THE LIMITS OF KINDNESS

Caspar John Hare

June 2010

CONTENTS

PART II

Introduction

1. Normative Ethics

Let's begin with some basics. This book is about *normative ethics*. What is that?

There's a natural distinction between claims like these:

> "The square root of two is irrational."
>
> "The economy is in decline."
>
> "Proxima Centauri is our Sun's closest big neighbor."
>
> "That really, really hurt."

and claims like these:

> "You ought to be nice to your grandmother."
>
> "Paris is beautiful."
>
> "Lying to little children to serve your own ends is wrong."
>
> "Television is bad for the soul."

The first sort of claim pertains only to the way things are, the second sort of claim pertains, at least in part, to the way things ought to be. In philosophy-speak: the first sort of claim is *descriptive*, the second *normative*.[1]

Some normative claims overtly concern morality. For example:

> "Cyril is a dishonorable bounder."
>
> "The administration is irredeemably corrupt."
>
> "Harry wronged you by taking your hat."

Others do not. For example:

[1] I should note (the first of many qualifications) that though the basic distinction is natural and easy to grasp, drawing a precise boundary between normative and descriptive claims is an exceptionally interesting and difficult job. We need not worry about that here.

"Eating your hair is a bad way to conserve energy."

"On clay, Roger Federer ought to adjust his backhand to compensate for the higher bounce of the ball."

Moralists are in the business of making normative claims of the first sort, *moral-normative* claims.[2]

Moral theorists are moralists with a systematic bent. They are unsatisfied with claims about particular things, like Cyril, or Harry's action-of-taking-your-hat. They have grander ambitions. They want to make very general claims about kinds of things:

"All people of kind __ are dishonorable bounders."

"All actions of kind __ wrong you."

"All punishments of kind __ are unjust."

Their idea is for these general claims to serve as the basis of a theory that will entail claims about particular things.

Normative ethicists are moral theorists with a special interest in individual people – in how they are, and in what they do. Normative ethicists are not directly concerned with when institutional structures are fair, when policies are discriminatory, when aspirations are noble, when customs are deviant and contrary to nature… and so on and so forth. They want a theory of the conditions under which people are good or bad, honorable or dishonorable, admirable or loathsome…etc. And they want a theory of the conditions under which people's actions are right or wrong, permissible or impermissible, praiseworthy or blameworthy…etc. They are, not to mince words, high-minded busy-

[2] Again, I should note that, though the distinction between normative claims that have a moral flavor and normative claims that do not is natural enough, there is no uncontroversial way of drawing a precise boundary between the realm of the moral and the realm of the normative-non-moral. Again, we need not worry about that here.

bodies. Their goal is to tell us how we ought, morally, to be and what we ought, morally, to do.

2. Why Normative Ethics is Hard

Suppose we try to join the high-minded busy-bodies, and set about building a theory of normative ethics. How hard can this be? Quite hard, it turns out. Let’s start by addressing the morality of action. Our first job is to find some general claims about the moral status of actions to build our theory around. But this is not a simple matter. Trusting our instincts, looking for general claims that have a *plausible ring* to them, will not help, because many general claims about the moral status of actions that have plausible ring to them turn to be inconsistent with one another. Normative ethicists have proven to be very adept at rooting out and exposing such inconsistencies. I will discuss here, and over the coming chapters, three celebrated examples of pairs of inconsistent, right-seeming normative claims. Here’s one:

> *Betterness*
> You do wrong only if things overall would have been better if you had acted in some other way.

> *Violation*
> It is wrong to inflict grievous physical harm on an unconsenting, innocent person who poses no threat to anybody, unless you can secure massively disproportionate benefits by doing so.

Betterness seems right. If I accuse you of doing something wrong, then I can't very well say "…though, of course, things would have been no better if you had done anything else." By accusing you I seem to be committing myself to the idea that things would have been better if you had done something else. And *Violation* sounds right too. Of course it is not okay to hurt unconsenting, innocent, unthreatening people.

But the claims are inconsistent. Consider:

> <u>Human Fuel</u>
> While piloting a steel, steam-engined boat across the bleak South Seas, you receive a distress call from Amy, who has been left to die of thirst on a nearby, bare island by cruel pirates. Knowing that you are the only boat in the area, you pick her up, and then receive another distress call from Brian and Celia, left to die of thirst on a not-so-nearby, bare island by more cruel pirates. You do not have enough coal to get to them, and no part of your steel boat will serve as fuel… but Amy is both fat and dehydrated… Knock her on the head, shove her in the furnace and you will make it over there. And nobody but you will ever know.

If *Violation* is true then it would be wrong to use Amy-power to save Brian and Celia. But if *Betterness* is true then, seemingly, it would not be wrong. If you knock Amy on the head, then one person will have been killed in a quick, relatively painless way. If you don't, then two people will have been killed in a lingering, relatively painful way. It is no better that two people be killed in a lingering, relatively painful way than that one person be killed in a quick, relatively painless way.[3]

Here's another celebrated example of inconsistent, right seeming claims:

> *Harm*
> An action is wrong only if it harms something with morally significant interests.
>
> *Optimizing the Health of Your Child*
> Whenever you have made it your business to conceive and bear a child, it is wrong to choose that your child be unhealthy, rather than healthy, unless you have strong reasons to do so.

Harm seems right. If I argue that you have done the wrong thing, and you reply "Where was the harm in that?" then I must answer your question or lose the argument. And *Optimizing the Health of Your Child* seems right too. Of course parents ought have some concern for the health of their children. Of course they ought to act on that concern, absent strong reasons to the contrary.

But the claims are inconsistent. Consider:

> Not Postponing Conception
> Mary is recovering from German measles. Her doctor recommends that she postpone her efforts to conceive a child for a couple of months. If she conceives a child in the next couple of months then the child will, most likely,

[3] This problem goes by various names. It is sometimes referred to as 'Scheffler's Paradox', after an influential treatment of it in Samuel Scheffler, *The Rejection of Consequentialism*, Oxford University Press 1982, sometimes referred to as 'Foot's Problem' after an influential treatment of it in Phillippa Foot, "Utilitarianism and the Virtues", *Mind* XCIV(374): 196-209, 1985. But philosophers were aware of some form or other of the problem well before that. See, for example, HJ McCloskey "A Note on Utilitarian Punishment", *Mind* LXXII (288): 599, 1963. I will discuss the problem in more detail in Chapter Seven.

> have significant health problems. Mary has no strong reasons to conceive a child immediately, but she does have a mild preference for *getting on with it*. She gets on with it. Nine and a half months later baby Mariette is born, with significant health problems. This is not a disaster – Mary is a woman of means, so Mariette's health problems do not impose a burden on wider society, and, on balance, Mariette has a rewarding life. But it is not great either – Mariette's health problems are a chronic source of anxiety, pain and frustration to her.

If *Optimizing the Health of Your Child* is true then it would seem that Mary did wrong by conceiving her child immediately. But if *Harm* is true then it would seem that she did no wrong. If Mary had not conceived immediately then she would, most likely, have conceived a baby some time later, as genetically different from Mariette as typical non-twin siblings are genetically different. On any plausible view of essence, that child would not have been Mariette. On any plausible view of harm, you do not harm somebody whose life is, on balance, rewarding, by making it the case that he or she exists rather than not.[4] And if Mary didn't harm Mariette, then who or what did she harm?

Here's the third celebrated example of inconsistent, right-seeming claims:

> *Sacrifice*
>
> Whenever you are in a position to save the life of a child at a relatively small cost to yourself and no cost to others, you are morally obliged to do so.

[4] This has been known as the 'Non-Identity Problem' since Derek Parfit, "Rights, Interests and Possible People," in *Moral Problems in Medicine*, Samuel Gorowitz ed., Prentice Hall 1976, 369-75, and Derek Parfit, *Reasons and Persons*, Oxford University Press 1983, chapter 16. I will discuss it in detail in Chapter Five.

> *Sackcloth and Ashes*
> People living in affluent societies are not morally obliged to give away almost everything they have, for the sake of people in distant, poor societies.

Sacrifice sounds right. What kind of a monster would let a child die for the sake of his cufflinks? *Sackcloth and Ashes* sounds right too. Surely we do not have to give away almost everything we have! Can't we hold onto at least a few small indulgences – a car when we could take the bus, freshly squeezed orange juice when we could get squash?

But the claims are inconsistent. Consider:

> <u>Oxfam</u>
> The charity Oxfam is soliciting donations for a program that will vaccinate impoverished children against disease. Craftily, the administrators of the program have arranged their finances in such a way that the marginal benefits of further donations are clear: For every $100 you give, around ten more children will be vaccinated. For every ten children vaccinated, around one of them will live through an epidemic of disease that would otherwise have killed them.

If *Sacrifice* is right then it would appear that you are obliged to give your first $100 to the program – by doing so you will save the life of a child at a small cost to yourself. And you are obliged to give your second $100 to the program for the same reason, and your third $100… right down to the point where the marginal cost to you of giving away a further $100 is relatively large. But by that point you will have given away almost everything you have. So if *Sacrifice* is right then you are obliged to give away almost everything you have in this case, which contradicts *Sackcloth and Ashes*.[5]

3. Reflective Equilibrium

The moral to draw from these examples is that not all general normative claims that seem right, at first blush, are right. How, then, are we to decide which to accept? Many normative ethicists would give this advice: "Building a theory of normative ethics is about arriving at *reflective equilibrium*. Start by taking all the normative claims that sound right to you, whether they be very general claims about the nature of the good and so forth, or very particular claims about particular cases. Then test them against each other. Look for inconsistencies. If you find inconsistent claims then discard one or the other. When you make decisions about which claims to discard, do not dogmatically favor the more general over the more specific, or vice versa – consider the strength of your attachment to the respective claims, consider the unity, simplicity and explanatory power of the emerging theory. If you can find more general claims that will entail and explain disparate, surviving, more specific claims then adopt them. Repeat this process again and again, until you are left with a simple, consistent theory."[6]

I have never found this advice very helpful. I find it easy enough to identify claims that sound right to me, to test such claims against each other, and to find inconsistencies. But when it comes time to "discard one claim or the other" I often have

[5] The canonic presentation of this problem is in Peter Singer, "Famine, Affluence and Morality", *Philosophy and Public Affairs* I, no. 1: 229-243, 1972. I will discuss it in detail in Chapters Twelve and Thirteen.

[6] The name 'reflective equilibrium' was coined by John Rawls in his *Theory of Justice*, Harvard University Press 1971. Rawls traced the idea back to Nelson Goodman's discussion of methods for justifying rules of inductive logic in his *Fact, Fiction and Forecast*, Harvard University Press 1955. Nowadays many more normative ethicists use some form of the method than talk about it explicitly.

no idea which way to go. Take the Human Fuel case, for example. *Betterness* and *Violation* both seem right to me. The case shows they are inconsistent. How am I to proceed?

Attending, first, to the unity, simplicity and explanatory power of the emerging theory, I find that *Betterness* comes out ahead of *Violation*. Precisely explicating the virtues of unity, simplicity and explanatory power in a theory is a notoriously difficult problem. But any interesting explication of these virtues will cast act consequentialism, the theory that makes *Betterness* its centerpiece, in a flattering light. Act consequentialism can be stated in a sentence:

> *Act Consequentialism*
> An act is wrong if and only if some alternative to it has a better outcome.

When sharpened up a bit (we need to explain what *alternatives* and *outcomes* are) and combined with an appropriately precise axiology (a theory of what makes one outcome better than another) it nails down the moral status of all acts, performed by anyone, at any time, in any place. *Violation*, on the other hand, is a very local principle. It tells us about the moral status of certain acts that inflict grievous physical harm on unconsenting, innocent people, and nothing else. Some theorists who adopt *Violation* think of it as just one of a patchwork of independent moral principles. These principles together make up a moral theory that is far more complex and disunified than act consequentialism. Others think that *Violation* can be derived from principles that are (at least close to) as simple and unified as act consequentialism – variants of rule consequentialism, contractualism and Kantian rationalism. But I, for one, find these derivations less than convincing.

But unity, simplicity and explanatory power are not the be-all and end-all of everything in normative ethics. If they were then moral nihilism (for these purposes: the view that all actions are morally neutral) would be the best theory of all. The method of reflective equilibrium would have me weigh, also, the strength of my attachment to *Betterness* and *Violation*. How strongly am I attached to each?

Well, the first thing I must acknowledge is that, for me, *Violation* has a lot of pull. I expect it has a lot of pull for you too. Imagine ignoring it, and behaving as the act consequentialist would have you behave in the <u>Human Fuel</u> case. Imagine picking Amy up from the first island. Imagine her relief at seeing you. Imagine learning of the second island. Imagine shielding your calculations from Amy (better that she not know about them). Imagine creeping up behind her and smashing her head with a heavy spanner. Imagine smashing her head again and again to be sure that she is dead (very important, given what is about to happen). Imagine dragging her body across the deck and cramming it into the boiler. Imagine the temporary loss of power to your boat as the liquids in her body evaporate away. Imagine pungent new smells emanating from the boiler. Imagine the surge of power as her flesh begins to burn. Imagine arriving at the second island. Imagine Brian and Celia's relief at seeing you. Imagine steaming away with that secret smoldering in the middle of your boat… I expect that, the more carefully you fill in the details of this story in your imagination, the more disgusting you will find it. And I expect that your disgust will have a particular flavor. It will not be the sort of disgust that comes with imagining doing something morally admirable but gross – like diving into a cesspit to save the life of a toddler, or spooning the brains of an injured soldier back into his skull. It will be *moral* disgust. Your imaginary actions will seem, in visceral way, wrong.

Fine. Now, what should you and I do with our visceral judgments?

On one way of thinking, we should discard them. All that matters in the Human Fuel case is that if you do the one thing then two people will be killed, if you do the other thing then one person will be killed. Our visceral judgments about these sorts of cases are mistaken.

Advocates of this way of thinking owe us a debunking explanation of how we came to be mistaken. And they have many to hand. For example, they can say: "Doubtless the idea of knocking Amy on the head and throwing her in the boiler seems horrific to you, but this is because the immediate consequences of doing so are horrific, and we all have a tendency to focus on the immediate consequences of what we do. This tendency serves us well, most of the time, because most of the time we have far greater control over the immediate consequences of our actions than the distant consequences of our actions. But sometimes it leads us astray. It is leading you astray here. Immediate and distant consequences matter equally much."

On another way of thinking, we should preserve the judgments. Obviously we do not want to demand of our incipient moral theory that it represent all of our visceral judgments about right and wrong as true. We want to leave space for correction. But this visceral judgment deserves the status of a Moorean conviction. Its truth is radiant. Any claim with which it is inconsistent should be rejected.

Advocates of this way of thinking, too, owe us a debunking explanation of why we might be inclined to think otherwise, to think that killing Amy is in fact the thing to do. And they, too, have many to hand. For example, they can say: “Look, what matters in this case is that Amy, Brian and Celia each have a right not to be murdered. Philosophers may appreciate that rights matter, and infer that the appropriate thing to do in this case is ensure that as few people as possible have their rights violated. Malaria is bad, so we should minimize malaria. Violation of rights is bad, so we should minimize violation of rights. But this is a subtle mistake. Rights are not like malaria. The appropriate way to respond to the fact that rights matter is not by *minimizing the violation of rights*, but rather by *not violating rights*.”

For each of the two points of view, I think I see it quite clearly, and I can work myself into a frame of mind where it seems like the right one. But which is the right one? I find it very hard to judge. I know, of course, what I would do if I found myself in the Human Fuel case. I would leave Amy alone. I doubt that I would even allow myself to acknowledge that throwing her into the boiler was an option. It would be one of those thoughts that flits around behind the camera of my mind, too shameful and disturbing to be released into view. But would that reflect some deep practical wisdom on my part, or would it reflect moral cowardice? I find it very hard to judge.

Is this a weakness on my part? Maybe so. Certainly, many of my friends and colleagues have no difficulty with the question. They whole-heartedly embrace one of the views and renounce the other. They think it obvious that we should all do the same. But I take some consolation in the fact that they do not all go the same way. Roughly half of them embrace some form of act consequentialism. Roughly half of them embrace some form of deontology.

In my experience, feelings run pretty high in this domain. Many act consequentialists think of deontologists as wooly-headed weaklings, as people who would rather the obscure the landscape of normative ethics with enigmatic musings about Kant than face up to the fact that, sometimes, contrary to 'intuition', they must dirty their hands. Many deontologists think of act consequentialists as accountants-gone-wild, as people who have entirely lost touch with their moral sense, and come to care only about the books. (Six years ago, when I first arrived at MIT, I taught a graduate seminar on normative ethics. Judy Thomson, who I *enormously* respect and admire, attended the seminar, and came to suspect that I had an unseemly attraction to act consequentialism. One day, after class, she pressed me on cases like the <u>Human Fuel</u> case. When I confessed to feeling the pull of the act consequentialist way of thinking about such cases she threw her hands in the air, gave me a special sort of look, proclaimed my condition 'terribly sad', and walked away. The look made a big impression on me. It wasn't the fond, indulgent sort of look that you give your idealistic, wayward young nephew when he tells you that he is thinking of the joining the socialist party. It was the sort of look that you give your nephew when he tells you that he is thinking of joining the Klan.)

It is a curious phenomenon. These are very intelligent people. They are fully aware of the arguments on both sides of the issue. Ask the act consequentialists to make a case for deontology and they will say everything the deontologists say, with all the same passion, emphasis and conviction. Ask the deontologists to make a case for consequentialism and they will say everything the act consequentialists say, with all the same passion, emphasis and conviction. It is not as if the members of either group are *missing something* – not as if there is a decisive consideration that has passed beneath their attention. And they all know this. Yet some psychological mechanism causes roughly half of them to break one way and roughly half of them to break the other way.[7]

4. A Foundational Approach

If you, like me, don't break either way, if you share my ambivalence about these sorts of questions, then you might be interested in different methods. You might hope to set the 'prima facie plausible' normative claims aside and base your theory on firmer foundations. You might hope to fix on some abstract normative principles that are non-negotiably true. You might hope to take these principles as axioms and derive a normative theory from them. You might hope that the resulting theory would be as solid and pure as Peano Arithmetic or Zermelo-Fraenkel Set Theory.

What sort of 'non-negotiable' principles would do the job? An obvious place to start is with principles of practical rationality.

This is not an original idea. For almost as long as there have been philosophers, there have been philosophers trying to base ethics on rationality, broadly understood.

[7] An untutored outsider might feel that this reflects badly on what normative ethicists are doing. The three problems that I have drawn attention to here are among the most basic problems in normative ethics. They are routinely taught in introductory classes on the subject. The fact that there is no consensus on how to resolve them, and a firm consensus that our present methods of enquiry will not yield a consensus on how to resolve them, might suggest, to the untutored outsider, a weakness in those methods.
Philosophers tend to be wary of indulging such feelings (with some reason – in philosophy, deadlock is everywhere). A good part of our early education consists in stripping them away. David Lewis used to say:

> "Debates may deadlock. But it does not follow that they are not worth pursuing, or that there is no fact of the matter about who is right."

Back when I was young and impressionable, this struck me as deep wisdom, but now I am not so sure that the moral we were expected to draw from it (that, in philosophy, there's a kind of virtue in sticking to your guns in the face of implacable opposition) is a good one. Often, when philosophical debates deadlock, it seems to me irresponsible to come down firmly on one side or other. Often it seems to me that coming down firmly on one side involves willfully ignoring the powerful considerations that move your opponents.

Plato at least experimented with the thought that morality was about enlightened self-interest. The ancient stoics held as their ideal a life in accordance with reason, free of internal conflict. Kant claimed that moral requirements derived from his categorical imperative, and that his categorical imperative was a requirement of rationality. Generation after generation of neo-Kantians have argued that rationality requires us to be impartial in some way, and this requirement is the source of moral obligation. Most, following the letter of Kant, have argued that the resulting obligations are deontological. Some, like my unrelated namesake R.M. Hare, have argued that the resulting obligations are consequentialist.

Airily sweeping my hand across thousands of years of intellectual history, I say now that none of these projects does the job we want done. There's a reason why. To derive substantive moral principles from principles of practical rationality you need some very rich principles of practical rationality. But the principles of practical rationality that have the non-negotiable flavor that would make them suitable as foundations for a pure and solid normative theory are relatively impoverished.

When we think of non-negotiable principles of practical rationality the examples that spring to mind have to do with internal coherence: if you are rational then your desires are internally coherent (roughly: you do not want one thing and at the same time want another – I will unpack this precisely later) and your behavior is coherent with your desires (roughly: you do not want one thing and do another – again, I will unpack this precisely later). But it is easy to be internally coherent. Internally coherent people may be good people, who want good things and behave in good ways, or evil people, who want evil things and behave in evil ways, or just bizarre people, who want bizarre things and behave in bizarre ways. The history of philosophy is full of examples that illustrate this point. The most famous are due to David Hume. If we take ‘reason’ to place nothing more than coherency constraints upon us[8] then we must agree with him that

> ’Tis not contrary to reason to prefer the destruction of the world to the scratching of my finger. ’Tis not contrary to reason for me to choose my own total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. ’Tis as little contrary to reason to prefer even my own acknowledged lesser good to my greater, and have a more ardent affection for the former than the latter.[9]

It does not follow from my being internally coherent that I will behave in one way or another when placed in a morally portentous situation. I may be an internally coherent saint or an internally coherent psychopath. If we are to derive a substantive, interesting theory of normative ethics from rock-solid axioms then we will need to supplement principles of internal coherence with something richer.

What might do the trick? One strategy might be to add axioms concerning about the content of rational desires. Some philosophers have claimed that rationality places significant constraints on what we can desire. Derek Parfit, for example, has said[10] we can have reasons for desiring various things, and that we are more or less rational to the

[8] Did Hume himself think that reason placed coherency constraints upon us? He did write that someone who fails to choose the proper means to his end makes an ‘error’. This might appear to suggest that he thought of the instrumental principle (which says, roughly: *take the known means to your desired ends* – a sort of coherency constraint) as a requirement of rationality. But there are grounds for thinking that the appearance is misleading. See Christine Korsgaard “The Normativity of Instrumental Reason”, in her *The Constitution of Agency*, Oxford University Press 2008.

[9] From Book 2, Part 3, section 3 of *A Treatise of Human Nature*, David Hume 1740. The examples are under-described in odd ways. Does the first fellow prefer the destruction of his finger to the scratching of his finger? If not, is his finger not part of the world? – But the general moral is clear.

[10] In the first part of his as-yet-unpublished, but widely read manuscript *On What Matters*.

extent that our desires are more or less strongly supported by reason. Someone whose desires are wildly out of step with reason (his canonic example is a man who has a healthy desire that he not suffer pain, except on Tuesdays – a man who would rather that he suffer any amount on pain on a Tuesday than any amount of pain on any other day of the week) is, by any standard, irrational. So we might add an axiom that says that, if you are rational, your desires are not wildly out of step with reason.

Maybe this claim has the non-negotiable flavor that we want of the axioms of our theory. But to derive any interesting conclusions about what rational people do in morally portentous situations, we will need to supplement it further, with some specific claims about which kinds of desires are, and which kinds of desires are not, out of step with reason. (Jane kills one person to prevent two from being killed, because she cares more about *there being less killing* than about *her not killing*. John refuses to kill one person to prevent two from being killed, because he cares more about *his not killing* than about *there being less killing*. If we are to derive any interesting conclusions about their respective actions, we have to say that one of John or Jane has desires out of step with reason.) These sorts of claims certainly will not have the non-negotiable flavor that we want of the axioms of our theory.

In light of considerations like this, many contemporary philosophers think that the project of deriving a substantial theory of normative ethics from self-obvious axioms is hopeless. Normative ethics is not about cranking out results. It is an altogether subtler business. To do normative ethics well you must do reflective equilibrium well. And to do reflective equilibrium well you must have tact, experience, and sensitivity to the delicate nuances of moral life. This is why pimply adolescents are very good at number theory, but very bad at normative ethics.

I am no longer a pimply adolescent. But I find that the little tact, experience and sensitivity to the delicate nuances of moral life that I have gained since my pimply adolescence leave me ill-equipped to settle any interesting questions in normative ethics by balancing my intuitions about principles against my intuitions about cases. So, in this book, I want to pursue the foundational project further. I think that we can make significant progress in normative ethics by supplementing some very minimal assumptions about rationality with some very minimal assumptions about moral decency.

5. Moving Forward

In Chapter One I spell out my first assumption about moral decency. It amounts to this: If you are decent then you are at least minimally benevolent towards other people. *When absolutely all other things are equal*, at least, you would rather that other people be better off rather than worse off.

In Chapters Two and Three I spell out my first assumption about rationality: If you are practically rational then, when you don't know what will happen if you do one thing or another, your decisions are guided by the prospects associated with the acts open to you. This is a very intuitive idea, though tricky to put in a precise way, as we will see.

In Chapters Four to Six I put these assumptions to work in normative ethics. Chapter Four is about saving people from harm. There has been a great (albeit, to outsiders, rather mysterious) controversy in the normative ethics literature over whether and (if so) why, given the choice, we are obliged to save more people, rather than fewer people, from similar harms, and to save a multitude of people from a small harm rather than one person from a large harm. But it follows from the minimal assumptions that, in some circumstances at least (circumstances in which you don't know who you will save by doing one thing or another), if you are decent and rational then you will save more people, rather than fewer, from similar harms, and you will save the multitude from the small harm rather than the one from the large harm.

Chapter Five is about cases like Not Postponing Conception, cases that raise 'the non-identity problem'. It follows from our assumptions that, if Mary were decent and rational, then she would not have conceived unhealthy Mariette. It does not follow that she has done something *wrong*. But I will suggest that the two notions *being such that somebody decent and rational would not do it*, and *being wrong*, are closely connected.

Chapter Six is about cases like Human Fuel, cases that raise the problem of whether it is okay to kill-to-prevent-two-killings. I will argue that, by framing the issue in terms of what a minimally decent and rational person will do in these cases, we add significant force to a traditional objection to the deontologist's treatment of them: the so-called 'dirty hands' objection to deontology.

Thus far we have only made progress in cases in which you do not know who you are in a position to harm or benefit by doing one thing or another. In Part II of the book, Chapters Seven to Eleven, I will extend the treatment to cover cases in which you do know who you are in position to harm or benefit by doing one thing or another. To do it I will need fancier tools than before.

In Chapter Seven I spell out an assumption about people and their essences. Each of us could have been ever-so-slightly-different along any natural dimension of sameness and difference. You could have been a millimeter taller than you actually are. I could have been conceived a second before I was actually conceived. Barak Obama could have been slightly more irascible than he actually is. Our essences are not perfectly fragile.

In Chapter Eight I spell out a further, quiet assumption about rationality. If you are rational then your desires are coherent – which means, at least, that your preferences between maximal states of affairs (fully specific ways for everything to be) are transitive.

In Chapters Nine to Eleven I put these new assumptions to work. In Chapter Nine I argue that it follows that, if decency obliges us to prefer, in some circumstances, that some people be better off rather than worse off, then decency and rationality together oblige us to prefer, in some circumstances, that some people be better off rather than other people better off. I call this the *Morphing Argument*. I explore its consequences for the problems we have looked at so far, problems involving who to save, the non-identity problem, and the problem of whether to kill-to-prevent-killing, in Chapter Ten.

Chapter Eleven is about the limits of good-will towards others. One surprising consequence of the morphing argument is that it is impossible to be both rational and minimally benevolent towards *everyone*. Rationality itself places limits on how good-willed we can be.

Which brings us to the third major problem that I discussed in the Introduction – our moral obligations towards distant strangers. In Chapters Twelve and Thirteen I argue that the rational requirement that preferences be transitive and the moral requirement that we be at least minimally benevolent towards (some) strangers together place great pressure on us to attend to needy, distant strangers. It is much harder than we ordinarily think to be both decent and rational.

6. Two Goals

Part of what I hope to do here is to shed light on some interesting problems in normative ethics. Another part of what I want to do is to show off the benefits of approaching the subject in this way – by thinking about rationality and minimal benevolence. In typical introductory classes in moral philosophy we teach our undergraduates that there are three approaches to normative ethics. There's the consequentialist approach (usually traced back to Jeremy Bentham – with a nod at the ancient Epicureans), which has it that acting morally is about bringing about the best states of affairs. There's the deontological approach (usually traced back to Kant, with a nod to contractualist and rights-based alternatives), which has us focus on the character of the act itself, with an emphasis on whether it is universalizable. And there's the virtue-based approach (usually traced back to Aristotle), which has us focus on how the act reflects on the moral character of the agent – is it the sort of thing that a perfectly virtuous (honest, kind, loving, courageous…etc.) person would do? The approach I am exploring here does not fit neatly into any of these traditions. I think this is a good thing, as you will see.