, Alice’s duty to act as if she had a duty to endow Bob with X in C entails that she has a duty to endow Bob with X in C. But again, this is just to say that Bob really has a right to X over Alice in C.

Further, it is clear that if rationality is sufficient for full moral status, then dispositional rationality is probably sufficient for it on its own. Again, temporarily comatose people do not lose their moral status in virtue of losing their occurrent rationality. We could, strictly speaking, hold that the comatose person had his moral status in virtue of his occurrent rationality while he was awake, but in virtue of something other than his dispositional rationality when in the coma. But this hypothesis seems far too messy and ad hoc.

4.211 Rationally specific goods
So it is reasonably clear, I take it, that first-order rationality generates the right to life. What is less clear is how it generates the right to life. What sort of morally relevant difference does it make? Intuitively, it must make a morally relevant difference by being good in some way. And intuitively, the goodness of rationality has a lot to do with the goods that rationality generates. There is a long and diverse list of good things, some of them clearly very good, that can only fully and reliably be created or enjoyed by rational beings – things like electric cars, Beethoven symphonies, brain surgery, and second-order logic. The list goes on and on. Call such goods rationally specific goods.
Note that some rationally specific goods, such as certain types of natural beauty, can be created by non-rational beings, but can only be enjoyed by rational beings. (Further, many rationally specific goods can be enjoyed in several distinct ways, only some of which are exclusive to rational beings. It probably takes a rational being to appreciate the beauty of a pristine woodland stream, but a non-rational being can certainly use it to slake its thirst.) Others, such as veterinary medicine, can be enjoyed by non-rational beings, but can only be created by rational beings. Others yet, such as Plato’s dialogues, seem to be such that rational beings have a monopoly on both creation and enjoyment.

Nor, by the way, are all the goods we create and enjoy rationally specific. As rational beings, we partake of certain goods that non-rational organisms do not partake of, but there are also many goods we have in common with them – food, sleep, sex, and so on. Further, some rationally specific goods, such as the self-sacrifice of a philanthropist, are moral goods, while others, such as the paintings of a brilliant artist, only seem to be non-moral goods (aesthetic goods in this case). And, while some rationally specific goods are instrumental goods, others may well be intrinsically good. (For instance, aesthetic experience, which is plausibly a rationally specific good, has been included in several taxonomies of intrinsic goods, notably those of William Frankena and G. E. Moore.110) In short, rationally specific goods can, on the face of it, be good in just about any way, shape, or form – the only thing that unifies them is that they are good, and that, if there were no rational beings, these goods either would not exist or would not be good for anyone, at least not in the right way.

Note also that I do not claim that rational beings have a monopoly on the creation or enjoyment of good things, nor even on the creation and appreciation of very good things or the best things. I only say that they have a monopoly on the creation and appreciation of some of the best things. And though some rationally specific goods may be intrinsic goods, as just suggested, I am certainly not committed to the premise that any of them are intrinsically good. The only important thing is that they are very good, be it intrinsically or instrumentally. Hence, even someone like an axiological hedonist, who thinks that rational entities have no monopoly whatsoever on the creation or appreciation of those things that are intrinsically good, could coherently accept what was said above, since rational beings may still have a monopoly on the creation or enjoyment of some of the highest instrumental goods. Consequently, nothing that will be said here seems to me to imply that non-rational beings (and, in particular, non-human animals) cannot have significant moral status, nor even, in principle, that they can have the

110 Frankena 1973, pp. 87 & 88; *Principia Ethica* §113
same moral status as rational beings. Non-rational animals are indeed excluded from creating or enjoying certain very great goods, but as far as I can see, nothing about what I just said suggests that they cannot create and enjoy other goods, nor even that some of these goods may be just as good as (or even better than) the rationally specific goods. Further, and as noted, it is quite coherent to say that non-rational beings can create or enjoy a given rationally specific good. The only thing we cannot coherently say is that a given rationally specific good can be created by non-rational beings and that there is no way of enjoying it that is exclusive to rational beings.

Finally, it should be noted that as well as generating many very great goods, rationality also seems to generate many very great evils – clinical depression, industrialized genocide, and what have you. However, I do not think this should shake our conviction that rationality is, all in all, a very good thing. For my own part, this belief is based on two convictions. I freely admit I cannot prove these convictions, but I also trust that most of my readers will share them, at least on reflection. Firstly, though rationality produces evils, it probably produces much more good than evil, all in all. If it didn’t, the existence of rational beings would be a bad thing all in all, and some form of anti-natalism would apparently be called for. But that, I trust, is a bullet most people won’t be willing to bite (though some will, of course). Secondly, rationally specific goods seem to be essential to rationality in a way that rationally specific evils are not. I, at least, find it much easier to conceive of how there could be rational beings even in a world where no rationally specific evils were produced or suffered than how there could be rational beings even in a world where no rationally specific goods were produced or enjoyed.

4.212 How rationally specific goods generate rights

In short, then, our (first-order) rational dispositions include or entail certain dispositions to create very great goods. But in what way does this generate a right to life?

A simple and obvious answer is the following, broadly consequentialist one. If your rational dispositions include dispositions to create and enjoy very great goods, then to kill you will normally be to bring it about that you will not create or enjoy certain very great goods that you would otherwise have created or enjoyed. This is a very serious harm, not only to you, but also to all the people who would have enjoyed the goods that you could have created.

Note that this account does not seem to be consequentialist in the strong sense – that is, it does not seem to commit us to the view that the permissibility or impermissibility of an action is determined *entirely* by its consequences. All it says that the wrongness of a particular type action, namely killing rational beings, is determined at least partly by its harmful consequences.
But the view that we should not perform actions that result in serious harm is hardly unique to consequentialist ethical theories. Rather, it – or at least some type of regard for the consequences of actions – is a feature of any sane system of ethics.

Note also that while the consequentialist account bears some resemblance to Marquis’ account of the wrongness of killing (outlined in 2.1), it need not necessarily rest on the sort of heavyweight assumptions about personal identity that Marquis’ account seems to rest on. It is true that I just said some things which suggest that killing a rational being is wrong because it prevents *that very rational being* from creating and enjoying rationally specific goods in the future, but that can be regarded as loose talk. The essential point is surely that by killing a rational being, one normally prevents the creation and enjoyment of very great goods that would otherwise have been created and enjoyed. It’s not obvious that it matters *who* would have been doing the creating and enjoying, and, in particular, whether some of the creating and enjoying would have been done by the very same being as the one that was killed. For the same reason, *mutatis mutandis*, it does not, on this account, seem to matter whether the victim bears strong prudential unity relations (or prudential unity relations of any kind) to any future creators or enjoyers of rationally specific goods. The point, again, is just that there are (or would) be such creators and enjoyers at all.

The consequentialist account has at least three major problems. Firstly, there is the *instrumentality problem*. The consequentialist account makes the value of rational persons instrumental and extrinsic. On the consequentialist account, the reason why it is wrong for me to kill you is not because you have any inherent dignity or value, but because you are a producer of value. This seems unsavory.

Secondly and relatedly, there is the *equality problem*. The consequentialist account does not secure the sort of basic moral equality that many moral philosophers want to secure. This is not to say that it fails to secure full human moral equality – to make that a desideratum at this point would smack of circularity. Rather, the point is that it fails to secure the moral equality of human *persons*. In other words, it does not do justice to the fact that, in some sense, you and I seem to have the exact same right to life, on the exact same grounds. On the consequentialist account, the wrongness of killing me – and the badness of my death – seems to be a function of how many rationally specific goods (and other goods, for that matter) it deprives the world of. In other words, killing me can apparently be more or less bad, depending on how many goods would otherwise have been created and enjoyed in the future. But it is obvious that people differ radically in how many future goods their killings take out of the world. The consequentialist account therefore suggests a radical moral inegalitarianism. Killing a brilliant
artist or scholar – a person who would otherwise have created unusually great goods – would be much more wrong than killing an average person. Killing a person who would shortly die anyway, with or without his consent, would hardly be wrong at all, since it presumably prevents very few goods from being created and enjoyed. And so forth.

Finally, there is the nonconception problem. It is hard to see how the consequentialist account can explain why failing to bring a certain rational being into existence is almost always much less wrong than killing that being, and often not wrong at all. Suppose my parents had decided not to have children, or had decided to have children at a later or earlier time than they actually did, so that I never came into existence. (For convenience, and with all due respect to Plato, I’ll disregard the possibility that we exist before conception.) Then the world would have been deprived of all the value I would otherwise have put into it. In other words, the consequences of the action, which the consequentialist account identifies as the source of the wrongness of the act, would have been the same as the consequences of killing me now. In fact, the consequences would have been worse if I had never come into existence. Killing me at this moment would only deprive the world of all the goods I would have put into it from this moment, whereas failing to bring me into existence would not only have deprived the world of those goods, but also of all the goods I have so far put into it.

One possible response to the nonconception problem is that killing me is an act, while failing to bring me into existence is a mere omission. But even setting aside questions about the moral relevance of the act-omission distinction, this response does not solve the problem, for it suggests that not bringing me into existence is morally equivalent to causing my death by omission. But causing my death by omission – e.g. by not giving me the antidote to a lethal snake bite – is still seriously wrong under most circumstances.

A (partial) fix for these problems is available. So far, I have not tried to account for why rationality tends to produce such great goods. Rather, I have simply rattled off a list of such goods and assumed that the reader will share my conviction that the list could be expanded almost indefinitely. But an explanation is surely called for. If an author consistently produces great novels, we tend to credit those novels to the traits of the author himself – it would be a miracle if, like a literary Forrest Gump, he just happened to stumble into one brilliant work after another. Similarly, when a certain human faculty produces so many good things, it behooves us to ask what it is about that faculty that allows it to do so. And there is, it seems to me, a simple answer to this question: that rationality itself is an intrinsic good, and a very great intrinsic good at that. In fact, it is not unreasonable to suppose that rationality itself is at least as good as the greatest rationally specific goods, since you cannot give what you do not have.
In other words, the great goodness of rationally specific goods may, in a sense, be something they have on loan from rationality itself.

If rationality is intrinsically good, a revision to the consequentialist account is called for. On this revision (call it the intrinsic-value revision), rationality does not endow the right to life because of the goods it may bring forth, but because it is good in and of itself. Killing rational beings is not wrong because it prevents the coming-into-existence of rationally specific goods, but because it destroys an already existing good, namely their rationality itself.

The intrinsic-value revision seems to solve the instrumentality problem, since it identifies an intrinsic and intrinsically valuable feature of rational beings, rather than the external goods they will or might produce, as the source of their moral status. It might be said that to fully solve the instrumentality problem, we must say that what is intrinsically valuable are rational beings themselves, not their rationality as such. On this view, the rationality of rational beings is the thing in virtue of which those beings are intrinsically valuable, rather than being the very thing that is intrinsically valuable about those beings. (McMahan takes this sort of view.\textsuperscript{111}) I tend to think that this revision of the intrinsic-value revision is beneficial, for reasons we will shortly see.

Secondly, the intrinsic-value revision seems to solve the nonconception problem. The source of the nonconception problem was that, in its original form, the consequentialist account asserted that killing rational beings is wrong in virtue of an attribute that it has in common with failing to bring such beings into existence, namely that it normally prevents certain goods from being brought into the world. But on the revised version, killing is wrong because it destroys an already existing intrinsic good, which is clearly not a feature it shares with most acts of nonconception.

It might be objected that in a way, the original, non-revised version of the consequentialist account also claims killing is wrong because it destroys an already existing good, namely the (instrumental) good of a particular rational being’s first-order rationality. However, it seems to make a difference that on the non-revised version, the good in question is merely instrumental. For when it is wrong to destroy a merely instrumental good, it is surely wrong because it prevents that good from producing the goods it would have otherwise produced, so that it would have been equally wrong, all else equal, to prevent the production of those goods in some other, less destructive way. For example, it would, under normal circumstances, be wrong of me to blow up a truck carrying lifesaving medicines to a remote

\textsuperscript{111} McMahan 2001, ch. 5, sct. 2.2
village. But it would surely be just as wrong, for just the same reasons, if, instead of blowing up the truck, I went to the factory where the medicines are made and non-destructively sabotaged the machinery in a way that prevented the medicines from ever being manufactured. So, if rationality is merely an instrumental good – good because it brings forth certain other goods – then destroying any particular instance of rationality is wrong only because (or to the extent that) it prevents the coming-into-existence of those goods that would otherwise have been brought forth by that instance of rationality. But again, this is a feature it has in common with preventing the coming-into-existence of that rationality.

However, the fix does not seem to solve the equality problem. Even if rationality is an intrinsic good, it remains the case that rational beings have rationality in different degrees. And if killing rational beings is wrong because it destroys their rationality, it is hard to avoid the conclusion that the killing’s degree of wrongness is a function of the amount of rationality it takes out of the world, so that killing more highly rational beings will come out worse than killing less highly rational beings.

4.22 (How) does second-order rationality generate rights?

Do the rightsmaking properties of first-order rationality carry over to second-order rationality? I shall now argue that the preceding account suggests so. And, unlike the preceding account, an account that grounds the right to life of rational human persons in their second-order rationality stands a good chance of avoiding the equality problem. But of course, such an account will also extend the right to life to certain beings that do not have first-order rationality, including all members of the human species who do not presently have it. Finally, I will argue that this feature of the consequentialist account – the fact that the property it pinpoints as the rightsmaking property of first-order rationality is a property that is also had by second-order rationality – is a feature it shares with many other accounts of moral status.

4.221 The consequentialist account again

Recall that on the consequentialist account, killing a rational being is wrong for two reasons. Firstly, it deprives the world of future rationally specific goods. Secondly and much more importantly, it destroys an already existing intrinsic good, namely the rationality of that rational being. Do these features carry over to the killing of a bearer of second-order rationality?

The first point seems to carry over. Second-order rational dispositions are clearly also dispositions to create and enjoy rationally specific goods, albeit of a more indirect sort than first-order rational dispositions. According to the first point, the fact that we are rational...
rights because it causes us to create and enjoy certain very great goods. Our second-order rationality, for its part, causes us to be (first-order) rational. But causation is plausibly a transitive relation – that is, if X causes Y and Y causes Z, then X causes Z. Therefore, our second-order rationality causes us to create and enjoy rationally specific goods just as much as does our first-order rationality – it shares its “rightsmaking” properties. Of course, some beings will never actually realize their second-order rational dispositions, and will therefore never be in a position to create or enjoy rationally specific goods. But then, this is equally true of first-order rational dispositions: Some beings with first-order rational dispositions will never realize them again, and some will realize them less than others. In effect, this is just a higher-order version of the equality problem.

Does the second, more important point of the consequentialist account carry over? At first blush, it seems not. First-order rationality may indeed be intrinsically good, but of course, not all beings with second-order rationality are in possession of its first-order counterpart. Now, if our first-order rationality is intrinsically good, then our second-order rationality will presumably be instrumentally good, since it is our second-order rationality that produces our first-order rationality. But as we saw in the previous section, there are at least three reasons why this kind of merely instrumental goodness can’t secure the sort of moral status we’re after here.

But recall how I argued for the intrinsic value of our first-order rationality. Firstly, it helped us explain how our first-order rationality produces so many good things. Another, more implicit point in favor of this move was its explanatory power – the fact that it allowed us to mend two of the problems the consequentialist account started out with. But both of these arguments can also, mutatis mutandis, be given in favor of imputing the same sort of intrinsic value to our second-order rationality and making it the ground of our right to life. Like first-order rationality, second-order rationality produces great goods. In fact, our second-order rationality produces more great goods than our first-order rationality, for not only does it indirectly produce all the goods our first-order rationality produces, it also produces the good of our first-order rationality itself. Secondly, ascribing intrinsic value to our second-order rationality – and grounding our right to life in the badness of destroying this value – may let us mend, and will certainly let us ameliorate, the one problem that was not mended by imputing intrinsic value to first-order rationality, namely the equality problem. For as I argued in the previous chapter, while we differ in our degree of first-order rationality, it is plausible that we differ much less, if at all, in our degree of second-order rationality. Killing a rational human

---

112 Though for arguments to the contrary, see e.g. Hall 2000
being will not always destroy the same amount of first-order rationality, but it will, at least to a much closer approximation, always destroy the same amount of second-order rationality. Grounding our right to life in our second-order rationality, unlike grounding it in our first-order rationality, therefore solves or ameliorates the equality problem and secures the basic moral equality of all human persons. And in doing so, it also secures the basic moral equality of all or virtually all human beings, since all or virtually all human beings have second-order rationality, as I also argued in the previous chapter.

It may be objected that this seems like the start of an unsavory regress. If our second-order rationality is intrinsically good because it produces our first-order rationality, then surely whatever produces our second-order rationality must also be intrinsically good, and likewise for whatever produced that, and so on to infinity.

Notice, though, that the character of this regress changes radically once we move from second-order rationality to its source, in such a way that we seem quite within our rights to end it at second-order rationality, before it has really even begun. For when the regress moves beyond our second-order rationality, its relata cease to be higher orders of our rationality. There is no such thing as “our” third-order rationality. There is, of course, something that produces our second-order rationality – normally a sperm and an egg – but it is not something about us. We do not have third-order rational dispositions, for there is nothing intrinsic to us that can serve as the base for such dispositions. Rather, our genomes – and hence our second-order rationality – come about through a complex interplay between the genomes of our parents and the external environment. Thus, when it moves past second-order rationality, the regress moves from our dispositions to (possibly non-dispositional) features of the external world. And thus, imputing intrinsic goodness to whatever produces our second-order rationality, unlike imputing it to whatever produces our first-order rationality, does not improve our account in any way. It does not serve to further explain what we wanted to explain here, namely where the value and dignity of rational beings comes from, and does not seem to mend any further problems with the consequentialist account. In short, then, we have good reasons to end the regress at our second-order rationality.113 These reasons become even better if, as suggested earlier, we assert that it is rational beings that are intrinsically good, and that they are intrinsically good in virtue of their rationality. For again, while our first- and second-order rationality are intrinsic

113 Also, even if do keep the regress going for a bit longer, I am not sure that this makes it all that unsavory, nor that it needs to go on forever. Perhaps we should instead be led to conclude that if we follow the regress of goodness far enough back, it must eventually terminate in an ultimate, highest good from which all other goods get their goodness. I note without further comment that this sounds more than a little like the beginning of an argument for the existence of God, and in fact bears a passing resemblance to the fourth of Thomas Aquinas’s five ways.
properties of ours, whatever produces our second-order rationality is not. But it is hard to see how a being could possibly be intrinsically good in virtue of something external to it.

There is an important caveat here. If we have essences, in the Aristotelian sense introduced earlier in this work, then there is plausibly some sense in which they are the substrates out of which our genomes flow, much as our genome is the substrate out of which our rationality flows. If we do have such essences, therefore, we do have an internally grounded third-order rationality, since these essences can then serve as intrinsic bases for this third-order rationality. If we do have such third-order rational dispositions, I see no reason not to make them the grounds of our right to life. In fact, such a move would also have an additional theoretical benefit (or at least it counts as a benefit if human moral equality is a desideratum), namely that, as alluded to in the previous chapter, we would no longer have to contend with what sort of moral status to bestow on human organisms who do not have the genetic basis for rationality (if such organisms do or could exist). In effect, it would remove the last lingering traces of the equality problem. And, for the same reason that we previously stopped the regress at our second-order rationality, we should now stop it at our third-order rationality. For the sources of our essences (if such there be) are plausibly external to us, and so, we have no fourth-order rational dispositions. But enough about that – I was supposed to stay officially agnostic about whether we have Aristotelian essences.

Another objection is that it is not as clear as I may have made it seem that first-order rationality, let alone its second-order counterpart, really is intrinsically good. The standard test of whether something is intrinsically good is whether it is desirable for its own sake, and not as a means to other ends. But is it obvious that rationality is desirable in this way? Most of us find rationality desirable, of course, at least insofar as we would like to retain our own rationality. But is that because we value our rationality for its own sake, or is it simply because we value the rationally specific goods it brings forth?

Two things can be said against this. Firstly, I am not, in fact, committed to the claim that (first-order or second-order) rationality is intrinsically good. As mentioned previously, I could say that, but I could also say that it is rational beings themselves that are intrinsically good, in virtue of their rationality. And to my mind, at least, imputing intrinsic value to rational beings seems more immediately plausible than imputing it to rationality itself. Secondly, the theoretical costs of denying the intrinsic goodness of both rational beings and their rationality seem great enough that they outweigh any of the prima facie implausibility of either of these claims. Even if they are both somewhat implausible, denying them will only end up committing us to other views that are deeply implausible. For if we deny both the intrinsic goodness of our
rationality and the intrinsic goodness of rational beings, it seems impossible to avoid the instrumentality problem, and difficult to avoid the equality and nonconception problems. Some people (notably certain utilitarians) will no doubt be willing to bite this bullet. But I am not—and more importantly, nor are the people on the other side of the conversation to which this work is meant as a contribution. We have already seen how McMahan complicates his account considerably in order to save the Equal Wrongness Thesis (i.e., in my jargon, in order to avoid the equality problem). Boonin, too, seems anxious to save the equal and inalienable value of persons: On the basis that “killing an innocent person is supposed to be impermissible even if it makes the world contain more good (say, by preventing two other innocent people from dying)”\textsuperscript{114}, he argues that the moral status of persons is grounded in what he calls their \textit{integrity value}, not in what he calls their \textit{existence value}\textsuperscript{115}, where a thing has existence value when it is valuable “in the sense that … [it is] good thing to have around,” and integrity value when it is valuable “in the sense that … it has a right not to be destroyed.”\textsuperscript{116}

As promised, the foregoing also does not commit us to the view that having the genetic basis for rationality must be necessary as well as sufficient for the right to life. In fact, it very strongly suggests the opposite. For on the foregoing, the genetic basis for rationality endows us with the right to life because it endows us with second-order rationality. Hence, any being that has second-order rationality also has the right to life, even if it does not have the same genome as us. And second-order rationality, like its first-order counterpart, is very arguably multiply realizable. Rationality of whatever order could certainly be realized by entities whose genomes differed radically from ours, and probably also by entities with no genomes at all.

My account also avoids several pitfalls that any defense of human moral equality should best avoid. Firstly, for reasons we already saw Liao outline in Chapter 2, it is not a species-norm account, since it grounds human moral status in a property that humans have \textit{universally}. Nor is it a speciesist account in any interesting sense. It does not entail that non-human animals are inferior to humans in moral status, nor that we have our moral status simply in virtue of being members of the human species. Rather, our moral status comes from our second-order rationality, which is a feature we certainly could have had (and perhaps do have) in common with certain non-humans. Further, as I have emphasized several times, non-human (and non-rational) animals also create and enjoy goods. Of course, the goods they create and enjoy will

\textsuperscript{114} Boonin 2002, p. 54
\textsuperscript{115} Boonin 2002, p. 41
\textsuperscript{116} Boonin 2002, p. 49
not be rationally specific, but as I have also emphasized, this does not mean that the goods that non-human animals create and enjoy cannot have considerable moral weight.

Nor does the foregoing seem to be a potentiality account. In other words, it does not rely on the fallacious inference from the claim that certain entities potentially have a certain right or rights-making property to the claim that these entities actually have the corresponding right. Rather, it grounds human moral equality in a property that all human organisms actually have, namely second-order rationality. Now, since second-order rationality is a dispositional property (or a set of dispositional properties), it also counts as a certain sort of potentiality. But the exact same is true of first-order rationality, and I don’t think anyone would accuse an account that grounded our right to life in our first-order rationality of being a potentiality account. The point is that second-order rationality is a very different beast from mere potential rationality, particularly in that it does not seem to be had by any entities that are clearly not rightsholders. For example, the Statue of Liberty may be potentially rational, in the thin sense that there is a metaphysically possible world where it (she?) comes alive and begins to reason. But it does not follow from this that the Statue of Liberty has second-order rationality (or any other higher-order form of rationality), since the Statue of Liberty does not have a rational genome, nor anything functionally equivalent to a rational genome.

It could be objected that this only serves to show that my account is not a potentiality account of the crudest sort. More sophisticated potentiality accounts do not rest on the sorts of formal fallacies described above, nor on an equation of potentiality with mere metaphysical possibility. Rather, these sophisticated accounts usually rest on the claim that a standard fetus has the potentiality for personhood in the sense (very roughly) that if all goes well and nothing stops it from developing in the normal way, it will, either certainly or very probably, develop certain properties that will make it intrinsically valuable. However, the problems faced by these more sophisticated sorts of potentiality accounts are not unlike the problems faced by their cruder counterparts, or like other problems that have already been addressed.

For one thing, there are the concrete counterexamples. Of course, most of the counterexamples to the crude accounts will not work as counterexamples to the sophisticated accounts. It is not the case, for example, that the Statue of Liberty will in fact become rational if we don’t prevent it from doing so. But other purported counterexamples are available. McMahan, for example, offers the following case, call it Smart Dogs:

*Suppose that we were to discover that dogs have the intrinsic potential for self-consciousness and rationality but that until now we have failed to recognize this because the potential has never been realized. For suppose that, in order to elicit*
this potential, it is necessary for someone to cultivate and nurture the relevant cognitive capacities through an intensive and highly structured program of “cognitive therapy.” Only through years of patient work, taking virtually every waking moment of the dog’s life from earliest puppyhood on, can the relevant mechanisms latent in the dog’s brain be activated and developed. ... After five or six years, dogs subject to this program develop cognitive capacities comparable to those of a normal four-year-old human child. ... Assuming we were to discover this, ought we to conclude that all dogs have a high moral status—in particular, that all dogs are above the threshold of respect, so that killing a dog is just as wrong as killing a person, if other things are equal? ... I doubt that anyone would draw these conclusions. While we would (or should) accept that respect would be owed to any dog whose potential to become a person had been realized, the knowledge that all dogs had this potential would not require us to reassess our estimation of the actual worth of all those dogs whose potential was never cultivated or never would be cultivated.117

My account seems vulnerable to a counterexample of this sort. After all, couldn’t the Smart Dogs be interpreted as having second-order rational dispositions of a sort? Perhaps they could. The trouble, though, is that McMahan’s own account of our moral status is also vulnerable to a counterexample of this sort. For as noted previously, that account says that we are worthy of respect in virtue of our neurally grounded capacities for rational agency and self-consciousness. But no less than what McMahan calls “intrinsic potential” – and no less than second-order rational dispositions (or first-order rational dispositions, for that matter) – such capacities can go unrealized. Hence the following case, call it Smarter Dogs. Suppose we discover that a certain part of the canine brain (call it the schmerebrum) endows all dogs with all the same capacities for self-consciousness and rational agency that the cerebrum of a normal human person endows on its bearer. However, a quirk of evolution has put the schmerebrum out of commission without damaging it, in a way equivalent to the way the cerebrum is put out of commission when a human person falls into a deep but temporary coma. Luckily, there is a way to activate the schmerebrum of any dog, upon which that dog will reason with the best of them. Unluckily, the activation process is arduous and time-consuming, much as waking up and recovering fully from a long-lasting coma usually is. In fact, activating the schmerebrum of any given Smarter Dog consumes roughly the same amount of time and resources as does the process of “cognitive therapy” needed to realize the rational potential of the Smart Dogs. “Assuming we were to discover this, ought we to conclude that all dogs have a high moral status – in particular, that all dogs are above the threshold of respect, so that killing a dog is just as wrong as killing a person, if other things are equal?” As far as I can tell, McMahan must

117 McMahan 2001, pp. 315 & 316
answer this question in the affirmative (and if he doesn’t, it’s not at all clear how he can account for the high moral status of the temporarily comatose). But the idea that the Smarter Dogs are our moral equals seems neither more nor less credible than the idea that the Smart Dogs are our moral equals.

It might be thought a morally relevant difference that the Smarter Dogs have (presumably) never actually exercised their rational capacities in the past, whereas the typical comatose person has. But in the first place, this seems incompatible with McMahan’s own account, on which capacities are all that counts, if I understand him correctly. It is also deeply implausible just in general, as we can see by tweaking the case of the comatose clone from Chapter 3. Chafing from my last, bad experience with the Duplicator, I make some improvements to it and try cloning myself again. This time, the machine does a better job of copying my kidneys, and my second clone does not come into existence in a permanent coma. But the Duplicator is still a work in progress, so my second clone does come into existence in a coma – it’s just that the coma is temporary. Over the course of the next few months, and with all the difficulty that comes with awakenings from more ordinary comas (or, more to the point, with activating the rational capacities of the Smarter Dogs), my clone will come to be able to immediately exercise all the rational capacities that I now exercise. Thus, my new clone has never exercised its rational capacities, but in due time, and with a great deal of work, it will.

What sort of moral status does this second clone have? It seems clear that it has the exact same moral status I would have had if I were now to fall into a temporary coma exactly like the one my clone is currently in. It is irrelevant that I have previously exercised my rational capacities and the clone has not.

Along similar lines, we have seen Boonin criticize what he calls the species essence argument (which, as also already suggested, can be taken as a sophisticated sort of potentiality argument) on the grounds that it cannot explain why a severely brain-damaged human being has an unrealized capacity for personhood while the spider on Boonin’s window has no such capacity. This could be read as a scope-based objection or as an epistemic objection. On a scope-based reading, the objection is that potentiality accounts do, in fact, commit their adherents to the absurdity that such entities as spiders have an unrealized capacity for personhood, and therefore have the same moral status as you or me. On an epistemic reading, the objection is that such accounts give us no way to tell which entities do and do not have an unrealized capacity for personhood, to the point where they cannot even rule out (though they also don’t rule in) the possibility that the spider on Boonin’s window has such a capacity. My account is not vulnerable to either of these objections. With regard to the scope-based objection,
it is quite clear that the spider on Boonin’s window does not have the genetic basis for rationality, so that the Genetic Criterion does not commit us to the claim that it has the right to life. With regard to the epistemic objection, it is quite clear how we could, in principle, tell whether the spider had a right to life – namely, by discovering that it had the genetic basis for rationality or something functionally equivalent to it.

Now, there may be an epistemic problem as to what sorts of features do and do not count as functionally equivalent to the genetic basis for rationality. But this problem, I suspect, is mainly a result of the fact that I have not here formulated a full account of what it takes for something non-genetic to be functionally equivalent to the genetic basis for rationality. The reason why I have not done this – and the reason why this problem is of little practical importance, both in general and for my account in particular – is that there are no non-controversial examples of entities that are (first-order or second-order) rational, but which are not carbon-based life forms whose rational dispositions are realized by their genomes. Further, my focus here is on one particular life form of this kind, namely human beings. And even if we should be in doubt about how to tell whether angels, aliens, or AIs have something functionally equivalent to the genetic basis for our rationality, there is no doubt that the genetic basis for our rationality is functionally equivalent to itself.

Sophisticated potentiality accounts may also be vulnerable to the nonconception problem. For they suggest that the “wrongmaking” feature of abortion is that it prevents the coming-into-being of a certain intrinsically valuable trait or entity, which is a feature it has in common with acts of nonconception. But for reasons already explained, my own account does not face such a problem.

In short, and again, it is not clear that the above account of how second-order rationality generates the right to life is any more of a potentiality account – even a sophisticated one – than are those accounts which seek to ground the right to life of persons in their first-order rationality. It may be that in taking the potentiality of second-order rationality to produce valuable things as evidence of its intrinsic value or the intrinsic value of its bearer, my account does count as a potentiality account, in a very loose sense of the word. But yet again, if that is the case, any account that makes the same move with respect to first-order rationality will also count as a potentiality account. And it is not at all clear why or how an account’s being a potentiality account in this very loose sense should count against it.
4.222 General morals
So far, I have argued in what might be called an indirect way for the conclusion that second-order rationality generates rights. I have given a particular account of how first-order rationality generates the right to life and then argued that on this account, the “rightsmaking” features of first-order rationality are also features of second-order rationality, so that second-order rationality endows the right to life no less than does first-order rationality. A problem with this strategy, of course, is that it is only as plausible as the particular account it proceeds from. Some other account could well turn out to be more attractive, and this other account, it might be thought, could well turn out to identify the “rightsmaking” feature of first-order rationality with a feature it does not have in common with second-order rationality.

However, the morals from the preceding section arguably generalize to many other theories. Indeed, rather than thinking of the consequentialist account as a serious proposal for an account of how rationality generates rights, we may think of it as a sort of toy theory, chosen mostly because it is relatively simple and straightforward and because its salient features are also present in many other accounts of moral status. The basic point of the previous section – that the rightsmaking features of first-order rationality are also present in second-order rationality – will plausibly generalize to any moral theory which endorses the equal right to life of persons (i.e., what McMahan calls the Equal Wrongness Thesis), and which does so on the grounds that human persons either have certain distinctive and intrinsically valuable capacities or are themselves intrinsically valuable in virtue of these capacities. For again, these capacities will plausibly be rational capacities of some sort, and again, the reason why we find it plausible to say that the capacities are intrinsically valuable (or bestow intrinsic value on their bearers) is surely because of the many great goods they bring forth. But again, second-order rationality brings forth at least an equal number great goods, and therefore has at least as much of a claim to intrinsic goodness.

It might now be objected that we feel tempted to ascribe intrinsic goodness to the distinctive capacities of persons, not only because these capacities bring forth great goods, but also because such an ascription of intrinsic goodness can help justify our pre-existing intuitive attraction to the moral equality of human persons (and persons in general). But as I have already touched on a few times (first in Chapter 1, then in 2.5), there are good reasons to think that there is a comparably strong and widespread intuition in favor of the moral equality of human beings.

The critic may respond to this claim in a few different ways, most of which we have already encountered. Firstly, he may simply deny that the intuition is as widespread as I claim. Perhaps many of the people who claim to believe in (or at least to be intuitively drawn to) the
moral equality of human beings really mean that they believe in the moral equality of human persons. I dealt with this line of argument in Chapter 1. Alternatively, the critic may claim that our intuitive judgments about particular cases show that few people really believe in (or have intuitions that support) human moral equality, even if they claim to do so. Here, early embryo death and the Embryo Rescue Case are favorites. But as I argued earlier in this chapter, our intuitive reactions to these cases may well be perfectly compatible with human moral equality, provided human moral equality is defined as applying only to the wrongness of killing, not to the badness of death. Similarly, McMahan argues that the fact that most pro-lifers are “pro-life with exceptions” – that they are willing to at least tolerate abortion in cases of rape, incest, risk to the life of the mother, and so on – shows that they do not really believe that abortion is morally equivalent to murder. 118 Two things can be said to this. Firstly, McMahan arguably underestimates the widespreadness of the most conservative, least exception-tolerant versions of the pro-life point of view. Such a version of the pro-life point of view is, e.g., the official view of what is almost certainly the largest and most influential pro-life institution in the world, namely the Catholic Church. 119 Secondly, as we will see in the next chapter, it is far from obvious that human moral equality is incompatible with the claim that abortion is permissible under some (or even most) circumstances.

Further, even if human moral equality is inconsistent with some of our other intuitions or beliefs, this does not show that we don’t really believe in human moral equality after all. For one thing, the fact that our intuitions about particular cases sometimes contradict our general theories does not, in general, show that we don’t really believe in those general theories. From the Banach-Tarski paradox to the double-slit experiments, the hard sciences contain plenty of results that almost everyone finds deeply counterintuitive. But I doubt that anyone would take this to show that almost no-one really accepts any of those results. Secondly, even if many adherents of human moral equality really do accept (rather than merely being intuitively drawn to) concrete judgments that really are incompatible with human moral equality, this does not show that they don’t really believe in human moral equality. By parity of reasoning, it might just as well be held to show that they don’t really accept those of their concrete moral beliefs that contradict human moral equality.

The critic might also concede that we are intuitively drawn to human moral equality in the same way we are intuitively drawn to the Equal Wrongness Thesis, and that this is prima facie evidence for human moral equality. However, he might add, this prima facie evidence is

118 McMahan 2001, p. 271
119 See e.g. Catechism of the Catholic Church §§2270-2279
overruled by the fact that our attraction to human moral equality has arational sources like speciesism. But for reasons also covered in Chapter 1, I don’t think accusations of speciesism can prove very much here. And even if they can, they must first be substantiated. Discounting the intuition that membership in the human species has moral relevance by simply asserting that this intuition has its source in “speciesism” is no more convincing than, e.g., discounting the intuition that sentience has moral relevance by simply asserting that this intuition has its source in “sentientism,” the purportedly irrational privileging of sentient entities over other entities.

In general, any account that grounds our intrinsic value in our second-order rationality will also have a much easier time with the equality problem. If persons have their intrinsic value in virtue of their first-order rationality, it is hard to avoid the conclusion that their intrinsic value – and therefore their moral status – varies depending on their degree of first-order rationality. As we have seen, McMahan deals with this problem by grounding the moral status of persons in the imperative to respect them for their capacities, an imperative which, as he argues, generates the same basic obligations towards persons in which those capacities are present to different degrees. But by doing so, he also complicates his account considerably, grounding the general prohibition on killing persons and the general prohibition on killing non-persons in two completely distinct moral imperatives (the “requirement of respect” and the prima facie prohibition on thwarting time-relative interests, respectively). If the moral status of persons is grounded in their second-order rationality, it becomes easier to hold the Equal Wrongness Thesis in conjunction with a more parsimonious and unified ethics of killing – one that accounts for the wrongness of all killing on the basis of a single imperative. The consequentialist account, for example, can be described as grounding a person’s right to life in a prohibition on the wanton destruction of goods. As we have seen, this prohibition takes a particular – and particularly stringent – form when those goods are intrinsic, but it still applies, in a somewhat different and less stringent form, when the goods are instrumental. It is therefore plausible that this imperative can also account, e.g., for our weaker obligation not to kill non-rational, non-human animals. There, the obligation is presumably weaker because the beings themselves are presumably only instrumentally good – good, not in and of themselves, but only because of the goods they will create and enjoy.

If we do not want to give up our conviction that persons – first-order rational beings – are intrinsically valuable and morally equal, it is hard to avoid the conclusion that second-order

120 McMahan 2001, ch. 3, sect. 3.3
rational beings have the same sort of status, no matter what sort of theory we use to account for this conviction. And once this status has been granted to second-order rational beings, it is even harder to deny that all human beings are intrinsically good and morally equal.
In closing, I want to consider what sorts of conclusions we can – and can’t – draw from the preceding chapters. More specifically, I want to consider what implications human moral equality does – and doesn’t – have for a variety of concrete moral questions.

Human moral equality may be thought, not without justification, to have huge implications for applied ethics. The focus of this thesis has been abortion, but the implications obviously extend far beyond that, touching on all questions that have to do with the taking of human life. Most obviously, it touches on the ethics of infanticide, in vitro fertilization, and stem cell research. A bit farther afield, it might also be thought to have implications for how we should think about euthanasia. And beyond bioethics, it might be thought to have implications for the ethics of war and of self-defense.

If what was said in the preceding chapters is right, killing a fetus under a given set of circumstances is only permissible if it would be permissible to kill you or me under equivalent circumstances. But there are some cases, notably self-defense, in which it very arguably is permissible to kill you or me. However, while there is an almost universal consensus that it is sometimes permissible to kill in self-defense, the particulars are much more controversial. At this point, it will be useful to introduce the distinction between culpable aggressors, innocent aggressors, and innocent threats. A culpable aggressor is an agent who intentionally and culpably puts you in danger, such as a killer rushing towards you with a knife, fully intending to stab you to death. An innocent aggressor is an agent who intentionally puts you in harm’s way through no fault of his own, such as a person who intends to stab you to death, but who has only formed this intention because he has been drugged or brainwashed. Finally, an innocent threat is an agent who puts you in danger unintentionally and non-culpably, such as a driver who has lost control of a trolley that is now careening towards you. (There might also be a fourth category: culpable threats, who put you in danger in a blameworthy but unintentional way. A careless driver, for instance, may not intend to harm you, but will still be at fault if he does.)

A particularly difficult – and particularly relevant – question is under which circumstances, if any, it is permissible to kill innocent threats in self-defense. The question is relevant because insofar as killing a fetus could ever count as self-defense, it could only count
as self-defense against an innocent threat, since no fetus is yet in a position to form intentions or incur moral blame. The most promising candidate for a situation in which killing innocent threats is permissible, it seems, is the case in which one or more lives are at stake. But even here, there is no consensus. Michael Otsuka, for instance, has argued that killing innocent threats in self-defense is morally equivalent to killing innocent bystanders, and is therefore never permissible. When it comes to abortion, the analogous case here is the case in which the life of the mother is at stake. Such cases, however, are quite rare today, and most of them do not involve the sort of “direct” abortion that is the main focus of this thesis, but the sort of “indirect” abortion I mentioned in Chapter 1 (where I also questioned whether indirect abortions should really be counted as abortions at all). A trickier case is the one in which the autonomy or well-being of the mother is at stake. Is it, at least under certain circumstances, permissible to cause the death of an innocent threat to one’s own autonomy or well-being? Judith Jarvis Thomson’s well-known “Violinist” case can easily be read as an argument to that effect.

The point I want to make by raising these questions is not to answer them, but precisely to show that a mere assertion of human moral equality can’t answer them either. Since it is unclear when, if ever, you or I pose innocent threats of such a severity that we may be killed in self-defense, it is also unclear when, if ever, a fetus poses an innocent threat of such a severity that it may be killed. Perhaps never, as Otsuka suggests, or perhaps quite frequently, as Thomson suggests.

And if human moral equality does not deliver a clear verdict on the moral status of any given abortion, it says even less about what sort of legal status any given abortion should have. Intuitively, the grave immorality of an action is prima facie evidence that that action should be illegal. For instance, the grave immorality of rape seems to be at least part of the reason why rape is (and should be) illegal. But not all acts that are wrong – and perhaps not even all acts that are gravely wrong – should be illegal. For instance, most of us would agree that adultery and lying are usually wrong, and sometimes gravely so. But they aren’t – and probably shouldn’t be – illegal (except for a few special instances in the case of lying, such as perjury and slander). When considering the question what the law should say about abortion, therefore, there is not only the unresolved question of which particular acts of abortion are wrong – there is also the question of whether they should be illegal even if they are.

When it comes to euthanasia, these troubles compound. In the case of voluntary euthanasia, the central question is not normally whether the patient is on a moral par with you

---

121 Otsuka 1994
122 Thomson 1971, pp. 48 & 49
or me, since almost everyone would presumably agree that anyone who has the whereabouts to consent to euthanasia is, by dint of that, a full member of the human moral community. (The big exception to this are the cases in which consent is given through an advance directive, the ethics of advance directives being a can of worms unto itself.) Rather, as has been noted by David Oderberg, among many others, the central question is whether it is possible to renounce one’s own right to life, or, perhaps equivalently, whether (or when) the purported right to autonomous decision-making can trump the right to life.\textsuperscript{123} Again, the foregoing offers no immediate answers to these questions.

When it comes to involuntary euthanasia\textsuperscript{124}, there is the case in which the patient cannot consent at all, and there is the case in which the patient intentionally withholds consent. The latter sort of case (where the vast majority of philosophers would presumably hold that euthanasia is never or almost never permissible) once again raises questions about autonomy, while the former sort is more closely analogous to abortion than any other type of euthanasia, since both involve a party that is unable to give or intentionally withhold consent. Once again, it also raises questions about whether (or when) a concern for preventing suffering can trump the right to life. Finally, the ethics of killing in war once again raises questions regarding self-defense (here mostly against innocent or culpable aggressors) which, yet again, are not answered merely by an assertion of human moral equality. For in the case of killing in war, as in most cases of voluntary euthanasia, the central question is not whether the other party is your moral equal, but under what circumstances you may kill him, given that he is your moral equal.

In sum, human moral equality has fewer implications for applied ethics than we might think. But while it does not offer immediate answers to any of the questions raised above, it does suggest a methodology for how we should approach them. If all human beings are on a moral par, then showing that it is right or wrong, permissible or impermissible, to kill a certain human being – any human being – under a certain set of circumstances suffices to show that it is equally right or wrong to kill any other human being under any morally equivalent circumstances. Again, this does not settle everything, but it does allow us to argue in new ways.

For instance, if combined with an assertion of human moral equality, Otsuka’s argument becomes an argument for the moral impermissibility of all forms of abortion. Thus, the philosophical case against abortion, in its standard form, is a two-step process: First one argues for human moral equality, then one draws out its practical implications. I hope this work has shed some new light on how we might take the first step.

\textsuperscript{123} Oderberg 2000, p. 54 \textit{et passim}

\textsuperscript{124} Also sometimes called “medical murder” – see e.g. Materstvedt & Magelssen 2016.
Bibliography


