The Justification of Deontology

by

Gaurav Alex Sinha

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

© Copyright by Gaurav Alex Sinha, 2011
Agent-centered restrictions are widely accepted both in commonsense morality and across social and legal institutions, making it all the more striking that we have yet to ground them in a compelling theoretical rationale. This dissertation amounts to an effort to fill that gap by seeking out a new principled basis for justifying such constraints. I devote each of the first three chapters, respectively, to the three established deontological normative ethical theories: Rossian intuitionism, Kantianism, and Neo-Thomism. In each of these chapters, I lay out the relevant portion of the view's deontological apparatus, analyzing it both for its plausibility as a whole and for its ability to justify constraints of the appropriate shape. After assessing and rejecting all three approaches, I devote the next two chapters to developing a new rationale for grounding constraints—one that avoids the pitfalls indicated in the prominent historical alternatives. Specifically, I anchor constraints in the distinction between the agent-neutral and agent-relative points of view, basing them in the widely accepted psychological fact of the natural independence of the personal point of view.
Acknowledgments

I started tinkering with the key ideas presented in this thesis a couple of years before I began writing it, but only after I received guidance from my committee was I able to put those ideas into context and fully understand them. While I wish very much for this project to contribute something of note to the body of normative ethical literature, I hope at the very least to have done justice to my committee. I especially hope to have lived up to the standards and expectations of my supervisor, Tom Hurka. Were it not for his invaluable assistance, there is no question that the project would have failed before it even had a chance to succeed. This thesis would not exist but for his efforts to keep me on course. I am also deeply indebted to the other members of my committee, Wayne Sumner and Gopal Sreenivasan, for their thoughtful comments and their patient beneficence in the face of my countless inquiries, even (and especially) when they were confronted with far more pressing matters.

Additionally, I am grateful for the helpful suggestions provided on numerous occasions by Chris Langston, Diego Silva, and Luke Gelinas. I am thankful to all three of them for their enthusiastic and unflinching willingness to discuss various elements of this project with me over the course of several long years. I also owe much to my mother, who encouraged me unconditionally, coaxing me past the various frustrations I encountered along the way. Lastly, but certainly not least, I cannot imagine completing this project at all
without the boundless and ever-present support of Janani Umamaheswar, who was there every step of the way, never shrinking from the opportunity to help—whether I needed an interlocutor, an editor, or an anchor.
# Table of Contents

Acknowledgements \hspace{1cm} iii

Introduction \hspace{1cm} 1

## Chapter 1: Intuitionism

1.1. Intuitionism in Historical Context \hspace{1cm} 5
1.2. Ross-Style Pluralism: The Good \hspace{1cm} 9
1.3. Ross-Style Pluralism: The Right \hspace{1cm} 12

## Chapter 2: Kantianism

2.1. Kant’s Theory of the Right \hspace{1cm} 38
2.2. The Formula of Universal Law \hspace{1cm} 41
2.3. The Formula of Humanity \hspace{1cm} 62
2.4. The Formula of Autonomy and the Formula of the Kingdom of Ends \hspace{1cm} 76

## Chapter 3: Neo-Thomism

3.1. John Finnis’ Normative Ethical Theory \hspace{1cm} 82
3.2. The Implications of Neo-Thomistic Deontology \hspace{1cm} 89

## Chapter 4: Tapping The Natural Independence of Human Agency

4.1. Background \hspace{1cm} 107
4.2. Developing the New Approach \hspace{1cm} 112
4.3. What Normative Features does Agent-Relativity Yield? \hspace{1cm} 133
4.4. Frances Kamm’s “Patient-Centered” Justification For Constraints 157

Chapter 5: The Viability and Merits of Agent-Centered Deontology 160

5.1. The Paradox of Deontology 160
5.2. A Reply to Samuel Scheffler 170
5.3. Shelly Kagan’s Skepticism 185
5.4. The Promise of the New Approach 190
5.5. Intuitionism, Kantianism and Neo-Thomism Revisited 196

Works Cited 202
Introduction

“Individuals have rights,” Robert Nozick famously declared, “and there are things no person or group may do to them (without violating their rights).” This line frames the starting point of Nozick’s magnum opus, Anarchy, State and Utopia; it is the first sentence of Nozick’s memorable preface, and it is noteworthy not for its originality but rather for the moral commonplace it expresses. Rights populate our world, both in moral and legal form. We speak of them with facility and remarkable certainty. We assume them; we invoke them; and most of all, we take them for granted. Indeed, only in such a world could Nozick begin a book with this line and offer a subsequent argument that is anything but a vanity exercise. His project requires that almost no one find his premise fanciful.

Of course, common acceptance of an idea does not guarantee its veracity. We may be mistaken about the existence of moral rights, and we may be mistaken about the shape of the legal rights that we bestow upon ourselves. But appropriately or not, we have adopted Nozick’s assumption and put it into practice. Our intuitions and our laws carve out a space around agents, granting them the liberty to make a variety of choices about how to act and guaranteeing them freedom from various forms of interference as they pursue their goals. We find broad support for these options and constraints even in cases (or because of cases) where they interfere with the production of impersonally optimal
states of affairs. In short, we have institutionalized deontology.

Unfortunately, it is far from obvious that deontology is “true.” It is not clear, in other words, that any deontological system accurately captures our objective moral duties, assuming we have any at all. But I think that we do have such duties, and we can proceed from the knowledge that the intuitive and institutional abandonment of deontology are not forthcoming. Given our reliance on deontological devices, it would be helpful to have a deontology that we can believe in. The purpose behind the pages that follow is to seek out just such a system. The institutional canonization of options and constraints, central as they are to the moral regulation of interpersonal interaction, offers its own independent rationale for pursuing the rational basis of these deontological devices. We might never know that deontology is “true” in the same way that we can verify the atomic weight of gold, but understanding our prospects for giving a rational account of our moral intuitions is worth nearly as much.

At minimum, properly assessing the basis of options and constraints requires an evaluation of the most prominent deontological systems—namely Intuitionism, Kantianism and Neo-Thomism. I will describe and assess each of these three major deontological proposals, in this order across the first three chapters, with an eye toward their plausibility and intuitive fit. All three of these systems suffer from a variety of deficiencies, and my evaluation of these major approaches will lead me to conclude that none are satisfactory. (I leave aside indirect consequentialism entirely, even though in some forms it may be coextensive with deontology for practical purposes. My reason is that the intuitive support for deontology extends beyond such practical purposes to the theoretical justification that lies in the background, and those intuitions do not appeal, at
bottom, purely to consequences.)

I will proceed in Chapters 4 and 5 to develop the beginnings of deontological account that I suggest is superior to the alternatives. I will use these chapters to argue that the intuitive appeal of options and constraints is not illusory or irrational, for that appeal issues directly from a particular fact about our psychological constitution. I will claim that this fact, which I follow Samuel Scheffler in calling the “natural independence of the personal point of view,” is of normative significance: it warrants some degree of recognition in a well-formed theory of the right. I will suggest that normatively adequate reflection of the natural independence of the personal point of view justifies both options and constraints.

After illustrating this new account of the justification behind options and constraints, I will turn to a discussion of how both devices serve crucial roles in reflecting the natural independence of the personal point of view, and in particular, how constraints overcome the conventional challenges they are thought to encounter in doing so. I will also advocate for an adjustment in our terminology regarding these devices, pushing for the introduction of the term “imposition” to serve as the obverse of “options” and the term “protections” to serve as the obverse of “constraints.” Finally, I will close by elaborating on the advantages I see in an account of constraints that possesses a structure of the sort I have in mind.

My aim, in sum, is twofold: first, to defend commonsense morality by arguing that the sense that commonly lies behind it has moral significance that ought not to be ignored; and second, to establish that commonsense morality can, and does, coherently place a premium on the normative ethical device we refer to as “constraints.” I view the success of
some such project as essential to justifying (properly, at any rate) a wide range of claim
and liberty rights—most directly moral ones, but legal ones as well. I thus see my purpose
as sketching, in some detail, the beginnings of a viable deontology.
1

Intuitionism1

Introduction

Among the dominant deontological approaches, the intuitionist model represents
the simplest way of grounding agent-centered restrictions. W.D. Ross is the most famous
proponent of such a view, though several of his contemporaries defended similar
positions. Because of Ross’ philosophical prominence, I will focus on his version of
intuitionism, largely as presented in his book, The Right and the Good. More specifically, I
will begin this chapter by outlining the historical context in which Ross wrote before I
proceed to give a detailed exposition and critique of his view.

1.1. Intuitionism in Historical Context

The qualities shared by the primary proponents of intuitionism extend beyond
their philosophical inclinations. Historically, the chief advocates of intuitionism were
British analytic philosophers who were posted at Oxford and Cambridge between the late
1800’s and the first half of the 1900’s. Their physical and temporal proximity made for

1 I am particularly indebted to Tom Hurka for the content of this chapter. As an expert on
this historical period, he has taught me almost everything I know about the intuitionist
movement.
lively exchanges, and they disagreed with one another on a number of significant matters. For example, some of them, like Hastings Rashdall, G.E. Moore, and J.M.E. McTaggart, applied an intuitionist approach to consequentialism—favoring “ideal consequentialism” or perfectionism. (We might include Henry Sidgwick among this sub-group, though he seems to stand out as the lone utilitarian in the club.)

Yet even these consequentialists shared certain core assumptions with their more deontologically-oriented counterparts, such as H.A. Prichard, E.F. Carritt, C.D. Broad, A.C. Ewing, and, of course, W.D. Ross. Both sides of the group thought that thick concepts are reducible to thin concepts, so they did not speak in classic Aristotelian terms when motivating their views. Instead of appealing to a broad array of virtues, like courage or justice, they generally appealed to a small number of thin concepts instead. For example, Sidgwick boiled everything down to “what one ought to do,” and Moore spoke primarily of “intrinsic good.” Others appealed to the concept of “fittingness,” while Ross highlighted his favored pair of thin concepts in the title of his book: the right and the good.

The group as a whole also accepted the idea that at least some moral claims are objectively true or false (a position that we can follow convention in calling “moral realism”). That the intuitionists were realists is not particularly surprising, given that they defended full normative ethical theories whose prescriptions hinge rather heavily on objective support. The overwhelming theoretical interests of the intuitionists aimed at the normative level, and they seem to have generally assumed their realism without argument. More interesting is that they paired their realism with non-naturalism, rejecting the idea that moral truths could be reduced to or derived from any other class of truth—whether scientific, mathematical, geometric, or whatever. By classifying ethical truths as
_sui generis_—accepting what is sometimes called “the autonomy of ethics”—they invited the question of how agents can come to know moral truths.

This is where the name “intuitionism” seems to have been born. The group did not posit a mysterious new faculty for grasping moral facts. Instead, they favored the idea that we come to know moral truths by intuition. It is not that they thought intuition is infallible; it did not escape the intuitionists that intuitions diverge across different agents, and therefore that some of our intuitive judgments must be incorrect. But the basic idea that they accepted is that careful reflection on (and organization of) our intuitions can yield knowledge of objectively true moral facts. While there was some disagreement about whether intuitions are clearest in evaluating concrete cases (as Carritt thought) or, instead, general principles (ala Moore), the group tended to accept the idea that careful handling of intuitions could produce moral knowledge.

Reliance on intuition often brought with it an emphasis on the validity of commonsense morality.² This concern manifested itself in somewhat different ways for ideal consequentialists from for deontologists, but here we may follow the deontological vein. Among the deontologically-inclined intuitionists, there were a pair of significant concerns: consequentialism frequently churns out incorrect verdicts on particular cases, and when it does happen to provide the right answer, the consequentialist explanation for that answer is plainly incorrect. Ross endorsed both of these criticisms, but he emphasized the latter.

Suppose, as is likely the case, that we oughtn’t to sacrifice an innocent man in order

² As a utilitarian, Sidgwick was a notable exception.
to appease an unruly crowd, even if doing so may prevent a deadly riot. Many consequentialists (or direct consequentialists, at any rate) must contort themselves to uphold this prohibition. But even if they can, their justification for it must be deeply implausible—or at least it would be deeply implausible according to the deontological intuitionist. When pressed, the (direct) consequentialist would have to maintain that the ultimate reason for protecting the innocent man derives from the negative consequences for other people of permitting his murder. But that isn’t why our intuitions seem to cry out against killing an innocent man; intuitively, it seems as if it’s something about him—perhaps his own moral status or his innocence—that imposes the restriction on killing him. According to the intuitionists, similar concerns militate against derivative deontological approaches. They worried that we distort our moral duties by channeling them through explanatory back-stories, such as those offered by the neo-Thomistic and Kantian accounts. To take a very rough example, facts about my own flourishing seem as inadequate for explaining why I oughtn’t to kill the innocent as are concerns about the overall consequences of my doing so.3

To avoid this problem, the intuitionists opted for underivative duties.4 Essentially by definition, underivative duties come with no detailed explanation of what makes them duties. If we want to know more about why we have certain underivative duties, we’re out of luck. If we want to know why we should fulfill our (underivative) duty, the intuitionists

3 This is an important point, and it goes to the heart of the intuitive fit of an ethical theory. We shall return to it in more detail later.

4 Tom Hurka has introduced terminology to mark this distinction. He calls an account “structural” if it uses underivative duties, and “foundational” if it derives its duties from some source.
would respond that that very question reveals a fundamental misunderstanding of duties. To have a duty is just to have something you ought to do. A structural account lacks the resources to say much more than that, though of course, the intuitionists did not regard that as a weakness of their approach.

1.2. Ross-Style Pluralism: The Good

This is the context that shaped W.D. Ross' *The Right and the Good*, which appeared in 1930, towards the end of the “intuitionist era.” Ross accepted the basic package of intuitionist positions, and, in his book, articulated what has come to represent one of the best-known formulations of a deontological intuitionist view. He defended a variation of intuitionism that some have called “Ross-style pluralism,” which posits six different duties oriented around four basic goods. Though some of Ross’ contemporaries also accepted multiple duties and goods (such as Prichard, Broad and Ewing), Ross’ version of such a view is the most famous.\(^5\)

Ross began his book by addressing his selection of thin concepts, or more specifically, explaining his reasons for separating the right from the good. His argument for the separation essentially takes the form of the Open Question Argument; if “the right” and “the good” refer to the same things, then we should be able to use the two phrases

\(^5\) Other intuitionists rejected the notion that we have a plurality of duties, and even, in one major case, a plurality of goods. Sidgwick is the prime example of the latter; he accepted pleasure as the only good, and thought our sole duty was to maximize it. Meanwhile, idealists like Moore and Rashdall accepted multiple goods, while still opting for one duty (again, maximizing the good). Perhaps surprisingly, even Ross’ view, with its plurality of duties, still carries significant consequentialist implications. Now that we are ready for a detailed exposition of Ross’ view, we shall see why.
interchangeably. Ross thought it clear that the two expressions are not functionally equivalent. As he put it:

If ["the right" and "the good"] meant the same thing we should be able to substitute, for instance, 'he is a right man' for 'he is a morally good man'; nor is our inability to do this merely a matter of English idiom, for if we turn to the sort of moral judgment in which we do use the word 'right', such as 'this is the right act', it is clear that by this we mean 'this act is the act that ought to be done', 'this act is morally obligatory'; and to substitute either of these phrases for 'morally good' in 'he is a morally good man' would obviously be not merely unidiomatic, but absurd.6 (Ross, 3, emphasis in original)

Ross proposed using "right," most roughly, to describe doing one's duty, while using "morally good" to denote an agent's acting from the proper motives. He made a further distinction between "acts" and "actions": Ross thought it best if "we used 'act' of the thing done, the initiation of change, and 'action' of the doing of it, the initiating of change from a certain motive." (Ross, 7) That would allow us to speak of "a right act but not of a right action" and "of a morally good action but not a morally good act." (Ibid)

Ross also thought it clear that the concept of moral goodness—the same concept as intrinsic goodness, except applied to states of character—stood distinct from the concept of the good. As noted above, Ross held a pluralistic view of the basic goods in the world. He ultimately accepted four primary goods: virtue, knowledge, pleasure, and the apportionment of pleasure in accordance with virtue. He thought of the first three of these as "complex states of mind that we think of as good in themselves,"; the apportionment of pleasure in accordance with virtue, on the other hand, is a state of affairs that stands

outside of the minds of individual agents. (Ross, 141)

Ross used a simple test for deciding which items belonged on his list of basic goods. The test involves comparing two hypothetical worlds for intuitive appeal. The idea is to imagine a pair of worlds, alike in all respects but one, and to select the one that is more appealing. That way, if either world seems better than the other, its seeming so has to follow from the single difference between them. Ross would design the comparison so that the sole difference between the two worlds in question was the presence of some small measure of a putative good. If, intuitively, it seemed that that small measure of putative goodness made one world better, then the item responsible for the verdict made its way onto the list.

Imagine two worlds—call them “World A” and “World B”—that are alike in almost all respects. Each has the same physical features, and is inhabited by exactly one agent. The only difference between them is that in World A the agent is virtuous, while in World B the agent is vicious. Does World A seem better than World B? Ross thought so; and as such, he put virtue on his list of basic goods. Ross used the same process to determine the inclusion of pleasure, knowledge, and the apportionment of pleasure with virtue—as well as to disqualify various other contenders that he left off of the list.

7 Strictly speaking, these two worlds may not be ideal for our comparison, as one contains virtue and no vice, while the other contains vice but no virtue. It appears, then, that there are two variables in our comparison (virtue and vice), instead of just one (either virtue or vice). Perhaps it would be better if one world contained virtue while the other contained neither virtue nor vice. Still, I’ve set it up this way to preserve a clear-cut symmetry, and I don’t think it results in a material change in the outcome of the test.
1.3. Ross-Style Pluralism: The Right

So much for the good; but what about the right? On the Rossian view, the relationship between the right and the good is relatively complicated, largely because Ross thought that there is no singular right-making property that can unify all duties by giving all of them an identical justification. Ross also rejected the possibility that there is some “general character which makes right acts right,” as suggested by such views as egoism and hedonism.⁸ (Ross, 16)

Ross dismissed the idea that right acts are necessarily the ones that maximize the good—irrespective of how we define the good. His reason for rejecting this quintessentially consequentialist formulation of the relation between the right and the good is thus not that he was a pluralist who believed his list of goods to be incommensurable with one another. Rather, Ross thought it intuitively obvious that consequential considerations neither give us nor explain many of our duties. For example, “When a plain man fulfills a promise because he thinks he ought to do so, it seems clear that he does so with no thought of its total consequences, still less with any opinion that these are likely to be the best possible. He thinks in fact much more of the past than of the future.” (Ross, 17)

Further, Ross thought that when it looks like we ought to violate what we take to be a duty, the justification for the ostensible violation may not best be thought of as relying on a contrary course of action that clearly produces the most good. He imagined a case where I have made a promise to meet a friend for some “trivial purpose.” (Ross, 18) On my

⁸ “I would contend that in principle there is no reason to anticipate that every act that is our duty is so for one and the same reason.” (Ross, 24)
way to the meeting, I pass the scene of a serious accident. I am faced with a dilemma: on one hand, I can keep the promise I made to my friend; on the other, I can stop to offer relief to a badly injured stranger, thereby missing my meeting. If we think that I should break my promise to meet my friend in order to help the stranger, Ross believed this is because we also recognize a duty of relieving distress. And if we think that I should help the stranger, Ross argued that it is not because we believe that doing so will produce more good than keeping my promise. Instead, we perceive that the duty of relieving distress “is in the circumstances more of a duty.”9 (Ibid)

This way of understanding duties led Ross to articulate his now-famous distinction between prima facie and actual duties, a distinction that many seem to find appealing.10 The idea behind the distinction is that we may stand in any number of different relations to other people, and each relation provides the grounds for considerations that may end up dictating how we should act in any given situation. Ross claims that consequentialists dramatically oversimplify the array of possible relations that people stand in to one another. For a consequentialist, everyone around me stands only as a potential beneficiary

9 It seems as if, at least in certain cases, our sense of duty tracks consequentialist reasoning quite explicitly, and it is unclear if Ross would accept such descriptions as legitimate. In classic trolley cases, where a conductor must choose between killing one or kill five, Ross might say we have more of a duty to abstain from killing five than we do to abstain from killing one. Perhaps this is a matter of semantics, but even if Ross is correct, he is correct purely on consequentialist grounds. In those cases, Ross’ description seems inaccurate.

10 We find at least one contemporary analog in Judith Thomson’s The Realm of Rights, except she describes it as a distinction between “oughts” and “all-things-considered oughts.”
of my action. But Ross thought that it was a mistake to omit the myriad other relations we can stand in to one another, such as that “of promisee to promiser, of creditor to debtor, of wife to husband, of child to parent, of friend to friend, of fellow countryman to fellow countryman, and the like.” (Ross, 19)

Each of these relations provides the ground for a *prima facie* duty: promisers have *prima facie* duties to promisees, debtors have *prima facie* duties to creditors, and so on. Each *prima facie* duty is “more or less incumbent on [us] according to the circumstances of the case.” (Ross, 19) According to Ross, then, moral decision-making involves carefully weighing the relevant *prima facie* duties against one another to determine which trumps the others in one’s present situation. As Ross puts it:

> “When I am in a situation, as perhaps I always am, in which more than one of these *prima facie* duties is incumbent on me, what I have to do is to study the situation as fully as I can until I form the considered opinion (it is never more) that in the circumstances one of them is more incumbent than any other; then I am bound to think that to do this *prima facie* duty is my duty *sans phrase* in the situation. Whatever we ought ultimately to do in any given situation is our actual duty in that particular case, or our duty *sans phrase*. (Ibid)"

Strictly speaking (and as Ross admits), the term “*prima facie* duty” is misleading on at least two counts. First, a *prima facie* duty is not really a duty at all; it is rather “something related in a special way to duty.” (Ross, 20) It is a consideration of sorts, a defeasible ground of obligation that counts in favor of the possibility that a particular course of action will ultimately be one’s duty.

Second, the phrase “*prima facie*” implies that *prima facie* duties are only illusory, or

---

11 Perhaps we should say that the consequentialist can also make sense of a second relation between agents—that of a potential victim.
maybe that they are moral features of the situation that dissipate after careful consideration renders some more important than others. This is also inaccurate, for Ross thought of *prima facie* duties as objective facts pertaining to the “nature of the situation” at hand. *(Ibid)* In some cases, *prima facie* duties even persist when they do not account for our duty. We might very well think that I should break my promise to meet my friend in order to help the injured stranger, but the *prima facie* duty I have to keep my promise does not disappear; it remains, leaving a residue. As a result of that residue, I might owe reparations to my friend—even if I have a good reason for standing her up for lunch. At the very least, I probably owe her an apology and an explanation.

Ross thought that the *prima facie* rightness of various acts—keeping promises, telling the truth, and so on—strikes us as self-evident.

“No to me it seems as self-evident as anything could be, that to make a promise, for instance, is to create a moral claim on us in someone else. Many readers will perhaps say that they do not know this to be true. If so, I certainly cannot prove it to them; I can only ask them to reflect again, in the hope that they will ultimately agree that they also know it to be true.” *(Ross, 21)*

But these propositions only strike us as self-evident when we have reached the point of “sufficient mental maturity.”*(Ross, 29)* Once we reach that point, and “have given sufficient attention to [a particular] proposition,” then Ross thought that its *prima facie* rightness would be “evident without any need of proof, or of evidence beyond itself.” *(Ibid)*

In this sense, *prima facie* rightness parallels mathematical axioms; both, thought Ross, are self-evident in the very same way, hinging on the proper sort of reflection paired with

12 We will review Ross’ list of *prima facie* duties shortly.
sufficient mental capacity.

Moving from recognizing our *prima facie* duties to recognizing our actual duty is a bit tricky, however. Ross likely thought that there is an objective fact about what we ought to do in any given situation. In other words, an objective truth identifies our actual duty (or duty *sans phrase*) in any particular set of circumstances. But that fact is not available to us *qua* fact. Unlike the principles that determine our *prima facie* duties, which Ross took to be self-evident, he claimed that we could never identify our actual duty with absolute certainty. In aiming at our actual duty, we have to settle for a “considered opinion” instead. A considered opinion is just the result of a careful evaluation of the situation, appealing to common sense, and past precedent, and so on. Ross offers no specific list of criteria for locating our actual duty in any given case; he actually thought that no such list could ever be provided. As he put it, “For the estimation of the comparative stringency of [our] *prima facie* obligations no general rules can, so far as I can see, be laid down.” (Ross, 41)

The reason for the indeterminacy at the level of duty *sans phrase* is that, according to Ross, our actual duty in any given situation is neither self-evident, nor does it logically follow from any self-evident premises. (Ross, 30-31) It is not self-evident because, “any particular act will in all probability in the course of time contribute to the bringing about of good or of evil for many human beings, and thus have a *prima facie* rightness or wrongness of which we know nothing.” (Ross, 31) Meanwhile, the only self-evident

---

13 Ross thought that there are objective facts about the volumes of pleasure to be gained or lost as a result of any given choice, even if we cannot pinpoint them. I’m inferring Ross’ position on the good from those comments about pleasure.

14 Note the overwhelmingly consequentialist tone Ross adopts here, a point that we will return to later.
premises our actual duty could follow from are the general principles that determine *prima facie* duties. But Ross took it as obvious that the answer to various moral dilemmas did not logically follow from any set of *prima facie* duties such that even if we could foresee the most distant consequences of any action, we would be unable to deduce our actual duty just from the relevant set of *prima facie* duties. *(Ibid)*

Ross offers an analogy for helping us to understand the fuzzy relation between *prima facie* duties and actual duties. He claims that “the judgment as to the rightness of a particular act is just like the judgment as to the beauty of a particular natural object or work of art.” *(Ross, 31)* More specifically, “A poem is... in respect of certain qualities beautiful and in respect of certain others not beautiful; and our judgment as to the degree of beauty it possesses on the whole is never reached by logical reasoning from the apprehension of its particular beauties or particular defects.” *(Ibid)* The idea seems to be that we can easily pick out various appealing and unappealing features of a piece of art, and while the ultimate verdict on whether we find the artwork beautiful as a whole depends in some way on its various parts, there is no clear-cut, logical relation between the appealing features and our overall impression.

Similarly, Ross thought that we detect *prima facie* rightness just in virtue of the clarity with which we recognize the relevant principles. But weighing everything out to determine our actual duty is not a precise exercise in deduction, which is why Ross thinks there is “much truth in the description of the right as a fortunate act.” *(Ross, 31)* Still, just as we can never be sure that what we think is in our best interests ultimately will prove to be, Ross thinks it sufficient that the very best we can do is reflect carefully on the various *prima facie* duties embedded in any given situation and then pick the course of action that
strikes us as most likely to be correct.

Yet Ross’ claim that we can never know our actual duty in any given case suggests a couple of problems. Most obviously, it strikes some as implausible that no conceivable choice situation would yield an array of foreseeable outcomes so conclusive as to give us a definitive sense of which one we should pursue. As we saw above, part of his concern hinged on the fact that the far-reaching consequences of our actions are inaccessible to us; and that as such, even if we could confidently foresee wildly disparate near-term consequences of various courses of action, we will always be working from incomplete information.

There is a sense in which it seems hard to dispute Ross’ view that the lack of complete information would have to impair our judgment to a certain degree. To modify an example given by Tom Hurka, we might think it is perfectly clear that we ought to abstain from bombing downtown Toronto, even if it is the last wish of a dying friend and we have promised her that we will honor her last wishes. All information available to us may obviously militate against gaining some measure of personal pleasure at the expense of countless lives. But we can imagine a situation in which bombing Toronto actually initiates a chain of events that validates the decision to do it, even if not for our initial motive. Strictly speaking, we cannot rule out the possibility that the unforeseen factors count in favor of bombing Toronto, irrespective of the fact that they are extremely unlikely to obtain. In other words, some dimensions of the prima facie rightness and wrongness of bombing Toronto will inevitably remain opaque to us simply in virtue of the number of people who stand to be affected by the act, both now and in the future.

Nevertheless, we might think that the level of indeterminacy built into our example
is effectively irrelevant, at least from a moral standpoint. We can concede that we do not know *everything*; but we seem to know enough to say that our actual duty involves abstaining from bombing the largest city in Canada. Interestingly enough, not only would Ross deny this, but he would go a step further: even if we could foresee all of the consequences of any given set of choices, and even if we had a full grasp of the *prima facie* rightness and wrongness of the various alternatives before us, he would still maintain that we cannot know our actual duties with full confidence. Specifically, Ross would say that our actual duties are never self-evident (like our *prima facie* duties are), nor do they follow logically from self-evident premises.

Principled though it is, this level of skepticism about our actual duties is troublesome. It is legitimate to question whether Ross oversteps his warrant in denying us direct access to our actual duties. For one, many might find that it *is* self-evident that we ought not to bomb Toronto—including Kant, neo-Thomists, indirect consequentialists, and even some of Ross’ intuitionist contemporaries, like E.F. Carritt. In pushing back, Ross could cite the impossibility of developing a clear-cut, deductive formula for deriving actual duties from the relevant *prima facie* factors. But that sort of reply is not responsive to someone who finds his actual duty to be directly and intuitively self-evident. In the face of someone who maintains as much, Ross runs out of things to say. Moreover, even if Ross is right that our actual duties are neither self-evident in the proper way nor logically deducible from self-evident premises, we take full confidence in many things that do not meet that standard. For example, many of us take that level of confidence in the dominant scientific explanations of phenomena like gravity and climate change, even when those explanations are neither self-evident nor deduced from self-evident principles.
Additionally, this entire discussion highlights the extent to which consequentialist considerations influence Ross’ deontology—and not just as a result of the limited list of duties Ross found to be self-evident. On Ross’ view, our actual duty in any given situation depends heavily on situational factors, including our reasonable estimates of the distant consequences of our actions. This feature of Ross’ view issues from his broad construal of which factors are *prima facie* right and wrong, and thus, which circumstantial factors ought to enter into any considered judgment about our actual duty. As a result, not only are Ross’ *prima facie* duties prone to circumstantial exceptions, but in classic consequentialist fashion, they are also evaluated against a grid of foreseeable effects. Here we see a stark contrast between Ross’ view and a view like Kant’s (which we will discuss in detail in the following chapter). Kant’s (in)famously inflexible injunction not to lie even if enormous good can be gained by it diverges sharply from the heavily qualified and conditional commands that fall out of Ross’ approach.

We can also ask serious questions about Ross’ test of self-evidence for *prima facie* duties. This is not an uncommon criticism, but it is piercing, and so deserves repetition. If *prima facie* duties are self-evident, it seems that Ross should be able to list all of them definitively—just as he does with his list of intrinsic goods and evils. Moreover, if our intuitions track facts when they capture self-evident *prima facie* duties, how is it that intuitions can diverge so widely across various agents? This is a significant criticism not only because it raises meta-ethical questions about the reliability of our intuitive faculty. Ross thought that *prima facie* duties are underivative, which means that they defy unification and further explanation. That they are particularly resistant to reasoned explanation (at least according to Ross) only underscores the importance of a reliable
faculty for detecting them. The key issue is that there is no valid set of criteria for mediating a debate over the intuitive salience of any putative *prima facie* duty. We have no concrete way of resolving a good faith disagreement over which duties belong on the list. That fact seems to drain the normative force of any list that Ross (or others) could provide by raising a nontrivial worry about the validity of any given list.

In any case, a major reason that Ross found this conception of duty appealing is that he thought it properly captures how we think about difficult moral decisions where we feel drawn in multiple directions. To draw this out, he explicitly contrasted his approach with the Kantian and consequentialist alternatives on this particular metric. According to the Kantian position, “perfect obligations” trump “imperfect obligations” when the two come into conflict, an arrangement that eliminates a large subset of potential moral dilemmas. On the consequentialist approach, there is only one duty in the first place: maximizing the good. As such, no actual dilemmas are possible—except to the extent that we cannot discern which course of action would maximize the good.

Ross did not find it plausible that “perfect duties” (such as telling the truth) should have lexicographical priority over “imperfect” ones (such as promoting another’s good), and from a purely intuitive angle, that view seems on point. Further, we have already seen how he outright rejected the idea that our duty would always align with maximizing the good. Thus, while Ross admitted that the Kantian and the consequentialist approaches are simpler than his own, he thought his was superior, claiming, “...it is more important that our theory fit the facts than that it be simple.” (Ross, 19)

---

15 I use “dilemma” loosely here to cover trilemmas and other, more complex situations in which we face an array of unappealing choices.
The question of factual support aside, Ross does capture a couple of important elements of our moral intuitions here. By framing dilemmas as a conflict of *prima facie* duties, he makes sense of the notion of defeasible “duties”—duties that can be outweighed by other considerations (as opposed to duties that are not even triggered to begin with). Moreover, Ross’ system accounts for the feeling we have that outweighed duties can leave behind a morally relevant residue.¹⁶ Justifiable promise-breaking represents the classic source of such a residue. Returning to our example from above, even if I have agreed to meet my friend for lunch, my duty to keep that promise may bind me with less force than a duty of beneficence to the injured stranger I pass on my way to the meeting. Yet even if it is incumbent upon me to skip my lunch appointment, I may still owe my friend reparations for having broken my commitment—notwithstanding the fact that, strictly speaking, I acted morally in ditching him in the first place.

It is significant that the defeasibility of duties follows from a structural feature of Ross’ view rather than from the nature of any particular duty on Ross’ list. Paradigm though they may be, promises are not Ross’ exception to some absolutist deontological rule; in theory, all of Ross’ duties are open to violation—under the right circumstances, of course. As a result, Ross captures a significant slice of consequentialism’s appeal by allowing us to acknowledge that the numbers matter sometimes. Many recoil at the thought of killing one to save two, but fewer resist the prospect of killing one to save a million. In the latter case, Ross might agree with the consequentialist by saying that we

¹⁶ I have put “duties” in quotation marks because, for Ross, dilemmas pit two competing *prima facie* duties against each other instead of two actual duties. As discussed above, in the strictest sense, *prima facie* duties are not duties of any sort. For convenience, I will drop the quotation marks going forward.
ought to kill the one. Ross can easily take stock of this intuition by affording the proper weight to the *prima facie* duties of beneficence and non-maleficence.\textsuperscript{17}

Unlike the consequentialist, however, Ross offers us a view with a formal feature recognizing that a regrettable sacrifice took place. At the intuitive level, *something* seems wrong with sacrificing an innocent person, *even if we think that doing so is justifiable because of the harm it prevents.* The consequentialist cannot account for this intuition in any clear way, but Ross can. Ross might say that the moral residue of the duty of non-maleficence remains here, even though the duty of beneficence wins out in the end. Notice that in this case, we cannot properly be said to owe reparations to the victim, as in the case of the broken promise. The victim is dead. Neither the most respectful treatment of his corpse nor the most kingly of burials can offset the harm he suffered at our hands. So it is not that the residue remains *only so that we can make amends*, as we might have thought based on the case of the broken promise. No; we have violated a duty (of sorts), and even though we made the right decision (all things considered), the moral stain of killing an innocent person does not simply wash away because of the surplus of good we purchased with his life. Beyond any reparations we can make, we ought to feel compunction at having had to violate the *prima facie* duty not to kill.

\textsuperscript{17} It is important to note (as mentioned previously) that Ross will not always want to ascribe the circumstantial dominance of any particular *prima facie* duty directly to the volume of agent-neutral good that comes with it. (He thought it self-evident that, for example, promises are meant to withstand certain gains in what we might call agent-neutral good.) Instead, he thought that the dominant duty is simply *more of a duty* in that particular case. Nevertheless, I slip into the standard contemporary philosophical parlance of tradeoffs, and I do so largely for facility. Strictly speaking, it is an independent question whether we endorse a particular *prima facie* duty because of the correlative tradeoff or because it is “more of a duty.” We need not commit to either interpretation here and now.
Moreover, Ross does not prejudge the strength of any given duty on his list, except to say that the duty of non-maleficence seems generally more binding than the duty of beneficence. More specifically, Ross seems to have regarded moral decision-making as complex enough to preclude the possibility of deriving a particular formula for the comparative strength of the various *prima facie* duties. That is because Ross thought that situational factors should play a role in shaping our deliberations. Still, we can bracket that issue to get at the underlying question. We might think that, all else equal, killing one to save 20 is as consequentialist as we want to go; or instead, we might hold that killing one cannot be justified except to save at least a million. A Rossian system is consistent with either view, and anything in between.

Whatever position we take on tradeoffs, the truths expressed in accurate propositions about *prima facie* rightness are “just as much a part of the fundamental nature of the universe (and, we may add, of any possible universe in which there were moral agents at all) as is the spatial or numerical structure expressed in the axioms of geometry or arithmetic.” (Ross, 30) This is a crucial point for the analysis that will follow below. For every normative ethical view, the buck stops somewhere, and for Ross, it stops particularly early. On his view, *x* is *prima facie* right (if it is) in the same way that two plus two amounts to four. Most significantly, Ross thought that there is no reason to prove these truths because they “certainly need no proof.”

---

18 It is unclear whether Ross thought it possible to prove truths about *prima facie* rightness. In *The Right and The Good*, he implies that they cannot be; but in a paper published three years earlier (“The Basis of Objective Judgment in Ethics,” *International Journal of Ethics*, 37 (1927), 121), he seems to have rejected the idea that propositions about *prima facie* rightness elude proof. For simplicity, we may take his more recent view...
While Ross offers less explicit discussion of the factual basis that substantiates our actual duties, we have every reason to think that his opinion on the matter parallels prima facie duties in every respect except for their self-evidence. That is, supposing that our actual duty in any given situation is to do $x$ rather than rather than $y$ or $z$ is a matter of objective truth, that truth is embedded in the fabric of the universe in much the same way as mathematical axioms. Unfortunately for us, we simply cannot grasp that we ought to do $x$ instead of $y$ or $z$ with anywhere near the same level of clarity that we have when we grasp the idea that $x, y$ and $z$ are all prima facie right.

It remains for us to review the list of principles that Ross takes to be so self-evident in the case of prima facie rightness. The complicated relationship between the right and the good ensures that Ross’ list of prima facie duties appears somewhat fragmented, for only in some cases are such duties tied to promoting one of his various goods. As noted above, Ross thought that to have a duty just is to have something one ought to do. (And Ross did not distinguish between a moral and a practical ought.) So he did not feel the need to explain why anybody ought to consider the following list of prima facie duties binding for everyone. The list he offers comprises six items: duties of fidelity and reparation; duties of gratitude; duties of justice; duties of beneficence; duties of self-improvement; and duties of non-maleficence.19

Ross was certain that each item on this list belongs there, but he was not so certain from *The Right and the Good* to trump that earlier position, but ultimately, I defer to scholars of the period on how best to reconcile this discrepancy.

---

19 For convenience, I often refer to prima facie duties simply as “duties” in the following discussion. This in no way is meant to undermine or blur Ross’ distinction between prima facie duties and actual duties.
that the list is exhaustive. “There is [supposed to be] nothing arbitrary” about this group of \textit{prima facie} duties; “each,” he says, “rests on a definite circumstance which cannot seriously be held to be without moral significance.” (Ross, 20) So Ross thought that each of these duties obviously applies to all of us, but he recognized that there may be additional duties that he had left off of the list. He merely “suggests” this taxonomy of \textit{prima facie} duties, “without claiming completeness or finality for it.” (\textit{Ibid})

Still, the diversity of this list is noteworthy. The first group (duties of fidelity and reparation) hinges on previous acts committed by the agent, while the second (duties of gratitude) depends on previous acts committed by those around an agent. The duty of fidelity includes keeping promises and telling the truth. The injunction to keep promises is only activated when we act to make a promise, and the requirement that we be honest, says Ross, arises from the implicit commitment we generally make to telling the truth when we open our mouths to speak. Duties of reparation kick in when we have wronged someone else. Duties of gratitude are structurally similar, but inverted, in that they enter the picture when other people have treated us in particular (good) ways.

Unlike these first two classes of duty, which bear no particular, distinctive relation to any of Ross’ intrinsic goods, Ross simply defines the third category (duties of justice) in terms of his final intrinsic good. Specifically, duties of justice deal with promoting the apportionment of pleasure in accordance with virtue. Ross acknowledges the possibility of distributions of happiness that are not in accordance with “the merit of the persons concerned” and this category of duty is meant to capture the intuition that we should “upset or prevent such a distribution.” (Ross, 21) This seems to be a relatively controversial way of phrasing a description of duties of justice, for it implies that there is
an intuitively obvious duty to interfere when good things happen to bad people just as much as when bad things happen to good people. It also fails explicitly to acknowledge a duty to help bring about good things for good people (if not bad things for bad people). It may be a stretch to say that this particular way of describing duties of justice is intuitively obvious, but perhaps Ross is using loose wording here.

There is a close relationship between the fourth and sixth categories of duty—which are beneficence and non-maleficence, respectively. Both flow from the fact that “there are other beings in the world whose condition we can make better [or worse] in respect of virtue, or of intelligence, or of pleasure.” (Ross, 21) The essential difference is that the beneficence is a positive duty while the non-maleficence is a negative one, a point that Ross is very emphatic about. Non-maleficence is, in fact, Ross’ only truly negative duty; it is the only duty that Ross understands as requiring that we abstain from doing something.

According to Ross, the duty of non-maleficence is also “of more stringent character” than the duty of beneficence. One reason for making non-maleficence stronger than beneficence is to block various consequentialist-friendly sacrifices, such as stealing from one to give to others. By giving non-maleficence greater weight, Ross shows his reliance on the distinction between doing and allowing: this way we can plausibly claim that my duty not to kill someone is stronger than my duty to save him, which generally seems appropriate.

The final category of prima facie duties offered by Ross covers duties of self-improvement.20 We might think of self-improvement as the duty of beneficence as applied

20 Ross labels duties of self-improvement his fifth category (out of six), but because his
to ourselves, since Ross offers a very similar description for the two. As he puts it, "we can improve our own condition in respect of virtue and intelligence." (Ross, 21) But there are two points worth noting about duties of self-improvement. First of all, Ross does not seem to recognize a parallel between non-maleficence towards others and non-maleficence toward ourselves (or more precisely, our own virtue and intelligence). As a matter of symmetry, one might expect him to prohibit harming our virtue and intelligence more stringently than he requires us to promote our virtue and intelligence. The second point is even more striking: Ross entirely omits a duty to promote our own pleasure, even though it is one of our duties of beneficence to help promote pleasure in others. This is no mistake; Ross actually did not think that there is a duty to bring about one’s own pleasure, even though he sometimes recognized duties to do things we would naturally be inclined to do anyway.21 In all likelihood, he simply thought it obvious that there is no such duty.

Another questionable element of Ross’s view is that he built in a surprisingly consequentialist default setting for agents who do not find themselves in the grip of *prima facie* duties.

”...if we are ever under no special obligation such as that of fidelity to a promisee or of gratitude to a benefactor, we ought to do what will produce most good; and... even when we are under a special obligation the tendency of acts to promote general good is one of the main factors in determining whether they are right.” (Ross, 39)

---

21 I think Ross would accept the duty of a mother to look after her child, even though mothers are naturally inclined to look after their children.
This is a remarkable claim to make, for it means that we are always positively obligated to do something. Such a demanding view is unusual for a deontologist, but recall that Ross explicitly acknowledges the possibility that his taxonomy of *prima facie* duties is incomplete. Any counterintuitive results that follow specifically from Ross’ list of *prima facie* duties are of only secondary concern, as the list that yields them is open to revision.

This is one of the major virtues of Ross’ approach: it provides us with precisely the set of prescriptions we want. While the list of duties offered by Ross may omit important duties, or may seem to misclassify some of its component parts, we have already seen that Ross explicitly grants his readers the flexibility to adjust and amend his list. We might very well wish to avail ourselves of that flexibility, as there are a number of plausible candidates missing from Ross’ list. For example, we might add an agent-relative permission (though Ross did not contemplate this possibility), a duty to use force in preventing maleficence, or a duty that preserves the possibility of self-referential altruism. In any event, the liberty to add those items rests with us.

The unfettered freedom to add duties to the list hinges on their being underivative. If Ross had posited a unifying principle or set of principles behind his duties, then any additions to his list of duties would either need to align with that framework, or would require modifying the underlying framework to accommodate our additions. But this freedom comes at a significant cost. What is perhaps the most significant problem with Ross’ approach falls out of the fact that he posits no unifying principle(s) behind his list of duties, in particular because that makes his view vulnerable to some of the very accusations that he levels against other normative ethical views.

A major reason why Ross favored underivative duties was a worry that offering an
explanatory back-story for our moral obligations would distort the nature of those obligations. Let us return to the example of promise-breaking, which we have used throughout this chapter. Ross did not seem to think that it was irrational to break a promise (as a Kantian might maintain), nor did he think it appropriate that I should keep promises only to promote my own flourishing (as a neo-Thomist might say). The first case at least seems to leave out the distinctive feature of acting wrongly (discussed above), while the second threatens to make acting well a matter of self-interest, which seems counterintuitive.

These points are well taken, though they surely deserve the more thorough treatment we will give them in subsequent chapters. But even in granting Ross’ point for the moment, we must recognize its limited scope: that is, even if several normative theories threaten to distort the nature of our moral duties by unifying them with an explanatory back-story, it does not follow that all possible reductive accounts will yield that same result. In fact, as we shall see, there are multiple reasons why we might favor recalibrating our reductive explanation rather than discarding it altogether.

This point is all the more important because Ross’ own description seems to under-describe our moral duties, leaving out their substantive content, and thus distorting them in its own peculiar way. To bring this out, let us zoom in on the constraints that Ross would be able to offer us. If I have promised to meet my friend for lunch, I activate a (defeasible) constraint that prevents me from missing the meeting. I probably cannot skip the lunch at the last minute (say) to help a random child with his homework, even if doing so would bring about more good overall.\(^{22}\) I must contend with a *prima facie* duty not to

\(^{22}\) We have to qualify this claim with “probably” because other situational factors might
break my promise. If I miss the meeting, I have wronged my friend (in some sense), even if I ought to have missed the meeting, all things considered. I may even owe him reparations as a result—again, notwithstanding the possibility that I acted appropriately in skipping the meeting. But Ross accepts no reductive explanation of these two duties—or any others, for that matter—for he views them instead as fundamentally disparate.

When I ask what it is that makes me in certain cases sure that I have a \textit{prima facie} duty to do so and so, I find that it lies in the fact that I have made a promise; when I ask the same question in another case, I find the answer lies in the fact that I have done a wrong. And if on reflection I find (as I think I do) that neither of these reasons is reducible to the other, I must not on any \textit{a priori} ground assume that such a reduction is possible. (Ross, 24)

Ross could be right that we should be careful in making \textit{a priori} assumptions about the reducibility of claims about the right, even if he overstates the impermissibility of optimism here. He offers this account up contra the perspectives taken by the Kantian and consequentialist (and maybe even the neo-Thomist); and he is partial to this description because he thinks it captures what is essential about our duties. Yet it seems like there is more we could say about promise-breaking (for example) than that it is just plain \textit{prima facie} wrong. For example, at the very least we could say that breaking promises seems to be a way of mistreating others who have a claim to certain behavior on our part. It does not just \textit{seem wrong} to break promises, full stop; it \textit{seems wrong because it involves treating others in an intuitively objectionable way}. Filling out this description does not require shifting from underivative duties to derivative ones, for whatever fuller description we actually make it more appropriate for me to skip the lunch in favor of helping a random child with his homework.
settle on could also be underivative. But the objection as I intend it isn’t just about the brevity of Ross’ description; rather, it’s about the fact that it is underivative.

All of the duties on Ross’ list seem quite clearly to dictate particular brands of treatment for those around us. More specifically, the *prima facie* rightness and wrongness of various acts seem quite closely to track (something like) the legitimate preferences of moral agents, or actors, or what have you. Ross was surely aware of this, so it is unclear why he didn’t make a more concerted effort to unify his list of duties (beyond listing a number of duties that aim at promoting the good).

It does not seem like we need to make any *a priori* assumptions to think that such a reduction is possible, either. We can locate *a posteriori* optimism about such a reduction merely from looking at the similarities among his various duties. The duties of beneficence and non-maleficence are especially striking in this regard. They are symmetrical, prescribing a consistent and predictable response to Ross’ list of goods and evils. That the only thing separating them is a matter of valence and a vague stipulation of their varying stringency invites the obvious question of why Ross would not bother to attempt to reduce them to some common base. A more comprehensive reduction may not be simple, but it certainly appears possible. The mere fact (if it is one) that we clearly intuit certain principles of right action does not mean there is no further baseline reason unifying the principles. In any event, hope for a reductive explanation need not depend on an *a priori* assumption.

Even if we are skeptical about unifying our duties, there is reason for thinking that we can make them underivative.\(^{23}\) One of the biggest selling points of consequentialism is

\(^{23}\) It is possible, I think, to do the latter without doing the former, though it is likely that the
that, while it offers us only a single, underivative duty, that duty is especially compelling. If something is good, then all else equal, more of it is better. That conclusion counts directly in favor of a duty to maximize the good. The consequentialist story makes such a duty underivative—consequentialists have nothing to say, after all, except that we ought to maximize the good because doing so is better than not; but it is reasonable to think that we have reached bedrock here. The force of questions about why we ought to promote the good is limited because, even if we ultimately reject consequentialism, there is an obvious and strong consideration in favor of complying. Ross himself taps into the force of this consideration by allowing the consequentialist principle to dictate what we do in a significant range of cases.

But duties that require doing something other than promoting the good are different. There are legitimate questions about why on earth we would want to permit the bringing about of suboptimal states of affairs (as we do when we give agents options), let alone why we would want to demand as much (as we do when we impose constraints). On Ross’ view, answers to those questions are not forthcoming. He essentially tells us that there are no answers; there are just certain duties, such as the (prima facie) duty to keep promises, that sometimes prevent us from bringing about the most good.

It feels, however, as if we need to drill deeper, and it feels that way not just as a matter of intuition, but also as a matter of precedent: we are familiar with the compelling underivative duty offered by consequentialists. It is hardly unfair to wonder why there should be such an asymmetry, and to pursue an explanation for deontological duties that two go hand in hand.
restores the balance.

Conclusion

We began this chapter by outlining the intuitionist approach to building a normative ethical theory. After reviewing the historical context in which intuitionism arose, we focused in on one particular instantiation of that approach by walking through a detailed exposition of W.D. Ross’ pluralist view. We concluded the chapter by considering some merits and demerits of Ross’ approach. In particular, we noted the value in preserving a distinction between prima facie and actual duties, which makes conceptual space for a commonsense interpretation of how we view many moral dilemmas. At the same time, however, there are serious questions about the Rossian appeal to underivative duties. Not only is it somewhat counterintuitive to think that our duties are underivative, but it seems like even Ross’ underivative list comprises a set of duties with a strikingly common core.

It is far too early to say what we may conclude from this result. If the other available approaches suffer from even worse troubles, then Ross’ method may be our best bet. And we cannot make the requisite comparison until we outline other deontological approaches and assess their viability. With that in mind, we will now shift our attention to the two other prominent deontologies: Kantianism and Neo-Thomism. After evaluating each, we will take stock of our situation and see how to proceed. Ideally, we will be able to build a new, more successful approach out of the lessons we have learned from the other views. In the meantime, we should merely note the strengths and weaknesses of Ross’ approach as we turn our immediate attention to Kantianism.
As we move forward, it may be useful to keep in mind a schema for relating the different deontologies both to consequentialism and to one another. Consequentialists think that all duties are grounded in claims about value, that value inheres in states of affairs, and that the proper response to value should always be simply to promote it. All of our deontologies (intuitionism, Kantianism, and Neo-Thomism) reject the last of these claims, but only Ross rejects the first as well. As we have seen, a number of Ross’ duties rely on relationships agents share with one another rather than on ascriptions of (or calculations about) value. The Rossian rationale for (say) keeping a promise when more good would come of breaking it has nothing to do with stipulating an attitude we must take towards value; it is, instead, simply that keeping the promise in the given instance is more of a duty than breaking it for the greater good. Framed in this light, the only consequentialist claim adopted by Ross is the second one: that value resides in states of affairs.

Thus, to the extent that we see the triad of consequentialist claims as sequential steps towards a normative theory, Ross’ deontology reflects the most significant departure from consequentialism of the theories that we will canvass. Ross rejects the first of these essential steps, parting ways with consequentialist reasoning at the earliest of three forks in the road. At least one criticism of Ross that I have raised fits neatly into this schema.

Recall the final point from the preceding section, which raised the troubling asymmetry between duties that are defined in terms of value and duties that are not. The doubts that

\[24\] I owe this useful way of mapping deontological approaches to Tom Hurka.

\[25\] Practically speaking, of course, Ross’ view resembles consequentialism to a much greater extent than standard interpretations of Kantianism.
such an asymmetry might evoke follow directly from Ross’ rejection of the first of the
three consequentialist claims. In denying that all duties are defined in terms of value, Ross
opens himself up to a nagging skepticism that can’t be talked away. Both Kantianism and
Neo-Thomism accept that our duties are defined in terms of value, however, and are thus
immune to one particularly nettlesome objection. But other objections that afflict the
remaining views track their acceptance of a particular permutation of the three
consequentialist claims, so this schema will become increasingly useful as we proceed.
2

Kantianism

Introduction

For better or for worse, Kant’s moral philosophy has come to represent the quintessential deontology. The complexity of Kant’s view and the breadth of scholarship on his work make it difficult to write a well-contained but complete assessment of his ethical theory. Even today, there are innumerable interpretations of the various facets of his position. Moreover, Kant’s views are comprehensive—offering a putatively coherent metaphysics, ethics, epistemology, philosophy of law and more. There are debates again about how heavily each portion of Kant’s general view rests on the other portions, but some significant interrelation seems beyond dispute.

My interest lies quite narrowly in evaluating how well Kant’s view of the right fits with our moral intuitions, both about what our duties ought to be and why we have those duties at all. More specifically, I want to explore the general plausibility of Kant’s account of agent-centered restrictions, as well as the assumptions that underlie that account. Fortunately, these topics can be isolated and addressed quite distinctly for the purposes of my investigation. We need not concern ourselves with the worry that a proper understanding of Kant’s views on agent-centered restrictions depends on familiarity with
his positions on seemingly unrelated questions. The implications of his views on the right are clear enough as is.

I will proceed by offering a sketch of the relevant portion of Kant’s ethics, drawing as directly as possible on direct citations from the Groundwork, and supplementing the discussion with the interpretations of leading Kant scholars. I will inject critical commentary throughout, highlighting what I take to be deficiencies in Kant’s approach, and I will conclude that the Kantian system does not succeed in offering us a compelling justification for constraints.

2.1. Kant’s Theory of the Right

Kant claims that we get our duties from “the moral law,” but to understand what that means, we need to take a brief tour of Kant’s views on the metaphysics of agency. Kant regarded all human action as purposive, so he thought of every action as grounded in volition. (4:427) His view is that, unlike non-rational wills (such as those represented in most animals), the rational will can act in accordance with a principle (or set of principles). Animals that lack rationality act only from inclination—the visceral and unsophisticated impulse to satisfy biological wants or needs. But man’s capacity for purposive action makes us candidates for possession of a good will, for we can choose to act from the right principle (which, for Kant, was acting from duty). Of course, we may choose the wrong principles, but it is our ability to select our principle(s) and act accordingly that makes us subject to the moral law.

Purposive action, on Kant’s view, is guided by a principle that is given by the will—specifically, we might say, a principle that takes a form something like, “I will do Action-A
in order to achieve Purpose-P.”26 (Korsgaard, 58) This principle is a maxim, a “subjective principle of volition.” (4:401) Of course, people are not perfectly rational. Like animals, we have natural inclinations. The difference between people and non-rational beings (like animals) is that people have the power to choose against those inclinations in favor of behaving on principled grounds instead.27

A perfectly good will—one that does not have to contend with natural inclinations—would not be pulled in different directions by the conflicting draw of duty and inclination. Since its maxims would not be polluted by inclination, it would naturally choose well all of the time. Therefore, it would not experience the moral law as a constraint at all.

“A perfectly good will would, therefore, equally stand under objective laws (of the good), but it could not on this account be represented as necessitated to act in conformity with law since of itself, by its subjective constitution, it can be determined only through the representation of the good.... Therefore imperatives are only formulae expressing the relation of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, for example, of the human will.” (4:414)

The sort of agent Kant has in mind here is highly idealized, and by Kant’s own lights, may not even exist. In particular, Kant has in mind God or maybe some other divine being. But

26 Thomas Hill, Jr. offers a slightly more specific formula for a maxim: “In conditions C, I shall do X in order to E from motive M.” (Hill 2006, 482) Hill acknowledges that Kant does not hew consistently to one format with his model maxims, and that some of Kant’s maxims have the less specific structure identified by Korsgaard.

27 “…[T]he pure thought of duty and in general of the moral law, mixed with no foreign addition of empirical inducements, has by way of reason alone... an influence on the human heart so much more powerful than all other incentives....” (4:410-411)
the possibility that there actually is no such thing as a perfectly good will is not a problem because the mere conception of a perfectly good will is sufficient for heuristic purposes. And all we need to note here is that a perfectly good will would act in accordance with duty without experiencing duty as a constraint. Unfortunately for us, as humans we are prone to choosing bad maxims. Our subjective principles of volition might involve acting to satisfy our inclinations, which, according to Kant, we tie up with our own happiness. Because choosing well for us involves choosing against the grain of inclination, Kant claims that we experience acting from duty—that is, acting from the moral law, an objective principle of the will—as a constraint.

The idea that an objective principle of the will lays down our duties is not unique; it is perfectly consistent with consequentialism, for example, even if consequentialists might typically use different terminology to describe it. The distinctive element—the area that we should focus on—is Kant’s view of what that objective principle must look like, and what commands follow from it. Kant selected as his objective principle what he calls the “Categorical Imperative,” which is essentially a maxim that contains no subjective content and could thus extend to all agents in all circumstances.  

This principle, according to Kant, is rationally binding on all of us, and it is supposed to yield rigidly deontological prescriptions.

Kant offers several formulations of the Categorical Imperative, though it is

28 It isn’t clear that this must be true. If our inclinations are sympathetic, then we probably will not experience acting from duty as a constraint.

29 There is only one “Categorical Imperative”—but there are numerous “lower case” categorical imperatives, each representing an action as good in itself, and therefore as necessary for a will aiming to conform with reason.
conventional to boil the list down to four: The Formula of Universal Law, the Formula of Humanity, the Formula of Autonomy, and the Formula of the Kingdom of Ends. The latter two of these are typically regarded as closely related, so we will treat them together as one. Much energy has been spent in trying to capture how the various formulations of the Categorical Imperative are meant to relate to one another. The standard Kantian view is that the formulations are coextensive in their prescriptions but distinctively illuminating with respect to Kant’s broader underlying picture. The real trick has been to arrive at a plausible understanding of each formulation that syncs up with the others in what it prescribes. This is hardly the place to take another stab at such a project. Instead, we shall take the various formulations one at a time in an effort to see if any of them might yield an appealing account of constraints. If none can offer such an account, then how the formulations sync up will not be of any significance.

It will thus be clearest if we take the formulations in turn, applying each to an example or two in order to see how the various versions of the Categorical Imperative might yield Kantian deontology. Quite naturally, then, the most useful examples for our purposes are duties that are unequivocally deontological—that is, those that cannot be explained in any directly consequentialist manner. (Perfect duties are thus preferable to imperfect duties.) I will focus on the duty not to commit suicide and the duty to abstain from making false promises, and I shall take the latter first.

2.2. The Formula of Universal Law

The first formulation of the Categorical Imperative given by Kant is the Formula of Universal Law, a popular version of which states, “I ought never to act except in such a way
that I could also will that my maxim should become a universal law.” (4:402, emphasis in original) Kant offers several other versions of this formulation. The second reads as follows: “Act only on that maxim by which you can at the same time will that it should become a universal law.” (4:421) The third is: “Act as though the maxim of your action were to become by your will a universal law of nature.” (4:421) Traditionally, the first two formulations are thought to express the “formula of the universal law,” while the third expresses “the formula of the universal law of nature.” (Hill 2006, 481, emphasis mine) The reason for labeling the third formulation differently is that it introduces the term “law of nature” into the mix, which arguably adds some substantive content to the formula it expresses. Yet the common view seems to be that, despite these variations, the principles effectively amount to the same thing. Following convention, we will treat them as equivalent.30

To illustrate the link between the Categorical Imperative and our various duties, Kant uses the Formula of Universal Law (and later, the Formula of Humanity—see below) explicitly to derive a series of duties, including the duty not to make false promises. Kant gives us the example of someone who badly needs money, and in order to acquire some, secures a loan by making a phony promise to repay what he has borrowed. In Kant’s example, the borrower is not attempting to steal for some trivial purpose; he believes himself to be in need, and he takes the steps necessary to fulfill that need. According to

30 Hill notes that there are other minor variations in the language Kant used to express the formula of universal law and formula of universal law of nature throughout his ethical writings—differences not captured in the three quoted expressions above. But for his part, Hill claims that these variations are incidental and not intended to change the content of the principles.
Kant, the problem with the borrower’s plan (in terms of the Formula of Universal Law) is that the maxim it relies on cannot be universalized in the right way:

“...the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses.” (4:422)

Before we can evaluate this passage, it is important to get a sense of how a maxim can fail to be universalizable. Kant was clear that there are two ways in which this can happen: if a maxim cannot, without contradiction, be conceived of as universal law, or if a maxim cannot be willed as a universal law. (Kant, 4:424) In accordance with the Formula of Universal Law, we may act on any maxim that can be both conceived and willed as a universal law, and we may not act on any maxim that fails either test. Very clearly, then, this pair of tests provides for negative duties; if you test a maxim, and it fails, then you oughtn’t to act on it. According to Kant, such is the case with making false promises. But this pair of requirements can yield positive duties too. (Hill 2006, 483) To confirm that a given permissible maxim amounts to a duty, all we would have to do is consider what our new maxim would be if we failed to comply with the old one. If the new maxim failed either of Kant’s requirements, then the initial maxim would express a positive duty.

There is much debate about what it means to say that a maxim cannot be conceived as a universal law, but there is one uncontroversial starting point. “[I]f everyone’s acting (or being permitted to act) on a given maxim is logically impossible, then the maxim cannot be conceived as a universal law in the relevant sense.” (Hill 2006, 484) Surely this much is right; if universalizing a maxim is logically impossible, then it cannot be conceived
of as universal law in any sense. While this is not the only interpretation of the conceivability test, any plausible alternatives would have to contain this version too. We will deal with some of the more robust interpretations below.

There is a similar debate about what it means to will a maxim as universal law. Once again, Hill captures the relevant starting point: if a maxim cannot be conceived of as a universal law, then surely it cannot be willed as one either. For the willing test to impose any of its own requirements, however, we obviously need it to do more. Hill notes that there are quite a number of views on what the willing test entails, but we will address only the most prominent in what follows.

Returning to Kant’s passage on making false promises, we see that Kant has attributed a particular maxim to the person who would make a lying promise:

“…everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping.” (Kant, 4:422) The rationale given is that “the universality of a law… [based on that maxim] is impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses.” (Ibid)

At first glance, Kant’s explanation seems to be that the maxim in question fails the willing test. After all, it appears that we can conceive of a world in which everyone adopts Kant’s maxim, since the reason for rejecting it seems to hinge on something unappealing about the state of such a world.

Yet a number of philosophers have argued that this line of thought is actually mistaken.31 (Wood 1999, 88) It is common to treat maxims that involve the violation of

31 These include Allen Wood, Barbara Herman (140), Onora O’Neill, and Jonathan Harrison, among others.
perfect duties as failing the conception test, while maxims that involve violating imperfect duties as failing the willing test. In Kant's Ethical Thought, Wood illustrates such an approach by homing in on one specific line, where Kant claims that universalizing the false promiser's maxim would “make the promise... itself impossible.” (Kant, 4:422) Wood notes that if we take Kant’s language at face value, and understand him to be saying merely that nobody would bother to make promises in a world where everyone feels free to make lying promises whenever it suits their purposes, then promises would not, strictly speaking, be impossible. (Wood 1999, 88)

Instead, Wood suggests an alternative understanding of how universalizing the false promiser’s maxim might be logically impossible: “Kant might... have meant to claim that a real or genuine promise... can occur only when it is possible for the promiser to have a reasonable ground to think that the promise will be believed.” (Wood 1999, 88, emphasis mine) If we universalize the false promiser’s maxim, then Wood’s necessary condition would be absent; and while that would not make the notion of promising contradictory, it might “render it impossible under the circumstances to engage in behavior that either the agent or others could regard as an act of promising.”32 (Ibid)

Wood’s line on the case of false promising represents an example of a stronger interpretation of the conceiving test than the minimalist approach described above. On Wood’s view, universalizing the promiser’s maxim is only impossible if we assume some

32 This is very much the sort of interpretation given by both John Kemp and Barbara Herman. (See “Kant’s Examples of the Categorical Imperative,” p. 238 and The Practice of Moral Judgment, p. 137, respectively.)
contingent facts; it’s not contradictory in an analytic sense.\(^{33}\) There is a problem that I don’t think Wood’s account can handle, which we will discuss below; but his angle on this passage alone is at least somewhat compelling, even though it relies on a weaker sense of inconceivability than we might prefer. The only glaring issue is one that Wood himself recognizes. The more plausible interpretation of Kant’s objection in this case is one that emphasizes the *practical* challenges of universalizing the maxim rather than the logical ones, for such an interpretation seems to align much more closely with the passage as a whole. (Wood 1999, 89)

In *Acting on Principle* and *Constructions of Reason*, Onora O’Neill likewise claims that the maxim behind false promising cannot be conceived as universal law.\(^{34}\) According to her, “[t]he project of deceit requires a world with sufficient trust for deceivers to get others to believe them; the results of universal deception would be a world in which such trust was lacking, and the deceiver’s project was impossible.” (O’Neill, 132) This description is rather broad, for it does not specify whether the contradiction is a logical matter or a practical one. In her earlier book, O’Neill seems to favor the latter interpretation, which distinguishes her stance from Wood’s. There she claims that universalizing the forbidden maxim would result in making “successful deception... an increasingly unlikely result of false promising, since public confidence would diminish and

\(^{33}\) This also seems to be the view of Bruce Aune, as presented in *Kant’s Theory of Morals*, though he does not distinguish between the conception test and the willing test. (See, specifically, pp. 53-54.)

\(^{34}\) O’Neill wrote the earlier of these two books, *Acting on Principle*, under the last name “Nell.” For clarity, then, I will denote citations from *Constructions of Reason* with last name “O’Neill” and citations from *Acting on Principle* with the last name “Nell.”
eventually vanish.” (Nell, 78) The reason the maxim fails the conception test even though the problem with universalizing it seems to be a practical matter is that she advocates placing an emphasis on the agent’s intention rather than on the specific structure of the maxim.\(^{35}\) For her, even conceiving of a maxim as universal law means intending it as universal law and intending the means sufficient to it. O’Neill’s approach is distinctive, and it certainly takes some liberties with Kant’s language. If we let it stand, however, it seems to handle the false promising maxim just fine.

In *Creating the Kingdom of Ends*, Christine Korsgaard develops an interpretation that she appropriately calls the “practical contradiction interpretation.”\(^{36}\) On her interpretation, the maxim supporting false promises can be conceived as a universal law but cannot be *willed* as a universal law. (Korsgaard, 96) Korsgaard thus deviates from the standard practice of treating maxims that run contrary to perfect duties as failing the conception test—a convention employed by O’Neill and Harrison, among others. (Harrison, 209; O’Neill, 132-133) Korsgaard claims that, in order to fail the willing test, a “maxim would [have to] be self-defeating if universalized: your action would become ineffectual for the achievement of your purpose if everyone (tried to) use it for that

\(^{35}\) Cf. Barbara Herman’s *The Practice of Moral Judgment*, specifically page 142, footnote 12.

\(^{36}\) We might think that objections to the logic of universalizing a maxim pair strictly with a maxim’s failure to pass the conceivability test, whereas objections to the practicality of universalizing a maxim pair specifically with a maxim’s failure to pass the willing test. After all, Wood’s view exemplifies the first sort of view while Korsgaard’s represents the second. But Wood suggests a third interpretation of the passage on false promises on which he claims Kant’s objection is one of practicality even though the maxim fails the conceiving test. (Wood, 88) Because he doesn’t defend this interpretation, and because it would also be vulnerable to the objection I raise below, we may leave it aside; but it is worth bringing up for the purposes of clarifying the relationship between the types of objections and types of tests that arise when discussing the Formula of Universal Law.
purpose.” (Korsgaard, 78) A direct reading of the passage certainly suggests such a view, seemingly claiming that the consequences of everyone adopting the false promiser’s maxim would be to render his action ineffectual. On Korsgaard’s interpretation, just as on Wood’s and O’Neill’s, it seems right for Kant to claim that his maxim fails.

It is certainly plausible that the convention of promising would dissipate if everyone made false promises whenever they pleased, or even that it is contradictory to think that promises would be possible in such a world. But surely Kant’s maxim is uncharitable. Someone like Ross, for example, might agree with Kant that we have a duty to keep promises even when some good stands to be gained by breaking them, but might also think, unlike Kant, that if there is a lot of good to be gained (or a lot of bad to be prevented) by breaking a promise, then we ought to break it. Someone with that view might also think that we ought to make false promises if doing so brings about sufficient good or averts sufficient evil. The maxim behind such a policy would not need to be as strong as the one given in the passage above. Ross would not say that we may promise “whatever [we] please with the intention of not keeping [our promise]”; rather, he would say that we should make promises with the intention of keeping them unless especially rare and dire circumstances arise. A maxim of this sort would not undermine the convention of promising in any significant way, for we can stipulate that the assurance that comes with a promise might drop from 100% to 99.9%. That’s hardly enough to erode people’s confidence in the convention of promising until it completely washes away (as O’Neill or Korsgaard seem to suggest), nor is it enough to render the act of promising

37 I’m grateful to Tom Hurka for helping me to see this point.
logically impossible (as on Wood's interpretation).  

We could be generous and accept this Rossian outcome as superior to the rigid prohibition on false promising that one typically associates with Kantian ethics. The Formula of Universal Law, we might think, can yield better results than Kant himself believed. But, of course, this trick can work against Kant too. I could adopt a maxim that I make false promises whenever I need to do so in order to see Halley's Comet. Even if everyone adopted such a maxim, the effect on the convention of promising would be similar to the effect from the previous example (which is to say, limited). Halley's Comet is visible about a dozen times per millennium, so the vast, vast majority of the time, promises would remain intact.

It is common for those writing in defense of Kant to try to disqualify more specific maxims in putative counterexamples to the Formula of Universal Law at least partly, one suspects, because allowing them would prove fatal. A good number of these attempts ring hollow, especially when the maxim in question meets one of the standard formats given by Kant (for he gives more than one). While I think there are counterexamples available to us that don't rely on specific maxims (see below), that fact alone is no reason to pass entirely

38 It seems to me that it would take much more than a .1% drop in the strength of promises to destroy the convention of promising, but we can illustrate the point with a conservative example.

39 There will, of course, be those who think the Rossian verdict is inferior, but I am not among them.

40 Note that this maxim meets the structural requirements given by Korsgaard: “I will do Action-A in order to achieve Purpose-P.” (Korsgaard, 58) We would only need to make it more specific to meet Hill's structural requirements: “In conditions C, I shall do X in order to E from motive M.” (Hill, 482) Our revised maxim could be: in periods when Halley's Comet is visible from earth, I will make false promises if they serve my purpose in seeing the comet because seeing such a rare event is more important than being truthful.
on those that do. O’Neill gives what strikes me as one of the most honest assessments of
the problem that specific maxims cause for Kant. Roughly, she suggests that the specificity
of the maxim becomes artificial to the extent that the details that make it specific are
unnecessary for shaping the agent’s plan of action. For example, she gives the case of
someone with a particular hair color who wishes to rob a bank located in a particular
location vis-à-vis his house, at a particular time of day. (Nell, 72) She thinks that a good
faith assessment of a would-be bank robber’s maxim would show that these details do not
add any substance; he would likely rob the bank were his hair a different color, or were it
located in a slightly different place, or were it a different time of day. She adds, “[o]n the
whole, I believe, agents who are honest will not claim to have highly specific maxims.”
(Ibid)

There is certainly some plausibility in O’Neill’s assessment. It is probably correct to
say that the maxim she uses as an example does not count against the Formula of
Universal Law in any compelling way. O’Neill’s view would block a large subset of the
possible maxims one might use to challenge the Formula of Universal Law. But, of course,
it will not block them all. There is a big difference between what is true “on the whole” and
what is true all of the time, and I think O’Neill is correct to limit the range of her rejoinder
to the former in response to the challenge of specific maxims. The Halley’s Comet maxim
seems to overcome her threshold of irrelevant specificity, and with minor tweaks, we
could make it slightly broader and even less controversially acceptable.

As noted above, though, we could avoid quibbles about the specificity of this maxim
altogether. Suppose, instead, that I adopt a maxim that I will make false promises
whenever I stand to benefit to some particularly high degree by doing so. If we set the
threshold high enough for how much good could justify false promising, then the
convention of promising will remain intact, as will the logical coherence of adopting the
maxim. None of the approaches we have canvassed seem to block either of these
problematic maxims, even though we would think it wrong to make false promises in
accordance with either policy.

Barbara Herman offers an interpretation of the Formula of Universal Law that may
dodge this category of problem. She proposes thinking of the Categorical Imperative
procedure “as applying not to actual maxims of action but to a type of action-justification
pair: to do x-type action for y-type reason.” (Herman, 147) These pairs, which Herman
calls “generic maxims,” do not represent the willing of any specific agent; instead, they
function at a more general level, “present[ing] a pattern of justification... in the form of a
possible principle of willing.” (Ibid) On Herman’s view, instead of plugging the maxim of
any given actor into the universalization procedure, we plug in a much broader
description of the sort of act in contemplation and the sort of reason behind it. For
example, instead of using Kant’s false-promising maxim, we would assess the pair “deceive
someone for personal gain.”

The immediate effect of switching to Herman’s approach is that maxims with
pronounced specificity will no longer trip up the Formula of Universal Law. Following this
strategy, we would abstract from the details from any maxim and shift, instead, to using a
generic action-reason pair. Presumably, then, Kant’s false-promising maxim would fall
under the same pair as my Halley’s Comet maxim (again, deceiving someone for personal
gain). Quite plainly, pairs of this sort are vague—too vague for us to know whether they
can be universalized without logical or practical difficulties. Herman notes this as well,
claiming that since “actual maxims are not the input... neither are duties its output.” (Herman 148) Herman suggests that we think of the output as a “deliberative principle in the form of a presumption,” thereby avoiding “the derivation-of-duties problem of rigorism.” (Ibid) This presumption is “open to rebuttal in a given case... [if the agent's] circumstances deviate in a morally significant way from those specified by [the relevant] moral principle. (Ibid)

A presumption against certain acts, if sufficiently strong, would likely generate some sort of defeasible agent-centered restriction. For that reason, Herman’s approach is of interest. What is unclear is how Herman’s procedure produces a presumption of any sort at all. She suggests that we engage in moral deliberation primarily in cases where “an agent perceives her circumstances as exceptional or as containing conflicting moral considerations and directives.” (Herman, 146) In those cases, we may need to invoke a generic maxim to figure out what we ought to do. Yet the only way in which using a generic maxim could be helpful would be if we already had reason to think that, for example, deceit for personal advantage is wrong. Herman's procedure does not tell us *that*; it only offers a schema for framing an analysis of our proposed course of action, “assum[ing]... moral agents with knowledge of deliberative principles.” (Herman, 148)

Herman’s strategy does not seem to derive directly from a specific reading of Kant. It appears rather that Herman is introducing her own variation on Kant’s Categorical Imperative decision procedure. Whatever the merits of this innovation, it surely will not do for our purposes. The entire point of assessing various ways of understanding the Formula of Universal Law is to see if any version of the concordant procedure can give us a proper account of agent-centered restrictions. If Herman assumes that agents *already*
know deliberative principles, then the procedure she describes is of no use to us. It is simply an aid for those who have already, somehow, discerned moral principles.

If the foregoing analysis shows anything, it is that there are legitimate questions about whether even the most prominent treatments of the Formula of Universal Law can give us a proper account of when we can and cannot make false promises. It is yet a further question whether the Formula of Universal Law can tell us why we oughtn’t to make false promises. For the moment, however, let us assume that our worries about at least one of the previous accounts are misplaced, and that the Formula of Universal Law does (among other things) impose a suitable ban on false promises. Let us revisit Kant’s passage on false promising one more time:

“...the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses.” (4:422)

There are two salient ways of reading this passage. Either what makes false promises wrong is that the maxims that underlie them are not properly universalizable, or the way we know that it is wrong to make false promises is that the maxims that underlie them are not properly universalizable. In other words, we can choose between taking the given passage as a description of what’s wrong with false promises, or as a neutral description of a heuristic device that tells us what is permissible and what is not.

The first of these possibilities is surely unsatisfactory. What can the Formula of
Universal Law tell us about why false promising is wrong? We may rule out one implausible interpretation immediately: the fact that I do x once does not entail—and will almost never lead to—everyone doing x all of the time, so it cannot be that we ought legitimately to worry that everyone would constantly do x. It seems more likely that the Formula of Universal Law expresses a fairness condition. If my maxim could not be adopted by everyone (either because it cannot be conceived as being employed universally or because it could not in practice be universalized), then arguably it is unfair in some sense.

But that in itself is not necessarily so objectionable. Barbara Herman notes that there are any number of times when we might aim to do something because others aren’t doing it, yet the fact that my aim in doing so would be frustrated if everyone did it doesn’t mean that what I’m doing is wrong. She gives the example of going to the movies at six to avoid the crowd at eight, or choosing a travel route that is easier to traverse because fewer others choose it. (Herman, 139) The problem with false promising is not just that we can’t all do it all of the time, especially considering the possibility of using a revised maxim that permits rare cases of false promising for everyone.

Herman claims that the Formula of Universal Law captures the wrongness of false promising if we interpret the false promiser’s maxim as being logically contradictory rather than practically contradictory. (Herman, 140-41) Herman makes this claim in the context of defending the former position over the latter. According to her, if we claim that the false promiser’s maxim is logically contradictory, we can distinguish between cases of “free riding” (as in false promising) and cases of “coordination” (as in going to a matinee movie). (Herman, 139) As we discussed above, Herman thinks that the false promiser’s
maxim fails because if everyone knew his maxim, then his purpose would be (logically) defeated. The same is not true of the movie case, she says, for people would not adjust their behavior if they knew that their preference for later shows affords me the chance to sit in an emptier theater earlier in the day. Thus, she says, the logical contradiction reading actually lets us explain the wrongness of false promising by highlighting the deception involved—something that entirely eludes the practical contradiction interpretation.

I am perfectly happy to accept Herman’s criticism of the practical contradiction interpretation, but she seems to overstate the conclusions we can draw from her own position. The logical contradiction reading only tracks the necessity of others’ ignorance of the operative maxim, and that is not the same thing as deception. The false promising maxim would be logically contradictory on Herman’s view no matter how the victim of the false promise finds out about the promiser’s plans. While the false promiser might need to intend deception for his plan to work, the Formula of Universal Law would give the exact same verdict on any example where the ignorance of others is key to the success of a given aim, whether morally objectionable or not. Suppose I adopt a maxim to throw a surprise party for each of my parents when they retire. (In some circumstances deception might be necessary for throwing a surprise party, but that is not always the case.) It would be logically impossible for everyone to adopt that maxim, for retirement parties would cease to surprise anybody. Thus, on Herman’s view both the maxim behind false promising and the maxim behind surprise parties fail the logical conceivability test, and do so in the exact same way. Yet not only should the party maxim be morally acceptable, but more importantly for present purposes, deception is not required to put the two cases on equal footing.
Even if the logical contradiction interpretation could track deception specifically—and I don’t see how it could—*that* is not the only thing that makes lying promises wrong. Making a false promise is not, for moral purposes, identical to misleading someone in general; it is morally objectionable (when it is) because of the way it treats the promisee’s claim on us, or perhaps because of the way it harms him. There are debates about how best to understand this mistreatment or harm, argued roughly as a proxy dispute between will theorists and interest theorists about rights, but the intuitively appropriate account is virtually assured to lie along either prong of this fork, and the Formula of Universal Law comes nowhere near either mark. The case of false promising alone strongly suggests that the Formula of Universal Law neither imposes on us the proper duties, nor explains what makes certain actions wrong.\(^{41}\)

Before moving on to the Formula of Humanity, however, it is worth touching on the second example of a perfect duty that Kant attempts to derive from the Formula of Universal Law: the prohibition on suicide. While many authors writing on Kant seek to defend some interpretation of Kant’s account of false promising, the opposite is true of his discussion of suicide, with a number of authors attempting to explain it away. As a result, there are a number of ready-made explanations for why an exposition of what is wrong with Kant’s account of suicide fails to indict Kant’s overall approach. Even so, there is some meaningful information to be gleaned from a brief discussion of the Kant’s application of the Formula of Universal Law to the case of suicide.

\(^{41}\) Korsgaard seems sympathetic to this conclusion, writing that the Formula of Universal Law is to be used in decision-making, while the other formulations give the moral law a more intuitive feel.
According to Kant, the problem with committing suicide is that “a nature whose law it would be to destroy life itself by means of the same feeling whose destination is to impel toward the furtherance of life would contradict itself and would therefore not subsist as nature; thus that maxim could not possibly be a law of nature...” (Kant, 4:422) Yet it seems both logically and practically possible for everyone to adopt this maxim. At the very least, there is no obvious reason why it would be impossible for each of us to kill ourselves if we thought ourselves better off dead. That is a rather large problem because Kant offers the prohibition on suicide as an example of a perfect duty, and such duties are supposed to be protected by a very strong sense of contradiction in the concordant maxim. Instead, the salient interpretation of this passage is teleological, which under the best of circumstances would leave us with a weak contradiction (if we could even call it a “contradiction”). As Jonathan Harrison puts it, what Kant claims in this passage is “not that the statement that everybody acts on this maxim is self-contradictory, but that it contradicts another statement, a statement of fact, namely that the purpose of self-love ‘is to stimulate the furtherance of life.’” (Harrison, 215)

Whatever Kant means by “self-love,” appealing to its natural purpose to uphold a prohibition on suicide is controversial on many levels. For this move to work, it must be true both that there is a natural phenomenon whose purpose is to keep us alive, and that we oughtn’t to act against the natural purpose of things. Neither of these points is obviously true, with particularly troubling questions surrounding the latter. There are also a number of related issues about how to describe acts (like suicide) that ostensibly contradict “self-love.” There is no need to adjudicate all of these disputes here, but it is worth noting just how unpopular this passage is, even among Kantians.
O’Neill agrees with Harrison that, contrary to Kant’s claims, the duty not to commit suicide must be an imperfect one. (Nell, 79) Herman stays away from the suicide example altogether, while Wood goes furthest of all. He questions whether we have reason to think that self-love aims to preserve our lives as long as possible, and raises the troubling possibility that some might even adopt Kant’s forbidden maxim out of self-love. (Wood 1999, 86) Finally, Wood dismisses the Formula of Universal Law’s prohibition on suicide as reflecting a view that “was widely accepted at the time.” (Ibid)

In his criticisms, however, Wood raises an important point: if our objections to Kant’s prohibition on suicide are based on doubts about various of his secondary premises—rather than on the Formula of Universal Law procedure itself—then perhaps our rejection of the passage does not indict the Formula of Universal Law so much as Kant’s misuse of it in this particular instance. Wood is right; for many, the most problematic dimension of Kant’s analysis of suicide is his reliance on the teleology of self-love, which suggests that we might be unable to conclude anything of significance about the Formula of Universal Law.

For those who are less inclined than Kant to condemn suicide, writing off this passage as an error may seem to help bring Kant’s view into alignment with their own. But doing so does not eliminate the troublesome implications for Kant. If the Formula of Universal Law cannot establish that suicide violates a perfect duty, then the Formula of

---

42 In her more recent book, The Practice of Moral Judgment, O’Neill goes further, suggesting that the Categorical Imperative may fail to show that “self-inflicted violence” violates a duty at all.

43 The flipside is that those who are strongly and uniformly opposed to suicide will not be able to accept the Formula of Universal Law as complete and accurate.
Humanity should not do so either. Kant, of course, is consistent in claiming that there is a perfect duty not to kill oneself (out of self-love, at least), and he presents an analysis (given below) of why the Formula of Humanity renders the same verdict. Either that analysis must also be mistaken, or the various formulations of the Categorical Imperative are not equivalent.

Moreover, even with this example, we can still find further support for our conclusion that the Formula of Universal Law does not tell us (properly, at any rate) why various acts are wrong. The cases in which suicide from self-love is most clearly objectionable appear to be those where, in killing oneself, a person abandons various responsibilities to others. Accounts of those responsibilities might vary greatly, ranging from special duties to family or friends, to professional responsibilities, to responsibilities given by God. In any of those cases, however, teleological accounts of the purpose of self-love come nowhere near the mark of explaining why suicide is wrong. It is the violation of one’s duties to others, in such cases, that counts most strongly against suicide—not the subversion of a natural impulse to sustain and extend one's life. Thus, Kant’s analysis based on the Formula of Universal Law could not tell us a complete story about why suicide is wrong—even if it could tell us properly when suicide is permissible.

Having dealt with the two examples Kant gives of perfect duties derived from the Formula of Universal Law, we may now step back and make some more general observations. Aside from the questions raised above, one major problem stands out: the Formula of Universal Law fails to explain why—or even capture that—certain failures are morally worse than others. If being able to universalize one’s maxim is the substantive marker of good action, then whether we have acted properly or not is a binary matter;
either our maxim is universalizable or it isn’t, so we can’t possibly say that the degree to
which our maxim is universalizable shapes how well or badly we act. Some further account
of why certain obligations are more rigidly binding than others would need to be
superimposed onto this particular Formula. On the other hand, if the Formula of Universal
Law is just a test that happens to tell us when actions are right and wrong, then it is a test
that is entirely silent on stringency.44

Either way, even if it is logically possible for Kant to supplement his view with
further details that help to explain the varied stringency of the duties that bind us—most
likely one that involves (at the very least) an account of the relationship between the
degree of harm to be inflicted and the strength of the correlative prohibition—it would
have a hollow and ad hoc feel. For it is almost certainly the case that the reason for which
certain things are wrong bears overwhelmingly on how wrong they are. No such
relationship is possible on Kant’s view, however, because the wrong-making properties
Kant identifies are not matters of degree. They are, instead, all or nothing.

One might think that, because Kant is an absolutist deontologist, he needs no
account of which bad acts are worse than others. After all, he could simply insist that we
are never to commit any prohibited action, and be done with it. But that will not do. Even
in cases where we recognize that it is wrong, all things considered, to perform a particular
action, we almost always have a good sense of how egregious a moral failure we exhibit in
performing it. That sense would have no grounding on a purely Kantian picture, and
would have to track irrelevant factors.

44 The closest Kant is able to come is dividing duties into two categories (perfect and
imperfect) depending on what sort of contradiction afflicts the concordant maxim.
On the other hand, if the universalizability of underlying maxims is just a test for determining which actions are permissible, then the test borders on useless. There are countless maxims consistent with any given action; everything hinges on how we describe both the act in question and its purpose. Suppose I adopt the maxim that, in order to be polite, I will always allow others to go through doors before me. According to Kant’s test, any actions that I execute as a result of that maxim will be immoral, for if everyone adopted that it, groups of people would never be able to pass through doorways. Something similar happens if I adopt a maxim that, in order to better myself, I ought to earn a Ph.D. (so long as we assume that there is not enough space in Ph.D. programs for everyone).

Kantians may say that I’ve botched the test. They might say that I haven’t properly identified the maxim that supports taking a given action; properly identifying the relevant maxim will yield the right result, so these examples indict my understanding of Kant rather than the universalizability test. Perhaps they are right. But the reason why these are the “wrong” maxims cannot simply be that they label as immoral a stunning array of benign (or even virtuous) acts; and even Kant scholars are deeply divided on how to distill appropriate maxims, suggesting that, at the very least, Kant was not clear enough on how we use it properly.

We have already noted (and complied with) O’Neill’s popular suggestion that we exclude irrelevant details from the formulation of our maxim. In addition, we might follow Hill’s suggestion to ensure that our maxims “only refer to facts of which we are aware, and... reflect honestly our beliefs about what is morally relevant in the situation.” (Hill 2006, 482) But while these sorts of guidelines may exclude a subset of troublesome
maxims, they will not exclude the types of maxims formulated above. Further, the actual maxim guiding an agent in any given instance will be a matter of fact. So long as it is possible for people, acting genuinely, to form and act from such problematic maxims, the Formula of Universal Law will fail as a useful normative test. At the very least, even if the universalization test would work well when used correctly, it seems clear that we have insufficient information to do that.

2.3. The Formula of Humanity

The language of the Formula of Humanity appears better suited than the Formula of Universal Law for explaining our various duties, if not at least indicating what they are. The Formula of Humanity commands us to “act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.” (Kant, 4:429) If any of the formulations stand a good chance of properly grounding constraints, the Formula of Humanity is likely our best candidate, advocating something like respect for persons. Yet, while it is an improvement over the Formula of Universal Law, this formulation also appears to miss the mark, both with explaining our duties and with identifying them.

We should take note that there are two parts to the second formulation: the first part emphasizes that we are to treat humanity always as an end, and the other specifically states that we are to avoid treating humanity merely as a means.45 As Allen Wood points

45 For more on the relationship between the two clauses of the Formula of Humanity, see Thomas Hill’s Dignity and Practical Reason in Kant’s Moral Theory, page 42.
out, the first phrase already implies the prohibition contained in the second; thus, the second expression, “never merely as a means,” actually “plays no role whatever in the actual content of [the Formula of Humanity].”46 (Wood 1999, 143, emphasis in original)

Going forward, I will take references to the first clause to imply the content of the second as well.

If we are to take apart the Formula of Humanity, it is also important to make sure that we have an adequate understanding of the term “humanity”—and, more specifically, what it means to “treat humanity as an end.” Gaining such understanding is not easy, especially because, as Thomas Hill puts it, “...as a set of rational capacities and dispositions, [humanity] is not the sort of thing which is an end, or goal, in the ordinary sense.” (Hill 1992, 42) There are a number of competing interpretations here, and we will focus on those offered by Korsgaard, Wood, and Hill.

It is common to abbreviate the Formula of Humanity—if not in writing, then at least in speech—as requiring us to treat people as an end (and never merely as a means). Kant spoke both of treating persons and treating their humanity as ends, so it is not implausible to think that “humanity” is meant just to stand in for “persons.” Hill agrees that the two terms are equivalent, but prefers the opposite interpretation: he thinks that “treating persons as an end” is elliptical for “treating humanity as an end” (Hill 1992, 39)

Korsgaard and Wood also seem to favor this view, where we understand “humanity” as a quality possessed by persons. Perhaps the reason for the relative consensus on this point

46 It has been pointed out by Robert Audi that treating someone as an end and treating him merely a as a means are not the only two morally relevant possibilities. He gives the example of speaking loudly during another’s lecture. (Hurka, 66) This is a point we’ll revisit below.
is that there are a number of ways of treating people merely as a means that are not morally objectionable. Ducking behind a stranger to avoid detection by another surely uses the stranger merely as a means (as a shield), but it is not necessarily objectionable. A more nuanced understanding of the scope of the Formula of Humanity’s injunction—one tracking humanity rather than persons more broadly—rules out this class of false positives.47

Hill, Korsgaard and Wood also agree, more or less, on the meaning of “humanity.” Considering that Hill claims “the texts are not unequivocal about exactly what characteristics make up our humanity,” this agreement may be somewhat surprising. (Hill 1992, 39) All three have narrow definitions in mind. According to Korsgaard, “...the distinctive feature of humanity, as such, is simply the capacity to take a rational interest in something.” (Korsgaard, 114, emphasis in original) For her that means being able to use reason to identify intrinsically valuable ends. (Ibid) Hill, likewise, thinks of humanity as “our rationality and capacity to set ends.” (Hill 1992, 50) Wood also thinks that the term “humanity,” as used in the Formula of Humanity, refers specifically to our “rational nature.” (Wood 1999, 144)

Korsgaard offers the sparsest description of what it is to treat humanity merely as a means. She emphasizes that, wherever we find humanity, it should enter into our deliberations “negatively,” which means that we may not act against it. (Korsgaard, 125) Humanity, according to Korsgaard, is not something for us to realize or bring into existence; it is not a value to be maximized. (Ibid) Hill generally seems to agree, but adds

47 I take it that a Kantian can accept this interpretation of the Formula of Humanity and still deny that value sits in states of affairs.
some more detail. He claims that, in addition to abstaining from impairing or destroying someone’s humanity (whether through physical attack or manipulation), we are to use our rational capacities to the extent that we can, and help develop them in others and in ourselves.\footnote{There seems to be a tension between the idea that we cannot promote humanity but recommendation that we ought to help develop it where we can. We will let that pass, however, in the interest of pursuing more germane points.}

It is Wood who offers the most specifics about what he thinks is required to treat humanity as an end. In addition to (more or less) endorsing the points given by Korsgaard and Hill, Wood emphasizes that it is still permissible for us to value ourselves under various descriptions that make no explicit mention of our rational natures (“wealthy, powerful, handsome, [etc.]”) so long as those “modes of self-valuation... are derivative from valuing humanity.” (Wood 1999, 145, emphasis in original)

Wood also discusses a salient concern that Korsgaard and Hill generally leave aside. The language of the Formula of Humanity raises legitimate questions about what we are to do with less-than-rational creatures. Presumably, young children, people with certain illnesses (such as Alzheimer’s), people in comas, and maybe even people who are sleeping all lack humanity in one sense or another—not to mention the majority of animals.\footnote{Wood offers a similar list of those who might find themselves in the penumbra of humanity.} Wood claims that “it would be absurd to suggest that [the Formula of Humanity] has nothing to say about how we should treat such people.” (Wood 1999, 144) He continues: “[W]e would surely dishonor rational nature if we did not cherish its development in children or failed to strive for its recovery in men and women who have
The reason for this is that treating humanity always as an end, for Wood, requires more than simply treating humanity in people always as an end. In making this claim, Wood attempts to round up many of those who we would not want to exclude from consideration in the Formula of Humanity. Wood even thinks that we should value our “capacities for feeling and emotion and even our animality,” because all of these are parts of our rational nature. (Wood 1999, 144)

There is reason to doubt that Wood’s reply here is successful. Samuel Kerstein has pointed out that the capacity to develop rationality is different from rationality itself, and it is unclear why the Formula of Humanity would require us to value the former. (Kerstein, 765) Given the relative consensus (at least among Korsgaard, Hill and Wood) that the Formula of Humanity applies to those who possess rationality, this point should not be overlooked.

It isn’t obvious how our animality constitutes part of our rational natures either, given that Wood does not endorse treating animality in animals with any special regard. There do seem to be situations in which it would be objectionable to harness, for our own purposes, someone’s animal nature or natural impulses (as opposed to their rationality). If it is not possible to roll these qualities into our humanity, then the Formula of Humanity has a serious problem. Trading on a person’s irrational fear of germs because it’s funny does not seem to treat anybody’s humanity merely as a means, but we would say it is morally wrong for much the same reason as tricking someone into making a fool of themselves. The same would be true of exploiting a person’s natural impulses for food or sex.

Still, general assessments of the Formula of Humanity can only take us so far. The
merits and weaknesses of this formulation will become more readily apparent when we consider Kant’s derivation of duties from it. We will, once again, use the prohibitions on false promising and suicide. Kant applies the Formula of Humanity to false promising as follows:

“...he who has it in mind to make a false promise to others sees at once that he wants to make use of another human being merely as a means, without the other at the same time containing in himself the end. For, he whom I want to use for my purposes by such a promise cannot possibly agree to my way of behaving toward him, and so himself contain the end of this action.” (Kant, 4:429-430)

There are several ways we could interpret this passage. Perhaps the simplest reading is that someone to whom you make a false promise would not consent to your doing so because people dislike being misled. As an explanation for our duties, this reading is unappealing. If what makes x wrong is just that the victims of x do not like the prospect of x, then we won’t be able to offer a strong prohibition against acts about which people’s feelings are divided. There is also a risk that we would be unable to prohibit any behavior towards ourselves so long as we are sufficiently pleased with it, including self-mutilation, voluntarily entering servitude, and any other form of self-degradation. On the other hand, if we expand the range of people whose feelings matter so as to block problematic self-regarding behavior, we will include the feelings of others who are only indirectly affected by x. That would yield its own unappealing consequences, as, for example, my choice to pursue a certain career might be wrong if it disappoints my father. For similar reasons, this is not a popular interpretation of how we identify our duties either. I know of no one who defends this view, and at least Hill explicitly disavows it. (Hill 1992, 45)
Another interpretation of the Formula of Humanity’s passage on false promises more or less channels the second formulation through the first. That is essentially Hill’s view: “[I]nsofar as they are used as means, [people] must be able to adopt the agent’s end, under some appropriate description, without irrational conflict of will.” (Hill 1992, 45, emphasis in original) For Hill, that simply means asking whether the agent's maxim could become a universal law, and we are thus to ask whether the maxim of making false promises could be universalized.50 But, whether we want to know if the Formula of Humanity explains or merely identifies our duties, this approach is vulnerable to all of the objections raised against the Formula of Universal Law discussed above.

Wood interprets the false-promise passage to mean that it must be possible for the person to whom I make a promise to share my end.51 He thinks this approach also helps explain the sense in which universalizing false promises is impossible under the Formula of Universal Law. “If... the giving and accepting of a promise requires the promiser and the promisee to share their ends, then it is clear why the [universalization of the maxim prohibited by the Formula of Universal Law] would be impossible [too], since they preclude such sharing.” (Wood 1999, 148, emphasis in original)

But it is unclear why the case of false promising precludes a sharing of ends in the relevant sense. Suppose a friend of mine wishes to borrow money from me, knowing that

50 There may be a slight gap on Hill’s view between the Formula of Universal Law and the Formula of Humanity in that the latter focuses on the recipient of action rather than the agent; but I see no relevant difference for the purposes of the objections raised here.

51 Korsgaard’s view can be understood to align either with Hill or with Wood. She writes, “If one’s end cannot be shared, and so cannot be an object of the faculty of desire for everyone, it cannot be good, and the action cannot be rational.” (Korsgaard, 116)
he can’t pay me back. He thus makes a false promise to repay the loan I give him. I could very well agree to give him the money, all the while hoping that he doesn’t repay me, either because I think he would be better off accepting the money as a gift (due to hardship), or because I want to prove a point that he is irresponsible. Further, suppose that I know he won’t accept the money as a gift, and thus that he will only accept the money from me if I allow him to promise to repay. Even without knowing for certain that his intent is to deceive me, I can coherently and rationally share his end: namely, that he secure money, for his own benefit, through a false promise.\textsuperscript{52}

This type of case may not be particularly common, but it is surely conceivable. This may mean that the Formula of Humanity would not always ground principles that stand without exception. Perhaps we would be able to justify making false promises sometimes, even under the Formula of Humanity. That seems like a good outcome to me, but I think it also suggests that Wood’s interpretation would, at the very least, not count against false promises made by someone under the conditions of our particular case. Whether to explain or to identify our duties, surely that isn’t right; my friend is still acting badly even if I share his end.

More dangerous is the possibility that, as a result of counterexamples like this one, Wood’s view would not be able to offer a prohibition on false promises at all. After all, it may not be true that, in principle, false promising cases necessitate the inability of the promiser and the promisee sharing an end. Further, Wood’s view suggests that accepting a

\textsuperscript{52} It is perfectly consistent with this sharing of ends that I prefer he tells the truth, rather than making a false promise, but would rather he have the money than not, so I accept the false promise.
promise that we suspect is false would be wrong because it involves “consent[ing] to be treated in a way that disrespects [our humanity] and reduces it to a mere means.” (Wood 1999, 149) That, also, is counterintuitive, as it does not seem to be immoral to loan money to a friend under the conditions we have described.

Much as we saw with the Formula of Universal Law, analyzing Kant’s Formula of Humanity prohibition on suicide may not give us the strongest of conclusions about the merits of the Formula itself. Still, a few points are in order. In particular, the different analyses offered of this passage are worth considering as possible explanations or tests for our various other duties.

Kant describes the wrongness of suicide in terms of humanity as follows:

“...someone who has suicide in mind will ask himself whether his action can be consistent with the idea of humanity as an end in itself. If he destroys himself in order to escape from a trying condition he makes use of a person merely as a means to maintain a tolerable condition up to the end of life. (Kant, 4:29)

As one might expect, there are once again multiple interpretations available here. The issue of consent raised in the prohibition on false promising seems ill suited for explaining a prohibition on suicide. For one, most (if not all) suicides involve consent. Additionally, Kant acknowledges a duty to promote our own happiness. (4:399) If we value our lives free from suffering, have a duty to pursue happiness, and consent to taking a self-regarding course of action that makes us happier than we would be otherwise, the false promising description is unlikely to be helpful in offering a prohibition on suicide.

For Korsgaard, suicide is positively contradictory because what makes a tolerable condition valuable, she claims, is the endorsement it gets from a rational being—the same
rational being who is destroyed in the act of suicide. Once you have killed yourself, then, the tolerable condition ceases to have value. This is a strange and rather gimmicky interpretation. The fact that, after you succeed, you are no longer around to confer value on the tolerable condition you pursue through suicide in no way makes it irrational to pursue suicide while you are alive. Put another way, ceasing to value a tolerable condition after you are dead does not constitute a reason against killing yourself if remaining alive is intolerable. No longer existing is precisely the point. More broadly, we cannot generalize from this description of the wrongness of suicide, either to explain what is wrong with various other acts or to indicate what we may and may not do. Korsgaard’s interpretation hinges entirely on the destruction of the rational agent who is in pursuit of a particular outcome; as a result, we won’t be able to establish any prohibitions through this test that don’t involve eradicating the agent in question.

Wood takes a more teleological view on the suicide prohibition that aligns his interpretation of the Formula of Humanity more closely with his understanding of the Formula of Universal Law. Citing The Metaphysics of Morals, Wood claims that Kant wants us to “respect the natural purposiveness involved in our instinctive impulses... since they belong to the predisposition to animality which is the underpinning of our rational nature.” (Wood 1999, 148) Acting against our natural impulses is thus, on Wood’s view, to show disrespect to our humanity.

This approach is most likely a dead end as well. Some of our earlier doubts apply here; specifically, it is not clear that “self-love” (or whatever natural impulse Wood appeals to) aims at the extension of life more than at preserving our happiness. Moreover, Wood’s position only works if it makes sense to say that our animality constitutes part of
our rationality when we have or stand to gain rationality, but it has no significance in a creature that cannot acquire rationality. I am skeptical on both counts, but there may be broader concerns indicated here. Hanging the wrongness of suicide on the purpose of some natural impulse threatens to indict other, benign subversions of such impulses. It is common to manipulate natural impulses for a variety of reasons, ranging from spiritual or religious celibacy to political hunger strikes. Whatever distinctions we might want to draw between these impulses and the impulse of “self-preservation,” if it is to be compelling, there has to be more to the suicide prohibition than a bare assertion that acting against natural impulses is wrong.

Hill’s view is somewhat less definitive, but also more plausible. He takes this passage to express a prohibition on trading humanity—which has no price, or rather, is above price—for a “tolerable condition,” which is of lesser value. Such a trade degrades humanity by treating it as if it has less value than it actually does. (Hill 1992, 51) Hill is clear that Kant does not expect every life to be preserved all of the time, and Kant does not claim that “life [itself] is irreplaceable.” (Ibid, emphasis in original) Yet he claims, as I think many Kantians do, that “the proper attitude about humanity is not that each bit of it has a value which one can weigh against the value of other bits to calculate reasonable tradeoffs.” (Hill 1992, 52)

Kant may think this, but such a stance doesn’t follow logically from humanity having unconditional value. It is a further claim altogether to say that we need not (and even more strongly, ought not) to aim at preserving more humanity rather than less. Korsgaard’s point above about humanity entering into our considerations “negatively” is likewise unsupported by any mere claims about the specific value of humanity. It is
unclear how, on Kant’s view, the prescribed attitude towards humanity can find support in anything but bare stipulation.

Neo-Thomists often claim that various goods are incommensurable, but such a claim is not available to Kant; we are told quite explicitly that we are to regard humanity as equally and unconditionally valuable wherever we find it. As a result, it’s surely possible to sum up the humanity at stake in various tradeoffs. Hill tries to deny this, claiming that “a parent of three children faced with the awful choice of saving two or one might on some ground choose to save the two without having to grant that two are worth more than one.” (Hill 1992, 52) But how can this be? That seems quite plainly to contradict any coherent understanding of what means to say that each person who possesses humanity has unconditional value.

Nevertheless, we can drop Hill’s latter claim and simply say that the proper attitude towards humanity is one of respect or non-violation. While it would be nice to know why that is the case, at least Kant could claim that we are not to do harm to humanity even if that means allowing more harm to humanity overall (that is, from other sources). There is still a question about whether this approach could explain our duties, or even justify them. While it is certainly a better justification than many of the other possibilities we have reviewed, it would not give us everything we need, for not everything one can do wrong involves acting against humanity or corrupting yourself for your future interaction with humanity. Intuitive prohibitions on harming animals cannot be justified this way (whatever Kant says to the contrary). In spite of these challenges, I think we can regard Hill’s interpretation as perhaps the most compelling, though it will not evade certain problems that we will discuss below.
Given the unpopularity of the suicide prohibition among contemporary Kantians (noted in our discussion of the Formula of Universal Law), there may be limited value in pressing the present matter much further. There are two reasons for us to engage with the Formula of Humanity’s ban on suicide, and we have fulfilled both already. First, for those who wish to maintain the prohibition on suicide, I think it was worthwhile to sort through the prominent interpretations in search of the best approach. Perhaps more importantly, however, we see that prominent Kantians seem happy to defend the Formula of Humanity’s ban on suicide while disavowing the similar ban given by the Formula of Universal Law. While it is not fatal to the Kantian approach, it is of nontrivial significance that many of the most vocal Kantians reject Kant’s stated view that the formulations of the Categorical Imperative are coextensive.

There remain some general problems with the Formula of Humanity that pose a greater threat to the Kantian justification for agent-centered restrictions. One comes to us via Tom Hurka, who has pointed out that Kant ascribes humanity to people on an agent-neutral basis. (Hurka, 68) If everyone has the same worth, he wonders how Kantianism can accommodate self-referential altruism. How is it justifiable (let alone morally required, as we commonly think it is) for us to treat some people, such as family and friends, with more care and consideration than we treat strangers? Hurka thinks the same question arises for agent-centered prerogatives (or options). On what grounds can we justify treating ourselves preferentially if the only valid ascriptions of value place humanity, which is of equal worth, in each of us?

There is another problem that I have claimed applies to the Formula of Universal Law as well: the Formula of Humanity is not well suited for describing the intuitive
variations in the stringency of various prohibitions. As noted above, a theory of the right should tell us more than what we may or may not do; it must provide at least some sense of how badly we have behaved if we do what we oughtn’t to have. Practically nobody thinks lying is as bad as murder, or false promising as bad as rape. Even if we can derive a complete accounting of where constraints should fall by appealing to the Formula of Humanity—something I have suggested is not possible—we would still need further information from Kant about why some acts are so much worse than others.

The only way I can see to make use of the Formula of Humanity in determining the stringency of various prohibitions is to appeal to correlative variation in the language that Kant has given us: it is immoral to treat others’ humanity merely as a means and not also as an end, and the more we use someone merely as a means in any given instance, the worse we behave. Of course, it isn’t obvious how to quantify the use of someone merely as a means, though that may not be a decisive objection. More serious is that, even if we could settle on a formula, it does not appear that treating someone as a mere means more rather than less would even be able to account for the varied stringency of our duties. For example, suppose that lying involves treating someone as a mere means at level $x$, while murder involves some greater level $y$. Surely a certain number of $x$’s (a certain number of lies) would come to a total greater than $y$ (a murder), even though intuitively we would think that, all else equal, telling a large number of lies is not nearly as bad as killing someone.
2.4. The Formula of Autonomy and the Formula of the Kingdom of Ends

The two remaining formulations—the Formula of Autonomy and the Formula of the Kingdom of Ends—are commonly treated as being closely related. The Formula of Autonomy emphasizes “...the Idea of the will of every rational being as a will that legislates universal law.” (4:431) The Formula of Kingdom of Ends, meanwhile, reads as follow: “All maxims which stem from autonomous lawgiving are to harmonize with a possible kingdom of ends as with a kingdom of nature.” (4: 436)

These formulations are, by most accounts, more opaque than the first two, and perhaps as a result, they seem to receive somewhat less attention. Both return in some sense to the language of the Formula of Universal Law, the former through its reference to “willing” and the latter through its mention of “maxims.” Yet both add new twists by emphasizing the act of legislation. Moreover, the Formula of Autonomy makes use of the word “autonomy” in an unintuitive and technical sense, specifically to refer to a will’s ability to choose actions not from inclination but in accordance with the Categorical Imperative. The Formula of the Kingdom of Ends also adds an eponymous term of art; according to Kant, the Kingdom is Ends is supposed to be “a systematic union of different rational beings under common laws.” (4:433)

The idea behind these formulas turns on Kant’s view that “the will is not merely subject to the law but subject to it in such a way that it must be viewed as also giving the law to itself.” (4:431) As we noted above, Kant viewed a will’s ability to give itself the proper constraints as an essential component both of its rationality and of its potential goodness. At the most basic and uncontroversial level, these two formulas of the Categorical Imperative seem to encapsulate that point.
But of course, there is much debate about how we are to understand the process by which the members of the Kingdom of Ends legislate their laws. Thomas Hill Jr. considers this question to be of the utmost importance for the future development of Kantian ethics. (Hill 2006, 497) Kant’s description of the Kingdom suggests a harmonious community whose members all share the same set of reasons. One question that arises here is whether what is distinctive about the community is the manner in which it actually legislates, or just the mere fact that everyone in it acts autonomously.

It is also unclear whether we should take the Kingdom purely as a heuristic device—indeed, it seems to have served as a model for Rawls’ original position—or more concretely as the idealized consequence of a society or union composed only of good wills. Kant believed that the Kingdom of Ends would in fact be realized if everyone universally followed maxims that comport with the Categorical Imperative, which counts in favor of the latter. But since Kant thought that we cannot verify the existence of a single good will (let alone a society built of them), we might consider the mere conceptual possibility of the Kingdom moot for all practical purposes.

In any case, it is doubtful that either of these formulations can accomplish what the first two failed to do—namely, offering a compelling account of constraints. More specifically, it is almost certain that making use of universalizable maxims would be essential, in Kant’s mind, to the process of exercising one’s autonomy as well as to the process of legislating for a Kingdom of Ends. If this assumption turns out to be true, then the criticisms I have leveled above would apply with equal force to these latter two formulations.

The vagaries surrounding the latter two formulations are not entirely a matter of
scholarly self-selection. Earlier parts of our discussion dealt heavily with the systematic effort Kant makes to illustrate the prohibitions on false promising and suicide through the language of the Formula of Universal Law and the Formula of Humanity. It is perhaps telling that Kant does not offer a similar explanation of how to understand these same (or any other) prohibitions through the use of the Formula of Autonomy and the Formula of the Kingdom of Ends. The salient implications of this asymmetrical treatment are either that the latter two formulas are so perspicuous that no elaboration is necessary, or that an extended discussion of the sort provided for the Formula of Universal Law and Formula of Humanity would be redundant. The former of these possibilities is so manifestly implausible—especially given the opacity of a phrase like “the Kingdom of Ends”—that I wholeheartedly endorse the latter.

**Conclusion**

In order to take stock of where we stand, it may be useful to revisit the schema introduced in the previous chapter. Recall: the consequentialist believes that all of our duties derive from claims about value, that value sits in states of affairs, and that the appropriate attitude to take toward value is to promote it. We have already noted that Ross denies the first and the third of these claims, while accepting the second. Kant, on the other hand, denies the second and third, accepting just the first. The Formula of Humanity offers us the clearest view of why it is that Kant thought of our duties as responding to claims about value. The incomparable value of humanity warrants certain treatment, and that is what shapes our moral obligations.
Perhaps the biggest problem with Kant’s account of agent-centered restrictions is that he locates value exclusively in humanity or dignity, and not in states of affairs. To put it rather glibly, even if value resides in people, people reside in states of affairs. To put it less glibly, unless we have some way of counting the relevance of states of affairs, we will be unable to take account of actual harms in shaping the rigidity of various prohibitions. It is quite literally impossible to generate a plausible account of agent-centered restrictions without the ability to differentiate the wrongness of acts and the stringency of prohibitions at least in part by reference to consequences. All else equal, stabbing is worse than punching, and punching is worse than pushing; the reason for this spectrum (and countless others) is just that there is a distinct range of harm generally inflicted by these different acts.

There is a further difficulty in shaping constraints without the ability to refer to consequences. It is common to think of Kant as an absolutist deontologist, which means that most regard the constraints produced by his system as inviolable. The problem, we might think, is that it is uncommon for people to have absolutist intuitions about constraints. But we must refine this concern, for Kant can accept constraints that have exceptions. We are to respect humanity, but we can kill in self-defense because someone attacking us has forfeited his right to have his humanity respected. What Kant’s approach cannot do is provide for duties that are outweighed by countervailing considerations, and this is another seriously counterintuitive result. Ross, by contrast, avoids this problem

53 For this reason, the fact that we might be able to formulate consequence-sensitive maxims that seem to pass the test of the Formula of Universal Law is not enough does not get us very far in making Kant’s view more plausible.
because he accepts the proposition that value inheres in states of affairs and thus that consequences can be of moral significance.

In short, the Kantian approach seems to fare worse than Ross’ in terms of its ability to generate an attractive account of agent-centered restrictions. Though Kant does not have to explain duties that are not grounded in value, he must confront other difficult problems. The core principles his constraints rely on—the Formula of Universal Law and the Formula of Humanity—gesture at morally important considerations (fairness and respect for persons, respectively); but alone, they seem insufficient to produce the sort of base we need in order to support the constraints we typically accept. The Kantian approach is defined, in part, by its ascription of value to people (or humanity) rather than to states of affairs, and that leads to further, perhaps even greater, difficulties. The only remaining prominent deontology is Neo-Thomism, which accepts yet another configuration of the consequentialist’s claims, and thus confronts a set of issues that is largely different from those we have discussed up to this point. That is where we will pick up in the next chapter.
3

Neo-Thomism

Introduction

Neo-Thomistic deontological views share more with consequentialism than either intuitionism or Kantianism. Like consequentialists, Neo-Thomists accept both that value inheres in states of affairs and that our duties are grounded in claims about value. But Neo-Thomists deny that we ought to aim at promoting the good, instead holding that we may not act against or destroy various goods. They claim that there is a limited range of attitudes that it is morally acceptable to take toward the good, and that range does not include one permitting their violation, even in pursuit of desirable consequences.

The standard Neo-Thomistic assertion as to why we oughtn’t to promote the good is that the goods are objectively incommensurable, and thus trading off among them in the interest of maximizing is impossible, or even incoherent. Some, like Finnis, go even further in denying maximization by rejecting subjective theories of rightness (doing the best we can with the information available) in favor of some version of an objective theory.

Neo-Thomists typically build their view around an array of objective goods thought
essential for undergirding some conception of human flourishing. They then derive a list of duties by applying the permissible attitude(s) to the items on the list. In coming to understand a Neo-Thomistic approach to deontology, it will be useful to focus on a specific example of such an approach. John Finnis defends the most prominent, fully developed theory of this sort, so I will devote this chapter to elaborating on and evaluating his version of it. In particular, I will draw on his arguments from Natural Law and Natural Rights, Fundamentals of Ethics, and, where necessary, Nuclear Deterrence, Morality and Realism (which he co-authored with Germain Grisez and Joseph Boyle).

3.1. John Finnis’ Normative Ethical Theory

While there are many ways of understanding human flourishing, an appeal to some such notion is a common feature of the Neo-Thomist approach. Finnis’ view epitomizes this strategy. He posits human “fulfillment” (or “full-being”) as the central good that we ought to pursue, starting from the following primary moral principle: “Make one’s choices open to human fulfillment: i.e. avoid unnecessary limitation of human potentialities.”

54 As far as I’m concerned, this makes neo-Thomistic ethical views objective list theories, but some, such as Derek Parfit, may prefer to restrict the use of the phrase “objective list theory” to theories whose list of goods cannot be unified under a broader heading. Under such a definition, the sort of approach I consider here would not, technically, represent an objective list theory.

55 I cite passages in Natural Law and Natural Rights and Fundamentals of Ethics using Finnis’ name and the year of publication; but because Finnis coauthored Nuclear Deterrence, Morality and Realism with Germain Grisez and Joseph Boyle, I cite that book using the abbreviation “ND.” To the extent possible, I avoid attributing to Finnis views that are expressed in the co-authored book, but I override that rule in cases where there is no analogous material in Finnis’ sole-authored books and where the content of a passage is particularly useful for the purposes of elaboration.
He develops a list of basic human sub-goods that conduce to such fulfillment, and he derives ethical precepts from the attitudes agents ought to take towards these goods. When we combine Finnis’ primary moral principle with the ways in which he thinks we can participate in basic human goods, we get intermediate moral principles that give us more specific guidance.

Finnis defends his choice of integral human fulfillment as the central quality supported by his list of goods in part by rejecting various alternatives. Relying heavily on Robert Nozick’s “Experience Machine” thought experiment, he concludes that, “...the notion that pleasure, in any sense, is the point of everything is absurd.” (Finnis 1983, 89)

Adopting John Rawls’ definition of the term, he also rejects all “thin theories” of the good. More specifically, he calls a theory of the good “thin” if (and only if) it “offers to identify as the basic human goods those goods which any human being would need whatever his objectives.” (Finnis 1983, 48, emphasis in original) Part of Finnis’ objection to thin theories of the good is that, he claims, they depend on thin theories of practical reason (whereby what is rational for a person will be determined by his preferences). Given Finnis’ Aristotelian commitment to a substantive view of practical reason, his objection on this score is not surprising. The objective goods that Finnis accepts directly conflict with both thin theories of the good and thin theories of practical reasoning.

Perhaps worse, according to Finnis, is that thin theories of the good are “arbitrary.” (Finnis 1983, 50) They ignore “whole ranges of human opportunity,” by “treating them as if they were... matters of taste, subjective opinion, individual desire... rather than what they are: opportunities to participate in real goods, in really worthwhile aspects of human flourishing. (Ibid) Finnis thinks that a fear of “authoritarian politics” drives the appeal of
thin theories—fear of illicitly imposing a conception of the good on agents. But he claims that such worries are unfounded because the content of accurate lists of goods will offer a sufficient degree of protection for the liberty of every individual.

All items on Finnis’ list of goods must connect fundamentally with personhood, and offer a source of intrinsic goodness. Finnis provides the following list: life itself, knowledge, play, aesthetic experience, sociability (friendship), practical reasonableness, and religion. (Finnis, 1980) Whatever one’s thoughts on integral human fulfillment, or on how the listed items might conduce to it, our interest lies in what makes the view deontological. As a deontologist, Finnis must reject the suggestion that we ought generally to maximize the good (or integral human fulfillment, more specifically). He thinks that there is only one appropriate attitude for us to take towards the goods; and while that attitude defies succinct description, and has no English synonym, it does not conduce to (and positively prohibits) maximizing. The idea is that we must show something like “respect for every basic value in every act,” and doing so rationally precludes choosing to harm any basic value.56 (Finnis 1980, 118)

Finnis’ understanding of what it is to harm a basic value seems to align with the Doctrine of Double Effect, though Finnis does not tend to employ that language. At multiple points Finnis speaks of avoiding actions that directly harm a basic good, and he refers to the prohibition on certain intentional acts. (Finnis 1980, 112) Moreover, in the book he co-authored with Joseph Boyle and Germain Grisez, we find a fairly explicit endorsement of the Doctrine of Double Effect. (ND, 292) We also find there a list of the

56 For simplicity, I will refer to this attitude simply as “respect”— notwithstanding the other possible connotations of the term.
three primary ways to participate in human goods. (These are not to be confused with the countless ways in which we might “share” in them). The first way to participate in a good is when “one chooses something for its intrinsic value.” (ND, 289, emphasis in original) Another way is “when one chooses something not for itself but as a means to some ulterior end.” (Ibid, emphasis in original) Finally, “one acts in a still different way in so far as one voluntarily accepts side-effects caused incidentally to acting in either of the two prior ways.” (Ibid, emphasis in original) Collectively, this evidence suggests rather strongly that Finnis himself places significant weight on the intention/foresight distinction. Accordingly, I will assume that we can interpret the view as conforming to the Doctrine of Double Effect.

Finnis arrives at his deontological conclusion by positing an especially tight relationship between practical reasoning and the basic human goods. He claims that “[p]ractical reasoning makes its claim upon us because it is a basic aspect of human flourishing” and the function of practical reasoning is “to direct the way in which we seek to participate in each and all of the basic human goods.” (Ibid) If true, this claim essentially defines appreciation of the goods into the process of engaging in any goal-oriented (or agentic) activity. When we pair that suggestion with Finnis’ view that the moral significance of human acts derives primarily from the actor’s involvement with the basic human goods, we get the outlines of a fuller deontological theory.

More specifically, recall that Finnis’ primary moral principle bars us from adopting attitudes that are inconsistent with the pursuit of integral human fulfillment. Finnis notes that this sort of formulation is too vague to be of much assistance in illuminating our moral responsibilities, so he attempts to explain morally forbidden and morally
discretionary actions by uncovering what he calls “requirements of practical reasonableness.” These requirements are essentially intermediate moral principles that Finnis derives from the primary moral principle.

Finnis argues for nine requirements of practical reasonableness. Although not all of them are directly relevant here, a quick glance at the list helps to give color to Finnis’ view and will let us pick out the principles of greatest importance. The first requirement of practical reasonableness stipulates the relation between practical reason and the other goods, as described above. The second enjoins us to form a coherent plan of life around those goods. Such a life can take innumerable forms so long as it is structured by the proper attitude of respect and an interest in integral human fulfillment. The third and fourth requirements tell us, respectively, to avoid arbitrary preferences among the goods and among people. The fourth quite plainly demands of agents an agent-neutral assessment of states of affairs. The fifth requirement demands fidelity to one's commitments without single-minded obsession. The sixth lays out Finnis’ views on incommensurability and indeterminacy in pushing back against intuitions to consequentialize, which we will discuss in detail below. The seventh is the essential deontological requirement, which posits the respectful attitude we must take towards the goods and insists that we may not intend to harm them, even with an eye toward good consequences. Finally, the eighth requirement of practical reasonableness produces a more specific list of duties with respect to the common good; and the ninth concerns heeding one's conscience.

The most relevant for us, aside from the seventh, are those addressing the equal weight of persons (the fourth), and the limitations on consequentializing (the sixth).
Finnis’ assessment of the fourth requirement is interesting in light of the preceding chapter on Kantianism. Finnis regards this requirement as taking the form of something like the Golden Rule, which he claims can also be understood as a principle of universalizability. As a requirement of practical reasonableness, the Golden Rule is meant to block egoism, and the unjustified preference for those closest to us. For Finnis, the impulses that drive us to favor ourselves constitute one class of impermissible attitudes, and the Golden Rule, applied loosely in Kantian fashion, is about ensuring that we treat others fairly.

Yet Finnis criticizes any ethical system (likely of Kantian form) that defends this sort of fairness requirement as the only principle of morality, and sharply criticizes the notion that we should focus—as Kant does—only on qualities we do not share with other creatures. (Finnis 1983, 122) Finnis thus regards Kant’s view as unduly focused on only one important requirement of practical reasonableness. According to Finnis, there has to be more to morality than mere fairness, and we need to make sure we capture all aspects of human flourishing.

As noted above, the fourth requirement of practical reasonableness also imposes on us an agent-neutral injunction of sorts. But obviously Finnis’ is no consequentialist. The sixth requirement, which tells us that maximizing is both impossible and incoherent, prevents us from falling into consequentialism. Moreover, the seventh requirement’s restriction on intending harm against the goods makes us adopt agent-relative attitudes; it gives us each an obligation to regard the goods in a certain way, making it more important to avoid harming the goods than to prevent harms to the goods at the hands of others. Yet as we shall see, it is not clear that Finnis’ normative stance is sustainable, for there are
serious doubts about the deontologically-oriented requirements of practical reasonableness.

The seventh requirement of practical reasonableness relies quite heavily on the sixth, which comprises both incommensurability and a secondary feature of his view that Finnis tends not to emphasize quite as much: an objective view of rightness. The ostensible rationality of acting against various goods in the interest of promoting a greater overall volume of good is illusory, says Finnis. Not only are such tradeoffs senseless because of the incommensurability of the goods; additionally, Finnis thinks that what is morally relevant includes possible instantiations of goods diverse in kind from one another, as well as the field of unrealized possibilities opened up by the basic human goods.” (Finnis 1980, 117) This feature of Finnis’ view injects a level of indeterminacy that, he claims, prevents us from measuring various quantities of good against each other.

On Finnis’ view, incommensurability and indeterminacy ensure that we can only choose arbitrarily to sacrifice a good (directly, by damaging it). We have no principled way of choosing among them, so we would be “playing favorites” without reason. (Finnis, 1980, 118-119) While we might think that we make principled choices between the goods, Finnis would say that we are mistaken. We are not in fact acting in ways that maximize good or minimize evil, he says, for that is impossible. Instead we are acting on the basis of our feelings about the various instances of the goods. (Finnis 1980, 120) Thus, Finnis justifies the ban on acting against the goods.
3.2. The Implications of Neo-Thomistic Deontology

Let us begin with the sixth requirement of practical reasonableness. Finnis argues for three specific points in defending the view that all of the goods are equally basic (and thus cannot be weighed against the others): first, each item on the list “is self-evidently a form of good”; second, “none can be reduced to being merely an aspect of any of the others, or to being merely instrumental in the pursuit of any of the others”; and third, “each [good], when we focus on it, can reasonably be regarded as the most important.”⁵⁷ (Finnis 1980, 92) On the basis of these three assumptions, Finnis concludes that there can be “no objective hierarchy amongst them.” (*Ibid*)

The third of these claims strikes me as the most dubious. There are a number of philosophers who defend objective list theories of the good while simultaneously maintaining that the various goods are not (in the abstract) of equal value. Ross thought that virtue, knowledge, pleasure, and justice are the four intrinsic goods, but that virtue stands alone as the most valuable. Thomas Hurka holds a view that is almost the reverse. He endorses a recursive characterization of good (and evil), on which goods like pleasure or knowledge or achievement have more value in the abstract than a good like virtue, which requires possession of an attitude of the appropriate valence directed at the other goods.

---

⁵⁷ Though he does not explicitly give it, we should also credit Finnis with the corollary: there is no other good in the world that satisfies all three of these conditions. Some might wonder whether it is fair for Finnis to defend the primacy of his list of goods in part based on how they look when we focus on them one at a time, for earlier we have been told to look at them globally and in the abstract. Even if we were to comply with Finnis’ directions here, it isn’t clear how incommensurability follows.
Perhaps worse, though: even if we grant Finnis his third point, it is not clear how incommensurability follows. The ostensible incommensurability of the goods in the abstract is not of particular concern when we are faced with concrete decisions. With relatively clear alternatives before us, why would we be unable to gauge courses of action by estimating the volumes of good to be gained or lost by pursuing the different options? Further, there might be a right answer about which option will yield the most good even if we do not always know it. And further still, certain cross-good comparisons seem beyond dispute. Which is better: the pleasure of eating some pudding or the knowledge of all of the facts of history? This is an extreme comparison, no doubt, but the fact that we can even make it contradicts Finnis’ view, and illustrates just how out of sync it is with our own intuitions.

Now, Finnis would say that this challenge misunderstands the nature of the relationship between the basic human goods and integral human fulfillment. He argues that there are countless different ways in which we can share in the goods, and that we pick between these various possibilities by relying on personal preferences, not moral principles. He claims that we each have “a subjective order of priority amongst the basic values” and this order reflects our choices. (Finnis 1980, 93) He gives the example of someone who, in choosing to be a scholar, subjectively prioritizes knowledge ahead of friendship over the course of his lifetime. These sorts of choices, most broadly, reflect our chosen life plans, and the factors that shape our choices are things like our “temperament, upbringing, capacities, and opportunities”—not “differences of rank of intrinsic value between the basic values.” (Finnis 1980, 94)

This is a compelling enough description of the way in which people choose life
plans. It seems true that we accept a variety of meaningful ways to spend our time and energy, electing to pursue those paths we find most appealing, based on a complex function of our personalities and the possibilities we encounter. The problem is not that Finnis has this particular description wrong; it is, rather, that a view that features this description as its centerpiece cannot account for cases where we feel, quite clearly, compelled to act one way or another—not because of a subjective prioritization of various goods but because we perceive such a sharp, objective disparity between the choices that we feel morality has made the choice for us.

Classic consequentialist dilemmas rely on this intuition for traction. While we might agree with Finnis that consequentialist reasoning does not properly account for the selection of life plans (or at least, struggles to do so), Finnis does not properly account for the sense we have that sometimes, or even quite often, consequentialist reasoning seems quite clearly to dictate what we ought to do.\(^{58}\) Recall the example we used in analyzing Ross’ view: on my way to meet a friend, I see someone who has been in an accident. I can choose to help the accident victim or choose to go meet my friend. Suppose further that it is important for my friend that I meet him on time, though it is not a matter of life or death. Even if I suspect that stopping to help the victim may harm my friendship somewhat, I may feel the obligation to stop. (Note also that this example does not require “acting against” the good of friendship in the sense objectionable to Finnis, at least assuming that he does indeed embrace the Doctrine of Double Effect.) Whatever Finnis says about this case, he cannot say that the balance of good objectively dictates what we ought to do. He

\(^{58}\) Obviously Finnis is aware that we feel like maximizing; he just feels that the impulse is misguided.
cannot even say what Ross does: namely, that saving the accident victim is more of a duty in this case than making it to the meeting with my friend. Finnis would have to appeal to various requirements of practical reasonableness that quite plainly mischaracterize the choice as we see it.

Worse yet, even if we grant that the basic human goods may be incommensurable, we have yet to account for the impermissibility of tradeoffs within a particular good. If, for example, “life” and “knowledge” are truly incommensurable, surely we can quantify life and knowledge independently, such that more lives are better than fewer, and more knowledge is better than less. This raises a sticky problem for Finnis because it seems quite clearly to give us a principled reason for picking certain states of affairs over others. Setting aside cases that call on us to “act against” a good, even on Finnis’ scheme it seems a stretch to say that we have no principled reason for picking the greater number of lives over the smaller.

Finnis’ response to this potential challenge is particularly inadequate. He claims that unrealized possibilities matter too—distant possibilities that we cannot conclusively evaluate. This is a more general version of Finnis’ problem with utilitarianism, which is that attempting to maximize even a single good is senseless because the ostensibly endless consequences of our actions elude us and cannot be factored into our calculations about which course of action is mathematically superior. It isn’t clear why this would be a good criticism of consequentializing, given that we can and often do distinguish between subjective and objective rightness. The right thing to do, all things considered, might be to bring about optimal states of affairs; but since we cannot be held accountable for unreasonably unforeseeable consequences, we might hold ourselves to standards of
subjective rightness: doing what we can reasonably conclude would promote the most
good.\textsuperscript{59}

Finnis’ idea is quite clearly to undermine the possibility of approaching maximization with any confidence at all. Yet there is just no reason to think that Finnis’ criticism even applies in the sort of case we are discussing. All else equal, if the reasonably foreseeable consequences of choosing one path will promote or preserve much more of one good than will the other, then we have a perfectly valid, principled reason for choosing the first option.

Put another way, Finnis’ response to this challenge entails incredibly counterintuitive results. Imagine a variation of the trolley problem: as a speeding trolley approaches a fork in the track, its driver realizes that the trolley’s brakes are out. Fortunately, contra the typical trolley case, he can control whether the trolley goes left or right at the fork. On the branch to the left, a single worker toils away, head bent low over the track. On the branch to the right, there are five workers, equally distracted. If we assume the certain death of all of the workers on whichever branch the driver selects, the common intuitive response is to say that the driver should aim to minimize casualties. Go left, and kill the one. If we accept Finnis’ view, however—thus denying the relevance of subjective rightness—it is a matter of indifference which branch the driver selects. After all, what if there are ten more workers on the left branch, out of sight beyond the horizon?

\textsuperscript{59} Finnis claims that the various limitations we place on the use of reasonable foreseeability are calibrated to matters of social commitment rather than consequences. (Finnis 1980, 117-118) That is entirely beside the point if we are defending the viability of consequentialism, for in doing that all we need it to point to the possibility of using a theory of subjective rightness paired with a consequentialist-oriented decision procedure (whether direct or indirect).
What if one of the five workers on the right is a future mass murderer? The driver simply cannot know that minimizing casualties at this particular instant will preserve life best in the long run.

On most moral views, none of that matters. Given what the driver knows, it seems perfectly uncontroversial that he has a principled reason to go left instead: for all that he knows, and all that he can reasonably be expected to know, going left will save four lives. If he neglected that reason on the grounds Finnis favors, and went right instead, he would be irresponsible at best, and morally depraved at worst. Anyone but Finnis would accuse the driver of hiding behind an indefensible agnosticism. This is a serious problem. Finnis says that it is unreasonable to choose a course of action on the grounds that your choice promotes more good than the alternative. Yet in this sort of case, it looks like as if Finnis has it exactly backwards.\(^6\)

Now, Finnis may resist this result. He suggests that, where damage is inevitable, we should favor preserving basic human goods to instrumental goods, and we should prefer causing less damage rather than more to the same instantiation of the same good. (Finnis 1980, 111) In short, Finnis accepts that we may be able to quantify the damage averted or inflicted on one particular instantiation of one particular good, for the purposes of choosing narrowly between courses of action that only affect that instantiation of only that

\(^6\) Tom Hurka makes the interesting point that conventional use of the Doctrine of Double Effect involves a weighing of consequences in deciding whether the foreseeable but unintended consequences of an action are proportionate to the gains it offers. That element is difficult to eliminate from the Doctrine because without it we open up the possibility of accepting extraordinarily negative foreseeable results for little positive gain. Finnis seems headed for one horn of this dilemma: either reject the proportionality element of the Doctrine of Double Effect or struggle to explain when a foreseeable harm is great enough to defeat our warrant for a given action.
good.

Yet that is not the sort of choice at stake for the trolley driver; he is not choosing between harming and killing the same individual. The trolley driver is facing different instantiations of the same good: one instantiation of human life on the left and five distinct instantiations of it on the right. If Finnis concedes that we can weigh the lives of the one worker against the lives of the five, then we clear the way for explicitly consequentialist reasoning in a nontrivial subset of cases, which is quite explicitly unacceptable to him, however intuitively appropriate it is for us.

While we may be able to describe this trolley case in a way that lets Finnis endorse the right outcome, that alone does not save Finnis because, first, the reason why we favor going left over going right is exactly the reason that Finnis wants to reject, and second, that sort of reasoning gives us a principled way of choosing courses of action. Finnis does not accept that there are principled reasons of this sort, which raises serious trouble for structure of his view. The trolley driver’s reasoning is most readily explicable in terms of quantifying the harm he stands to do by taking either fork of track. Any other explanation will ring hollow because, given the nature of the situation, the trolley driver will perform simple arithmetic in order to inform his choice.

It is worth noting that incommensuralists diverge on just which attitudes toward the various goods are acceptable. For example, T.D.J. Chappell has suggested a “Threefold Schema” for explaining the attitudes that one may take towards incommensurable goods: respect towards the good, interest in promoting the good, and interest in not violating the good. (Chappell, 74) These three attitudes are supposed to correspond respectively with pursuing good, doing good, and avoiding evil. (Chappell, 75) Respecting a good means
recognizing it as good—as when we can acknowledge the merits in someone's life choices without wishing to make those choices ourselves. Pursuing the good better describes what we do when we make certain (laudable) choices to bring various goods into our lives rather than others. Finally, interest in not violating the good is self-explanatory: a negative attitude toward the prospect of acting against a good.

For Chappell, respecting the good and promoting it are always permissible, whereas violating the good is always impermissible. He claims that positing three different attitudes (instead of merely one, as preferred by Finnis) simultaneously lets us accommodate both supererogation and moral absolutes. The distinction between pursuing and respecting the good accounts for the former (for choosing to promote the good rather than respect it can be supererogatory), while the distinction between respecting the good and not violating it makes space for the latter. Chappell's argument for moral absolutes is not particularly useful for our purposes here, as none of these modifications to the incommensurationalist view will allow it to escape the trouble with intra-good tradeoffs.

In any case, the effectiveness of an appeal to incommensurability, were it successful, lies in the power of the answer it gives to the obvious consequentialist question: why shouldn't we maximize? If the answer is simply that we can't or that it makes no sense even to try, then we might have good resources for grounding agent-centered restrictions. Further, one could, in theory, appeal to incommensurability regardless of the list of goods on the table (so long as there are more than one).

On the other hand, there is a striking tension in accepting that value sits in states of affairs while also accepting that the goods are incommensurable. These two positions are not logically contradictory, of course, but there is something odd in saying that we know
there is value in states of affairs but we have no idea whatsoever how much (either in absolute terms or relative to other possible states of affairs). One of the most intuitive moral judgments to make is that certain outcomes preserve or promote more good than others.

In the first instance, then, it is a mixed result for the Neo-Thomist that incommensurability is not especially compelling. If we discard incommensurability, we lose one of the strongest replies to the consequentialist. Once we arrive at the conclusion that weighing consequences is useful for making moral decisions—at least sometimes—then we open the door to legitimate and powerful questions about why we may not violate goods in the interest of promoting superior outcomes. But for the reasons just discussed, rejecting incommensurability also gives us the chance to adopt a reply that is more consonant with the underlying assumptions of the Neo-Thomistic view. The next step is to inquire about possible alternatives. Admitting that value sits in states of affairs gives away half of the game to the consequentialist, and only a strong bulwark can keep us from sliding all the way out of deontology. With that in mind, what options remain?

In the abstract, it sounds plausible to say that violating a good for the purpose of promoting more good, where prohibited, is bad because it means we are adopting an impermissible attitude of sorts. More narrowly, it is perfectly coherent to say that there are cases where it is morally wrong to violate a good, even where doing so would conduce to a state of affairs that contains a greater volume of good. Finnis would say that such violations are always inappropriate, and in the previous chapter I suggested that such a position is implausibly strong. Imagine that we soften Finnis’ position somewhat, taking a line like Chappell’s instead.
We could agree that it is not always wrong to violate a good, and we could try to come up with a principled reason for demarcating the range of cases where we feel that such violations are actually inappropriate. In doing so, we would be free to pick among almost any conceivable rationale that tracks the objective badness of committing those violations. Whatever rationale we use, we would say that the badness in question obligates us to treat the goods in those cases with something like respect. So long as we do not accept an attitude of acting against the goods, then we will have some form of deontology. The practical effect of banning such attitudes is to establish something with the formal features of an agent-centered restriction, blocking the sacrifice of some measure of good even in the interest of promoting more. What makes a restriction agent-centered, on a view of this sort, is that each of us has our own duty not to harm various instantiations of the goods regardless of the purpose for doing so.

There are two significant problems with this, the Neo-Thomistic approach to constraints. The first is that once we’ve stripped out incommensurability and the relevance of indeterminacy, the plausible options for justifying an attitude of respect toward the goods are rather limited. It is easy to invent an ad hoc attitude that validates our intuitions about when we may and may not violate the good, but the hard part is grounding that attitude. Once we strike incommensurability and indeterminacy from Neo-Thomism, we are left with some bare injunction to respect the good. As believable as that requirement may be, without some sort of basis, it is also no more than a hackneyed expression of common moral intuitions. A consequentialist might very well ask why respecting the good is inconsistent with maximizing it, and we want to have an answer more robust than a declaration by fiat.
One possibility might be to use respect for persons. The difference between simply respecting the good and respecting persons, for many people at any rate, flows from the fact that the good is not typically considered just to be equivalent to persons. Persons are more generally considered a locus for good, which may seem to make respect for persons more promising as a means of justifying constraints by preventing tradeoffs. Yet the consequentialist can still ask why respecting persons requires doing anything but treating their interests equally, as an agent-neutral state-of-affair valuation is bound to do. Kant's second formulation also appeals to something like respect for persons, but his view comes out looking quite different because he essentially claims that we should respect persons because persons house value. Finnis has already accepted that value resides in states of affairs, so the principle requiring respect for persons is not as easy to ground—especially in the face of the consequentialist's possible reply that treating people agent-neutrally is exactly what it means to respect them.

Various further difficulties present themselves depending on what sorts of constraints we want to justify. A solitary principle of respect for persons, even understood in a deontologically-friendly way, seems ill-suited to justifying non-absolutist deontology, which many find more intuitively plausible than absolutist alternatives. Such a principle will struggle to draw a line between cases where we want constraints to stand and cases where we want them to break. The effectiveness of an appeal to respect for persons also depends on what items make their way onto the list of goods, and how those goods relate to persons.

The bigger issue remains, however, that Neo-Thomists need a particularly strong justification to resist the consequentialist and, without incommensurability, there is not
much available except bare stipulation. Though this may be controversial, there seem to me two different kinds of replies (broadly speaking) that would do the work we need. One would be a deontologically-oriented normative principle that is as powerful as the one in favor of consequentialism; the other would be to uncover a particular fact of normative relevance that tempers the consequentialist’s conclusions—something on the order of incommensurability, but more plausible.

Our prospects on the first front are not particularly good. Everyone has deontological intuitions, but no one has yet expressed those intuitions in the sort of normative principle we would like. That was one of the criticisms of Ross raised in the first chapter; asserting deontological limits without giving a deeper explanation of them will leave a justificatory asymmetry between the consequentialist element of our view and the deontological one. Recall: there is a pressure to maximize the good because it is compelling to think that, all else equal, more good is better than less. The pressure to “respect” the good in Finnis’ sense has no analogous ground. At the very least, no such ground has been identified.

The second option is to identify a fact or set of facts that we have reason to regard as normatively significant. Incommensurability would be exactly that sort of fact, and indeterminacy might be another one—for both have the potential to undermine consequentialist commands. I have argued that incommensurability is false, and that indeterminacy, to the extent that it’s true, is not normatively significant because we typically and appropriately tend to apply subjective theories of the right. Aside from these
two invalidated options, no other such fact consistent with the Neo-Thomistic approach stands out to me.\textsuperscript{61}

Even if we were to find the sort of fact we need, however, there is a separate problem—the second major problem with a Neo-Thomistic approach to deontological constraints referenced above. More specifically, on the sort of attitudinal approach advocated by Finnis, the restrictions imposed on us thus emanate entirely from us. For any given good—whether another’s life or my own knowledge—the reason I may not intend to harm it is that I am bound, somehow, to take an attitude of respect towards it. Such an arrangement is not ideal for the simple reason that normally we think at least some element of constraints derives in a meaningful way from outside of us; merely stipulating that we cannot violate the goods, however, does not give us that element of the constraint in any meaningful way.

The point becomes clearer when we consider whether an intrinsic good just is something that cannot legitimately be violated. If pleasure is a good, as many people think it is, it does not seem to follow that it is morally wrong (for example) to cut off or act against our own pleasure for the purpose of focusing on some other worthwhile project. Or, to draw from Finnis’ list of goods, it is not convincing to think that intending the harm of our opportunities to engage in play or social experience is necessarily wrong, especially if done for noble purposes.

Put more clearly, the problem is that deontological intuitions do not support the view that constraints derive from the attitudes we must take toward the goods. At the very

\textsuperscript{61} There clearly remains the possibility that the relevant facts have eluded me.
least, if we are to get a compelling account of constraints by banning certain attitudes
toward the goods, we will need a variegated set of attitudes to capture the varieties in the
ways we find it permissible to interact with different goods in different situations.

Sometimes we find it permissible to violate a good and sometimes we find it
impermissible. Sometimes we find trades acceptable and sometimes we don’t. The verdict
in any given case depends less on which good we have in mind than on the situation we
find ourselves in. If correct, this observation pushes away from an account built around a
solitary attitude (like Finnis’) to a more nuanced account with several (such as Chappell’s).

Yet even a complex and well tailored account of such attitudes would not solve the
fundamental problem because we would find ourselves back where we started, asking the
question of why we should take any given attitude to the various goods in one situation or
another if not bare intuition.

There remains one promising resource that may allow us to solve both of these
problems. Specifically, I have in mind the distinction between the agent-relative and
agent-neutral points of view. The agent-relative point of view, as I understand it, reflects
the state-of-affairs valuations given by each agent to various possible outcomes. The
agent-neutral point of view is the view from which all people and their preferences get
equal weight.

Others have made use of this distinction before, but not in the way that would most
forcefully address the deontological pitfalls we have uncovered in these three chapters.
From that distinction we may pull a psychological fact, which I follow Samuel Scheffler in
calling the “natural independence of the personal point of view.” Such a fact, if true, could
serve to replace Finnis’ assertions to incommensurability and indeterminacy. The natural
independence of the personal point of view is the unwieldy name for (what I will argue is) the fact that agents naturally (and inevitably), from their own points of view, inflate and deflate the value of states of affairs out of proportion to what they are permitted according to the agent-neutral point of view. So long as we can show that this is plausibly a fact and that it is of normative significance, natural independence stands to anchor deontological devices in the face of consequentialist demands.

Further, natural independence appears to offer a more promising method of placing constraints around various goods. Instead of positing attitudes that we may or must take toward the goods themselves, we might be able to locate constraints by differentiating the various attitudes we could or should take toward the different points of view. The rough idea behind natural independence is that we cannot help but to have disproportionate interest in our own affairs, and this skews our state-of-affair rankings. If that is true, goods may become inviolable based more on how significant they are relative to salient alternatives or whether another person has a legitimate claim to them rather than whether they are good in the first place. This would be one of a several welcome results. It is plausible to think that we generally have more license to violate a good in our own lives than to violate a good in someone else’s. For example, I may be free to act against my own pleasure so I can focus but it is not so clear that I can trade off your pleasure. This is exactly the sort of observation that Finnis is not equipped to handle because he pins the deontological attitudes to goods rather than state-of-affairs valuations.
Conclusion

Neo-Thomism shares more with consequentialism than do either intuitionism or Kantianism. Neo-Thomists accept both the claim that our duties are defined in terms of value and the claim that value resides in states of affairs. Intuitionists and Kantians split on these two claims, with the former denying that all of our duties are explicable in terms of value and the latter denying that value sits in states of affairs. All three deontologies, of course, reject the third consequentialist claim, which asserts that our duty with respect to value is to promote it; but Neo-Thomism is the last of our three deontologies to get off of the consequentialist bus, so to speak.

There are advantages to accepting both the first and second of the consequentialist claims. Ross’ view, as we have seen, invites the question of why some of our duties can be explained as responses to value when others cannot. Because Ross does not reduce his duties to a unified base, he is unable to provide any deeper explanation; as a result, he leaves hanging a reasonable question we might have seeking a reason for the dichotomy between our duties. (We might want to know why our duties do respond to value, but that notion seems less curious than the notion that they do not.) Neo-Thomism thus dodges at least one set of challenges because it adheres to the first consequentialist claim.

Kantianism, on the other hand, creates all sorts of difficulties for itself by denying that value sits in states of affairs. The Kantian position on that question makes it impossible to count consequences as relevant in moral deliberation; and as a result, Kantianism entirely lacks the resources for a plausible account of the stringency of constraints, or at least a plausible account of the relative wrongness of different acts. Here, again, Neo-Thomism is immune from a serious class of concerns.
Like intuitionism and Kantianism, however, Neo-Thomism denies the third consequentialist claim. Neo-Thomists do not accept that we must promote value wherever we find it, and their reason is that committing to do so involves adopting an illicit attitude toward the good. I argued that the typical Neo-Thomistic justifications for those attitudes fail, and that even if they could be justified on other grounds, we would be left with poorly tailored constraints that are not intuitively acceptable. I have also suggested that there is no legitimate normative principle that requires us to abstain from trading off intrinsic goods just because they are intrinsic goods. At the very least, that sort of principle would not find broad intuitive support.

In the next two chapters, we will explore the possibility of a justification for agent-centered restrictions that avoids the problems we have identified in our review of the three prominent deontological approaches. Relying on the natural independence of the personal point of view, I will argue that Rossian, Kantian and Neo-Thomistic problems are not intractable, so long as we carefully built our deontology from the right components.
Tapping The Natural Independence of Human Agency

Introduction

Thus far, we have focused on prominent conventional methods of explaining constraints, starting with W.D. Ross’ intuitionism, moving on to Kant’s ethics, and finishing with John Finnis’ Neo-Thomistic natural law theory. I have argued that each of these three approaches is seriously lacking in at least one respect or another, and that we ought therefore to search for a new basis for constraints. In the previous chapter, however, I indicated an affinity for the Neo-Thomistic approach because it accepts the first two of the consequentialist’s propositions: that our duties are defined in terms of value and that value sits in states of affairs. Accepting this pair of assumptions allows the Neo-Thomist to avoid some particularly uncomfortable challenges. The problems with the Neo-Thomistic approach are just that it offers a weak explanation of why we ought to reject the third consequentialist claim, which advocates for the straight up maximization of the good, and it offers constraints that neither arise from nor end up in the right places.

In this chapter I will try to develop a novel foundation for deontology that avoids the problems that haunt Intuitionism, Kantianism and Neo-Thomism. I thus begin by accepting that our duties are defined in terms of value and that value sits in states of affairs. As a result, I will also remain agnostic about various theories of the good, though I
may occasionally indicate the implications of my approach towards the right for proponents of particular views about the good.

Very roughly summarized, the account that I wish to develop draws on observations made by Henry Sidgwick, Thomas Nagel and Samuel Scheffler. Each of these authors has gestured at a promising, core foundation for duties—including those responsive to agent-centered restrictions. To varying degrees, all indicate the joint significance of the agent-relative and the agent-neutral points of view. I will argue, however, that none of them have taken full advantage of what that foundation has to offer.

With that in mind, I will begin by highlighting how each of these authors has pointed in the direction of a new and effective justification of duties, as well as indicating where I diverge from their paths. Once that task is complete, I will move on to developing a fuller positive proposal, and I will conclude this chapter by illustrating how the new approach is different from the approach offered by Frances Kamm.

4.1. Background

Over a century ago, in The Methods of Ethics, Henry Sidgwick presaged an account of the sort that I have in mind. He claimed that there is an irresolvable conflict between egoism and universal hedonism, which leads to what he calls the “dualism of practical reason.” Sidgwick’s idea was that there is no way of persuading someone committed to egoism that he ought to step back from his self-regarding position and care about the

62 I will wait to address Scheffler a bit later because he raises a challenge to my approach that I can only deal with once I have outlined my position.
pleasure of all instead. There is, according to Sidgwick, a deep (and in some sense unbridgeable) chasm between a purely self-regarding point of view and a neutral, other-regarding point of view. Now, Sidgwick was a utilitarian, so he described this tension in terms of pleasure. In defending the relevance of the agent-centered and the agent-neutral point of view, I do not restrict my discussion solely to agent-relative and agent-neutral angles on the pursuit of pleasure. But that is a cosmetic difference, as far as Sidgwick’s view concerns us here.

The key point is that the structural tension between the two perspectives of practical reason identified by Sidgwick maps quite closely to the tug-of-war between agent-neutral and agent-relative considerations that I have in mind. (While I do not see agent-relativity as being entirely egoistic, the similarities are striking.) The agent-relative and agent-neutral point of view each offer different rankings of states of affairs; considerations relevant to one are not necessarily responsive to the concerns of the other. Deciding what to do, as a practical matter, often (though not always) involves deciding how to balance these competing sorts of considerations.

Moreover, Sidgwick seems to recognize the “intuitive” primacy of the egoistic or (loosely) agent-relative perspective: “Indeed, it is hardly going too far to say that common sense assumes that ‘interested’ actions, tending to promote the agent’s happiness, are prima facie reasonable: and that the onus probandi lies with those who maintain that disinterested conduct, as such, is reasonable.” (Sidgwick, 120) This observation also finds an analog in the approach I outline below. In general, though, the particulars of

63 Ultimately, Sidgwick rejected the relevance of commonsense moral judgments about concrete cases because those judgments fail his preferred test of “self-evidence”—the test
Sidgwick’s view are not overly relevant to the approach I defend below. The key point is that Sidgwick’s view establishes a historical precedent for asserting a fundamental, psychological duality of perspectives.

It is Thomas Nagel who, in The View from Nowhere, gives the first prominent description of the actual distinction I have in mind.64 Crucially, Nagel also speaks in terms of the duality of perspectives, though unlike Sidgwick, Nagel uses the specific perspectives I prefer—the distinction between the agent-neutral and the agent-relative. (I will, however, diverge appreciably from Nagel in the definition of agent-relativity.) The title of Nagel’s book has come to represent perhaps the most famous description of the agent-neutral point of view: the valuation of states of affairs we get if we hover above agents, assigning equal value to each for the purposes of guiding human action. The agent-relative point of view, on the other hand, takes on the first-person perspective of agents, and it offers different sorts of reasons for action from those offered by agent-neutrality. On Nagel’s view, at least some agent-relative reasons arise directly out of the personal preferences of agents, while some exist independently of the personal point of view but are not agent-neutral. Nagel claims that the former ground options and the latter ground constraints.

Nagel begins the relevant excerpt by defining his intention to defend “objectivity in ethics.” (Nagel, 142) He proceeds to outline three classes of reasons that might inject

he endorsed as crucial in coming to the proper moral conclusions. This particular claim does not accord with the approach I endorse below.

64 The excerpt from Nagel’s book that I’ll be using was reprinted under the title of “Autonomy and Deontology.”
agent-relative considerations into a moral theory. The fact that these reasons grip us
reflects our psychological constitution, and Nagel acknowledges that the baseline
assumption he works from is that “the shape of a moral theory depends on the interplay of
forces in the psychic economy of complex rational beings.” (Ibid)

The three types of reasons he gives are as follows:

The first type of reason stems from the desires, projects, commitments, and personal ties of the individual agent, all of
which give him reasons to act in the pursuit of ends that are his own. These I shall collect under the general heading of
reasons of autonomy.... The second type of reason stems from the claims of other persons not to be maltreated in certain
ways. What I have in mind are not neutral reasons for everyone to bring it about that no one is maltreated, but
relative reasons for each individual not to maltreat others himself, in his dealings with them.... These I shall collect under
the general, ugly, and familiar heading of deontology.... The third type of reason stems from the special obligations we
have toward those to whom we are closely related: parents, children, spouses, siblings, fellow members of a community or
even a nation.... I’ll refer to them as reasons of obligation.
(Nagel, 143-144)

Nagel is less confident in the existence of his third category (reasons of obligation) than in
the other two; but he thinks that these three classes of agent-relative reasons, when
combined with agent-neutral reasons, “cover much of the territory of unreflective
bourgeois morality.” (Nagel, 144) While Nagel acknowledges attempts to account for each
of these agent-relative classes of reasons in utilitarian or consequentialist systems, he
claims that those attempts only capture part of the truth. Instead, he wants to offer an
independent and agent-relative story that explains the validity of such reasons.\textsuperscript{65}

The most significant feature of Nagel’s view, as far as it concerns us presently, is not the distinction between types of agent-relative considerations. I have reproduced that discussion mostly to illustrate the scope of reasons that Nagel considers to be agent-relative, but I will offer a more detailed account of agent-relativity below that I think may allow us to collapse at least two of Nagel’s categories into one. The more important point is that Nagel sees a good theory of the right as one that recognizes the balancing act between agent-neutral and agent-relative considerations. As Nagel puts it, “I believe the human duality of perspectives is too deep for us reasonably to hope to overcome it. A fully agent-neutral morality is not a plausible goal. On the other hand, it is conceivable that deontological restrictions now widely accepted may be modified under the pressure of conflict with the impersonal standpoint.” (Nagel, 168)

Further, though Nagel does not offer a full theory of the right grounded in this distinction, his reasons for accepting the distinction appear to me to be good ones. First, Nagel accepts the relevance of agent-relative considerations because, he claims, “A thoroughly impersonal morality would require that victims as well as actors be dominated by impersonal, agent-neutral values in their judgments about how others treat them... [and] this seems an excessive demand to make of individuals whose perspective on the world is inherently complex and includes a strong subjective component.” (Nagel, 166) That seems right to me, and I will elaborate on why below. At the same time, Nagel points out that, “…in defending the legitimacy of agent-relative principles, we must guard against

\textsuperscript{65} Nagel focuses on autonomy and deontology, leaving the category of obligation aside after mentioning it at the beginning.
self-deception and the escalation of personal claims simply to resist burdensome moral demands. “(Nagel, 170) Again, this is a fair point, and one I will address in some detail as we proceed.66

4.2. Developing the New Approach

Grounding a theory of the right in the distinction between the agent-relative and agent-neutral requires addressing two main questions: why we should think that each perspective matters, and what kind of extensionality can we derive from that bipartite base? I think we can give compelling answers to both of these questions, so let me take each in turn.

As noted above, I follow Nagel in defining the agent-neutral and I follow him partially in defining the agent-relative points of view. I understand the agent-neutral point of view very much as “the view from nowhere”—the objective stance from which we simply sum up the hypothetical volume of good in the world in order to evaluate and rank possible states of affairs, irrespective of personal preferences about how that good is configured in our own favor. This is the view from which a rational being can understand that his life is not necessarily worth more than any other.

The agent-relative point of view is a bit more elusive; and here I think Nagel and I disagree in a rather significant way. Nagel seems to regard the agent-relative point of view as encompassing a set of three reason types (autonomy, deontology and obligation, all listed above). We can leave aside reasons of obligation because Nagel remains unsure

66 Though I think Nagel is onto something here, it will become apparent that I employ some of his claims differently from how he himself does. I will flag points of divergence as we proceed.
about whether they actually exist. At the very least, we might say, Nagel confidently distinguishes reasons of autonomy from reasons of deontology.

Reasons of autonomy are what Nagel uses to justify something like options (freedom to pursue some suboptimal outcomes). Nagel offers an example of such a reason in his famous example about climbing Mount Kilimanjaro. (Nagel, 145) Specifically, my desire to (say) climb Mount Kilimanjaro gives me a reason to do so. That reason, however, does not apply to others. It does not generally give them a reason to climb the mountain with me or, more importantly, to care terribly whether I succeed when I try on my own.

Reasons of autonomy, in short, are reasons for action a given person may have that flow directly from his preferences for certain sub-optimal states of affairs. Reasons of deontology, according to Nagel, constrain us from doing certain things—not because doing them would be objectively worse than not doing them, but because each of us may have a duty not to mistreat others in our dealings with them. So if I should not kill an innocent man to save three others, then the reason for Nagel is a reason of deontology.

On my view, what distinguishes the agent-relative point of view from the agent-neutral is that the agent-relative involves privileging one’s own interests. I do not mean “interests” to be synonymous with “self-interest” in an egoistic sense, but rather the aims and preferences we have for the success of people and projects closest to us. This is the standpoint from which most of us demonstrate a considered and overwhelming preference for configurations of goods in the world that advantage us and those we care most about.

My understanding of the agent-relative point of view thus aligns most closely with Nagel’s “reasons of autonomy,” and I will go on to defend that understanding below. The
purpose of noting where I diverge from Nagel is to highlight the fact, in terms of Nagel’s framework, I view the agent-relative point of view quite strictly as the point of view from which agents confer the value of personal preference on certain sub-optimal states of affairs. As a result I will be using something like a reason of autonomy (to use Nagel’s language) in an attempt to establish constraints, and I will not even argue for a category called “reasons of deontology.”

The agent-relative and the agent-neutral, on this view, are defined in terms of each other, rather than (as Nagel has it) the agent-relative being something of a catchall for reasons that aren’t agent-neutral. I take each of these points of view as basic; that is, neither can be reduced to the other. They are simply two different angles from which agents can evaluate and rank possible states of affairs.

The relevance of the agent-neutral point of view is the simplest to defend, so I shall begin there. Part of what makes it simpler is that very few are willing altogether to deny that we have moral obligations to promote agent-neutral good, at least sometimes. Those few who do exist—Shelly Kagan calls them “minimalists”—occupy an array of positions, ranging from the egoist to the hardcore libertarian; but relatively speaking, only a small minority stakes out a position on that end of the spectrum.67 It is far more significant for my project here that consequentialists and many deontologists endorse a general duty to promote states of affairs that are better from an impersonal standpoint—and crucially, because those states of affairs are better from an impersonal standpoint. Thus the objections I am most concerned with regarding the justification for options and

67 Here I borrow liberally from Kagan’s framing of the various positions, in large part because I will object to some of his arguments below.
constraints come largely from those who find the agent-neutral point of view morally relevant already.

One of the most compelling arguments for such a duty is that, when we abstract from our own interests and view moral matters dispassionately, we can quite clearly see the sense in assigning equal weight to every life and its interests. Objectively speaking, and all else equal, it seems as if fewer deaths are better than more, more happiness is better than less, and so on. We constantly employ the same reasoning within our own lives, for it is pure and potent: if anything is objectively good, it seems like there will be at least some cases where doing the right thing will involve promoting more of that good rather than less of it for the specific reason that more is better than less. (Likewise if we switch out “good” for “evil” and “promote” with “minimize.”) While there are various ways to try to soften or resist that conclusion—some of which we have canvassed in previous chapters—none seem capable of denying that there is some sense in objective aggregation. Claims about incommensurability appear overblown, and denying that states of affairs contain value is not entirely convincing. In the end, I suspect that there is no viable way of resisting agent-neutral pressures entirely.

Nevertheless, I think there are legitimate reasons for tempering those pressures by admitting the relevance of agent-relative reasoning. As Samuel Scheffler has observed, our own personal point of view stands naturally independent from the agent-neutral position that yields various forms of consequentialism.

From the impersonal standpoint, the fact that some person has a special kind of concern for his own projects and plans is relevant only in so far as it may affect the size of the harm or benefit incurred by that person when those projects fare poorly or well. Once that harm or benefit, however small or
great, has been figured into the overall evaluation, the impersonal standpoint has no further concern with his projects. But his point of view is independent of the impersonal point of view, and this independence is typically evidenced by the fact that he cares differentially about his projects just because they are his projects.... His own projects and commitments have a distinctive claim on his attention; he cares about them out of proportion to the relative weight carried in the impersonal calculus by having and caring about them.... (Scheffler, 56-57, emphasis in original)

Scheffler is basically describing (one dimension of) a universal and demonstrable fact about human agency. An agent’s perspective just is naturally independent of the impersonal standpoint, which leads us, among other things, to overvalue our own projects. “Overvaluing” of this sort amounts to preferring sub-optimal states of affairs—states that do not necessarily contain the greatest volume of good—to those that are agent-neutrally optimal. In most cases, it is not that we mistakenly believe that states of affairs that privilege our interests are agent-neutrally optimal. Rather (and rightly or wrongly), agents simply do not care exclusively about agent-neutral optimality.

To take a more concrete example, imagine a hypothetical hostage situation involving two hostages and one gunman. Trapped in a building, surrounded by police, the gunman has become desperate. In a last-ditch attempt to extract concessions from the hostage negotiator, the gunman has elected to execute one of his two hostages; but he has no particular preference about which hostage to kill. Let us imagine, plausibly enough, that neither hostage wishes to die. Each begs for her life, knowing that the success of her own pleas will entail the death of the other.

Let us also stipulate that the hostages know nothing about each other. In a sudden moment of abject terror, neither is likely to ask for mercy on the grounds that each
believes herself to be “objectively” more valuable to the world. It is much more likely that each simply and viscerally prefers her own survival. Each hostage vastly overvalues her own life in comparison with the life of the other. The first hostage very strongly prefers the state of affairs in which the second dies, while the second much prefers the state in which the first dies—and both of these preferences track nothing more than that each hostage confers an enormous amount of agent-relative good on her own life and almost none on the life of the other. Moreover, even if each hostage somehow believed that the survival of the other would be better for the world, it would be shocking if that belief resulted in either volunteering for execution. In the mind of each hostage, the expected deficit in objective, agent-neutral good of her own survival does not outweigh the enormous agent-relative value to be had in the state of affairs in which she survives.\(^{68}\) In short, each hostage defaults to the agent-relative point of view; and that point of view, as Scheffler asserts, is naturally independent of the agent-neutral point of view.

More importantly, however, the fact of “natural independence” is one that we seem to endorse morally at a visceral or intuitive level. Not only do most agents favor their own projects, but that they do so just seems morally acceptable. Except in rare and highly anomalous cases (such as that of Zell Kravinsky\(^ {69}\)), agents naturally tend to prefer the

\(^{68}\) At the same time, neither is likely to expect the other to volunteer for execution, fully recognizing that they each attribute agent-relative value to themselves and not to the other.

\(^{69}\) Zell Kravinsky is famous for attempting to live his life by making decisions from an agent-neutral point of view. He voluntarily gave the vast majority of his wealth away, reducing himself and his family from being multimillionaires to living like members of the working class. More radically, he has given one of his kidneys away and allegedly claimed that he would donate his second if the recipient would contribute more good to the world than he could himself.
wellbeing of their own families to that of strangers, the advancement of their own careers to those of others, and so on.\textsuperscript{70} And while there are plenty of instances where we expect agents to sacrifice for others, we do not generally condemn people for possessing the disposition to put (at least certain) personal projects first. Nor do we condemn them in the majority of the cases where they act on that disposition.

This independence is exceptionally resilient. Some of our personal projects can take on a value that seems \textit{drastically} inflated from an agent-neutral point of view. Silliness notwithstanding, an episode of the animated TV show \textit{Futurama} provides a terrific example of this. The show tracks the life of a pizza delivery boy named Fry, who was cryogenically frozen in contemporary times and subsequently revived in the distant future. In an episode appropriately entitled “The Neutral Planet,” Fry, who had always believed he was frozen accidentally, learns instead that he was frozen intentionally by an alien who traveled back in time specifically to preserve him. The aliens responsible for the plot had discerned that Fry was the only one who could save the universe a millennium after his natural life would have expired, so they sent one of their own back to ensure his survival for the next thousand years. But, by freezing him as a young man, the aliens severed Fry’s ties to his old life, causing him no small amount of emotional anguish once he was revived. “We had no choice,” the aliens claim in their defense, “You were the only

\textsuperscript{70} There appears to be a distinction between preferring state of affairs that are good for me, and preferring states of affairs that are good for my family: if I may act on the former preferences, doing so would seem to be largely discretionary, while acting on the latter may be a matter of agent-relative duty. I accept that distinction, but my view is that what makes \textit{both} sorts of preferences the sort that I can pursue is my agent-relative overvaluation of them; and what makes it a duty for me to support my family is a feature of what they can expect of me as the type of agent who does and should overvalue certain projects. I’m grateful to Tom Hurka for highlighting this distinction.
one who could help us. What is one life weighed against the entire universe?” Lip quivering and eyes welling with tears, Fry responds, “But it was my life.”

Fry’s response in no way disputes the aliens’ impersonal calculus. Instead, Fry seems to be talking past the aliens by giving an agent-relative response to an agent-neutral value claim. But we should note that our initial response to the example is one of compassion for Fry—or at the very least, understanding for his emotional reaction. Fry’s reaction seems natural and, while it is ridiculous from the agent-neutral standpoint, we are sympathetic. For better or for worse, it is impossible to deny that the natural independence of the personal point of view has a truly powerful effect on our moral judgments.

While I have just given a single, fictional example, the ubiquity of the intuitions that support the natural independence of the personal point of view is more or less beyond dispute. As noted previously, even Scheffler and Kagan will not contest that. But the truth of such a claim about agency, even if granted, does not (by itself) yield the conclusion that agent-relativity deserves a place in the structure of a theory of the right. At this juncture, the problem is to explain why that natural independence of the personal point of view is normatively significant.

Before we proceed, however, a qualifier is in order. There is a significant literature debating the semantics of the phrase “good for,” and some even question the coherence of the notion that a good can somehow be “located within” a person’s life. I do not intend to delve into that literature here, but I believe that my understanding of this phenomenon at the very least remains pliable enough to be fitted onto a variety of the semantic accounts. I also, quite plainly, reject the notion that a good cannot be located within a specific
person’s life. As I understand it, a good is in a person’s life when he has a disproportionate
interest in it—disproportionate, that is, when compared to its value from the agent-neutral perspective.\footnote{This formulation is a bit vague—partly because it is difficult to come up with a more
specific description that is flexible enough to capture the way in which people relate to the
variety of goods in their lives. Let us imagine that knowledge is a good. I may know a
particular fact—say, that the color of a flamingo’s feathers depends heavily on the protein
and plankton content of its diet—but I may not care about that fact disproportionately. (I
may think it is uninteresting information, or unimportant, or whatever.) Still, I think that it
is plausible to say that I have a disproportionate interest in retaining this fact, simply
because it constitutes a larger proportion of my total knowledge than of the overall body
of facts in the universe.}

For example, I have a puppy. Because I love him and take care of him, the happiness
of my puppy is a good located in my life. His happiness matters more to me than it
should—and gives me more pleasure than it would—if I were to follow a strictly neutral,
objective valuation of the happiness of a single, young pug. I do not see this as a
metaphysical claim about the abstract physical location of goodness; instead, it is a claim
about how we rank and valuate states of affairs from our own personal point of view.

In any case, unlike Scheffler, Kagan quite directly resists the conclusion that agent-
relativity demands recognition in a moral theory, so we must confront his argument
before moving on. It is worth noting that, for Kagan, acceding to the natural independence
of the personal point of view would give us a reason to justify options; but it is not so clear
from Kagan’s writing whether he thinks that natural independence can justify constraints.
(Kagan 1989, 266) As such, he writes quite explicitly about whether options (which,
according to him, can come out of that independence) help to reflect the personal point of
view. He does not, however, mention constraints in his discussion of natural
independence. Below I will argue that both options and constraints derive from natural independence, so I dispute that omission, but I will cite him accurately here and make the case for the joint justification of options and constraints later.

Kagan makes his point by presenting a hypothetical argument between a “moderate” (someone who accepts options and constraints, as well as some sort of general pro tanto reason to promote the good) and an “extremist” (who accepts the duty to promote the good, but rejects options and constraints). In the given context, it is more or less accurate to describe me as a moderate and him as an extremist—though, lest these terms appear tendentious, I emphasize that they are his labels for the respective views and not mine. Slightly restated, the argument he presents on behalf of the moderate is that we all generally prefer to favor our own interests, so no moral theory can fully reflect the nature of persons unless it includes options (which reflect the agent-relative point of view in allowing us to pursue our own interests at the expense of the greater good).

Kagan is happy to concede this point to the moderate because he thinks it is of no significant consequence. His response is that a moral theory should minimally reflect such facts, but need not do so fully. According to Kagan, a system minimally reflects a set of facts “if it does only what it must about those facts.” (Kagan 1989, 261-262, emphasis in original) Meanwhile, a system fully reflects a set of facts “if it does all that it can.” (Kagan 1989, 262, emphasis in original) This bare description still relies on some normative standard of what a theory “must” or “ought to” reflect:

“We might say that a set of facts often implicitly involves both demands upon a system and appeals to it. If the system recognizes the facts, the demands cannot be neglected; and provided that they are met, the system can be said to minimally reflect the facts. The appeals, however, can be
neglected; the mere recognition of the facts does not force the system to comply with the appeals. To fully reflect the facts, however, the system must go beyond what is forced upon it: the appeals too must be met.” (Kagan 1989, 264, emphasis in original)

The “extremist” response to the argument Kagan presented above, then, is that accepting the natural independence of the personal point of view as a causal explanation for why people fail to promote the good as much as they could (or, according to many consequentialists, as much as they should) is sufficient for minimally reflecting natural independence. Accepting options, on the other hand, would involve fully reflecting natural independence, and thus is not incumbent upon the extremist. In other words, Kagan denies that crude consequentialist (or extremist) theories “fail to minimally reflect the nature of persons.” (Kagan 1989, 265)

Still, Kagan realizes that, by his own lights, it is possible that the extremist’s position on natural independence is mistaken—in particular, if it can be shown that natural independence “necessitates certain features” that are lacking from the extremist’s system. (In this case, those features would obviously include options—and, as I will argue later, constraints.) Kagan says he does not see how natural independence could necessitate such features, and he claims that the moderate “has given no reason to think otherwise.”

On behalf of the moderate, I will now attempt to provide such a reason. Kagan appears to treat natural independence as something like a preference—albeit, a very strong one. Assuming the voice of a hypothetical moderate, he writes, “The [extremist’s largely unmitigated] requirement to promote the good leaves the agent no room to favor what he wants to favor, nor does it permit him to justify any of his decisions simply
through an appeal to his \textit{wants}.” (Kagan 1989, 263, emphasis mine) Instead, such a requirement “demands that agents devote only that amount of energy to satisfying their \textit{wants} which those \textit{wants} could command from the objective point of view.” (\textit{Ibid}, emphasis mine)

Kagan quite explicitly cashes out natural independence in terms of desires. While that is common, and although it captures one element of natural independence, such phrasing can also be dangerously misleading. It is true that the natural independence of the personal point of view is characterized by the agent-neutrally exaggerated value placed on each individual person’s interests—or “wants” as Kagan calls them here. But it is a mistake to infer from such language that natural independence refers to nothing more than a unique basket of preferences, calibrated to the idiosyncrasies of each given individual. The key difference, which I will elaborate on below, is that while the contents of one’s basket are generally contingent, the fact that one carries the basket in the first place is a matter of necessity.

One of the biggest worries about agent-relativity seems to be that people have preferences for all sorts of different things, some of which are good and some of which are not so good. In recognizing the importance of agent-relativity, the concern is that we are either endorsing the validity of all of those preferences, or, at the very least, failing to explain why some preferences are acceptable and some are not. After all, the racist favors members of his own race, but we do not think that is morally acceptable. Why should his preference for himself and his family be treated any differently?

This is the standard way of looking at the question of agent-relativity; and although Kagan does not explicitly connect the dots in this way, others have raised this sort of
The problem with approaching agent-relativity in this way is that there is more to it than the fact that we want certain things and we don’t want others. Understood as a counterpoint to the agent-neutral point of view, the agent-relative point of view describes an entire mode for ranking or assigning value to states of affairs—a mode that I will argue is impossible to escape. In other words, while Kagan seems to accept that we naturally overvalue our own interests, the mistake is to think that this fact in itself is somehow a matter of choice or personal preference.

On the contrary, it seems to me that we are psychologically incapable of leaving behind agent-relativity altogether—in part, at least, because of how agency manifests itself. Perhaps this is a controversial claim, so let us approach it in steps. At the very least, it is clear that we struggle to find cases of individuals who live their lives without regard for the agent-relative point of view. It also seems uncontroversial that the agent-relative point of view is more basic developmentally than the agent-neutral standpoint, in the sense that it defines the default setting of our individual psychology until we are mature enough to recognize impartiality. To the extent that it is appropriate to describe young children as privileging possible states of affairs, they seem to do so from the agent-relative point of view—starting out as “selfish” and eventually coming to recognize that others’ interests also carry moral weight.

Once we are capable of deferring to agent-neutrality, we still seem unable to make the switch permanently from agent-relativity to agent-neutrality. Even the most diehard consequentialists fail to live up to the agent-neutral standard they defend, prioritizing

I owe this (modified) objection to Gopal Sreenivasan.
their own interests over those of strangers. There is indeed a significant literature of consequentialists attempting to fight off their agent-relative intuitions with agent-neutral logic, some of which I will engage with below. I hope it is therefore uncontroversial that agent-relativity is a prominent concern, even among those who want to avoid it.

Strictly speaking, it is unnecessary to wade into the psychological literature in order to substantiate these claims because I want to make a stronger claim that is not actually dependent on these observations. I start with these observations because they constitute a type of claim that even some consequentialists (like Kagan) are willing to concede. The general tenor of the consequentialist replies to these sorts of points is that they do not seem sufficient on their own to justify deference to agent-relativity. The mere fact that we do, in fact, fall short of a standard does not mean that we should not hold it up as a moral ideal.

That is true enough. Still, what I want to suggest is stronger: that agent-relativity is actually inescapable as a matter of our psychological constitution. That is because it is a matter of *necessity* that acting as an agent involves conferring agent-relative value on various sub-optimal states of affairs. In other words, the act of making choices, whether for agent-neutral or agent-relative reasons, will of necessity cause us to prioritize suboptimal outcomes over optimal ones, at least sometimes.\(^\text{73}\)

---

\(^\text{73}\) It may seem slightly odd to suggest that we cannot escape agent-relativity even as we are permitted to maximize agent-neutral good in many cases, and even as people like Zell Kravinsky give it their best. I don’t think the tension here is as large as may seem at first glance, though. For one, objectively optimal states of affairs can be agent-relatively valuable to us depending on how we are situated vis-à-vis those states. As a result, we can have agent-relative reasons for bringing about objectively optimal (or at least good) states of affairs. Additionally, there will be moments when we are able to suspend agent-relativity long enough to effect legitimate change in the direction of agent-neutral
It is not important for our purposes what biological processes cause this to happen; it is more important to establish *that* it happens. Imagine someone who one day stumbles across a paper by Peter Singer. He finds himself compelled by the powerful simplicity of appeals to promote optimal states of affairs: “From this day forward, I will maximize agent-neutral value!” He pledges to make all decisions after that moment with an eye towards promoting the good. My claim is that, no matter how committed he is in his quest, he will inevitably find himself torn between maximizing the good and bringing about suboptimal states of affairs that favor his own interests. Moreover, it is not just that he will have overwhelmingly “selfish” desires (though he certainly will). Just in virtue of having endorsed certain options over others at various forks in the moral road after his conversion, he will start to care more about certain goals than his new ideology permits, even when those goals do not directly affect his general wellbeing.

Suppose his first consequentialist order of business is to find a charity to which he can donate his savings and free time. He narrows down the list to a variety of possible causes, none of which he has any reason to think will promote more or less agent-neutral good than any other. He will probably be tempted to choose a cause on the list for agent-relative reasons; maybe his aunt died of breast cancer, so he feels more strongly about optimality (for its own sake); that is perfectly consistent with the view that it is unreasonable to hold us to such a standard all the time because we can only live up to it periodically, at best. Finally, even if the rare individual is able to bring his personal state-of-affair rankings into relatively good alignment with objective rankings, I cannot conceive that he will succeed fully, and in any event, that is no reason to think the standard advocated for here is too low. That phenomenon appears very rare; even hardcore consequentialists who are generous with their resources (like Peter Singer) do not generally reduce themselves to the point of marginal utility. In any case, I do not find it troubling to permit people to maximize in general (so long as they violate no constraints), even though only a rotating handful may be able to do it at any given time.
contributing to a related project. But even if he chooses at random, his continued participation in that cause will eventually make it more important to him than the others, even though he would recognize the irrationality of his preference when considered from an agent-neutral point of view. If he makes friends with other volunteers for the cause, the very fact that he is friends with them will make him care more about their wellbeing than the wellbeing of strangers (otherwise, it is unclear how they would be his friends). If he falls for one of the other volunteers, you can bet that he will start to care about him or her more than would be permitted agent-neutrally.

Moreover, as Tom Hurka points out, this is just one of many ways in which agent-relative state-of-affairs valuations manifest themselves. For example, we care disproportionately about pain in our lives as against pain the lives of others, and we also care disproportionately about various projects or objects from our pasts. Even if we can imagine someone who lacks a particular element of the standard agent-relative impulse, it is hard to conceive of someone who lacks them all.

Whatever the mechanism, the point is just that we form agent-relative attachments to various features of the world that play an important role in our lives, whether we want to or not. The fact that our imaginary agent has committed himself in the strongest terms to making decisions based on the agent-neutral weighting of possible outcomes is never going to be enough to prevent him from forming agent-relative attachments. Those attachments simply happen, most likely because our very endorsement of a particular option confers agent-relative value on it, especially when we endorse it repeatedly (as when someone works for a cause for some extended period). Or consider the broader example of sentimental value: certain items with no objectively discernable special
qualities take on inflated meanings for agents in virtue of nothing more than their connection to some memory in the past.

While nobody seems to have used this sort of claim specifically to justify the normative significance of agent-relativity before, I am not the first to describe this type of phenomenon. Nagel essentially describes this process when elaborating on his category of “reasons of autonomy.” He offers the example of someone wanting to climb a mountain, and notes that, from the inside, his wanting to conquer Kilimanjaro is a reason to do so, even if that reason does not necessarily extend to making others care that he succeeds. (Nagel, 145-146)

The general idea here goes a bit further than Nagel suggests, though: investing time, energy or care in bringing about a particular state of affairs will make that state more valuable to us than merely allowing the exact same state to come about without our effort. From the objective point of view, the states are equally valuable in virtue of being identical, but that does not capture the entire story. It omits the agent-relative side, which explains (and legitimizes) our preference for certain states of affairs over others on grounds other than the neutral configuration of goods we find in them.

This does not mean that we cannot conceive of or understand agent-neutral valuations of states of affairs. To the contrary, I have explicitly claimed that I want the agent-neutral point of view to play a crucial role in our understanding of the right.74 Instead, the point is that we must work to adopt the agent-neutral point of view, and we

74 How exactly we balance the two points of view is a further question, and one that is difficult to answer. It is plausible that people can be conditioned to expect greater or lesser degrees of agent-relative freedom, and that will complicate clear-cut attempts to locate a fulcrum for the relative/neutral seesaw.
cannot operate from such a standpoint indefinitely. Eventually we will slip back into the agent-relative point of view. In this respect, viewing the world from the agent-neutral point of view is rather like holding one’s breath: with conscious effort, we are capable of doing it for short stretches; but soon enough we start breathing again. Without conscious effort we breathe constantly, just as we constantly occupy the agent-relative point of view unless making a deliberate effort not to do so. The extremist position defended by Kagan is the moral equivalent of asking each of us to persist while indefinitely holding our breath.

As I understand agent-relativity, then, it issues from the fact of the natural independence of the personal point of view. A moral system that minimally reflects natural independence will quite rigidly necessitate features of a moral system that Kagan’s extremist rejects. The fact of natural independence demands that self-preferential space. Fully reflecting the fact of natural independence, in Kagan’s terms, would require giving each of us unlimited latitude to favor our own interests. We might say that independence appeals for indefinite space, for complete freedom from agent-neutral demands. If Kagan is correct about how we must handle the demands and appeals of various sets of facts, then there are compelling reasons to think that a proper theory of the right will balance agent-relativity and agent-neutrality in the way that the moderate intuitively seeks to do. Such a theory will have to give us some latitude to breathe, and I will argue later for the conclusion that options and constraints accomplish this purpose nicely.

For now, I want to establish that accepting natural independence merely as a causal explanation for our stubborn refusal to promote the good is insufficient. In part because of how we experience agency, we simply cannot consistently rank the desirability of states of affairs from an agent-neutral point of view, nor can we promote agent-neutral good on
anything like the scale necessary to satisfy the extremist. It is absurd to hold us accountable for either of those failures. Indeed, they are not failures at all; we are simply not constituted in the way necessary to be bound by extremism. If ought implies can, then we cannot be pure consequentialists.

In short, the fact that agents naturally overvalue their own interests runs deeper than the facile observation that we all prefer certain things to others. The racist and the civil rights activist may have widely divergent preferences, but each dwells largely in the world of agent-relativity, and neither has the power to change that. No moral theory can have normative purchase on us without recognizing and reflecting that fact.

There may remain a worry that the mere possibility of stepping outside of the agent-relative point of view is enough to justify something like consequentialism. In cases where it occurs to us to maximize, why should we not be held to an agent-neutral standard? If there is no good reason to oppose this suggestion, perhaps conditioning ourselves to “hold our breath” more often will then impose on us consequentialist obligations in at least a significant proportion of cases.

We should not think that such an approach, even if reasonable, would give us something that looks terribly like consequentialism. Given the natural primacy of agent-relativity, it is hard to imagine that even the most well trained individual would be able to view more than a small percentage of his choices from an agent-neutral perspective. The

75 Of course, the racist’s preferences are also quite likely to involve violating constraints around others.

76 We might even draw indirect consequentialist conclusions here, suggesting that an attempt to maximize the good would fail if it required imposing such high demands on
bigger problem, though, is the normative impropriety of forcing agents to adopt and act purely on an agent-neutral scheme for ranking states of affairs (even if only in a subset of cases). My understanding of the agent-relative point of view is that it reflects a basic element of our psychological constitution, and as a result of its centrality in our agency, ignoring it is simply unreasonable.77

Similarly, lest there be any confusion, I should emphasize that the justification for accepting the relevance of the agent-relative point of view is not a matter of prudence (though it probably is prudent). All of us have our own agent-relative state-of-affairs rankings, and that may seem to suggest that we each have reason merely to promote those states that we individually find appealing; but there is nothing in the agent-neutral rankings or in my agent-relative ones to give special consideration to the agent-relative points of view of other people, which some may find problematic.

The view I have in mind, however, does not rely on the concern for other people’s agent-relative points of view being captured by the relative stance I take or the neutral one we all may access. The idea is supposed to be that the fact of natural independence is of normative significance, rather in the way that agent-neutral maximization is of normative significance. Recognizing the natural independence of others’ points of view is no more a matter of our self-interest than is recognition of agent-neutral considerations.

people who could not help but feel disproportionate pain at being made to comply with agent-neutral prescriptions.

77 This, quite plainly, constitutes a further normative claim; but it seems to me fairly plausible. It also suggests that there is a significant difference between ignoring agent-relative weightings and acknowledging them but finding them overruled by agent-neutral considerations. We will revisit this below.
We thus have an objective reason to recognize the role of natural independence in human agency. But this is not such a strange demand, for natural independence is not an opaque phenomenon that we must take as an article of faith. We each personally experience it, and that is why accepting natural independence, at least as I understand it, is not supposed to lead to a harsh confrontation with egoism.

The tension between the agent-neutral and agent-relative points of view actually explains a lot of the morally unclear situations we encounter. Balancing the two sorts of considerations is a difficult matter because, while we can survey the exact same scenario through either lens, the weights they each assign to different components of various scenarios can be radically different. Agent-relative and agent-neutral valuations can thus prescribe different courses of action. Part of the difficulty of being a morally good agent is that it is tough to sort out each type of consideration on its own, let alone in cases where the two weigh against each other. The interplay between them will become more apparent as we turn to an investigation of the moral features such an approach will yield.

Before we do that, however, it might be useful to revisit Sidgwick and Nagel for a moment, for by now it should be clearer how the view I propose resembles those that they suggested. Sidgwick famously argued that two distinct, irreducible perspectives shape our interpretation of moral questions: the egoistic and impartial points of view—which are roughly analogous to the agent-neutral and agent-relative points of view. There is thus a striking structural similarity between the two views. Further, in virtue of balancing the two perspectives, this sort of approach constitutes one method of filling out the sort of view endorsed by Nagel, who indicates that both agent-relative and agent-neutral
considerations are essential for a good moral theory because the canyon between them is too deep to overcome.

4.3. What Normative Features Does Agent-Relativity Yield?

This balancing approach will only be of limited use if we cannot show how it yields an appealing theory of the right. So the next step will be to illustrate how grounding a theory in the agent-relative/agent-neutral distinction can account for agent-centered prerogatives (options) and agent-centered restrictions (constraints), essential elements of commonsense morality. Later I will sketch out the beginnings of an agent-relative explanation for other intuitively appealing moral features.

I have claimed above that minimally reflecting the fact of agent-relativity will give us both options and constraints, but how this works becomes clearest if we start by considering what it would take fully to reflect the fact of agent-relativity. If we were to honor a person’s agent-relative perspective to the exclusion of all other (agent-neutral) considerations, we would need to give him three agent-relative dispensations:

(D1) The freedom for him to pursue preferred states of affairs, even at unlimited agent-neutral expense

(D2) A barrier around him, preventing others from undermining the coming-to-be of preferred states of affairs, even at unlimited agent-neutral expense

(D3) A moral requirement that others assist him in bringing about preferred states, even at unlimited agent-neutral expense
These three dispensations, taken together, would offer any given agent maximal support in pursuit of his preferred states of affairs—as opposed, specifically, to states that are objectively the best. The reason that there are exactly three such dispensations is simply a matter of logic; these three, when enacted in tandem, exhaust all of the possible ways in which an agent could have moral primacy in pursuit of his own favored states of affairs.

(D1) would grant him what we have been referring to as an “option” (or agent-centered prerogative), allowing him to choose his various goals and projects, even when he would bring about (much) more good by doing something else. (D2) would impose a constraint (agent-centered restriction) on others, preventing them from interrupting the coming-to-be of circumstances he finds favorable—again, even if interfering would result in agent-neutrally optimal states of affairs. Finally, (D3) would ensure that one’s personal limitations do not necessarily block off states of affairs that are preferable to him, specifically by allowing him to rely on others for assistance no matter how much better the world would be (objectively) if he were to go unassisted.

Note that (D2) and (D3) draw on facts about an agent to impose duties on others. That is a distinctive element of this approach, which sets it apart from Nagel’s, as well as from others who seem to have advocated fitting a moral theory around facts about agents

---

78 Again, I want to emphasize that a person’s preferred states of affairs need not be defined solely by selfish or self-interested desires. Wishing for a state of affairs that is favorable to others would make it one of your preferred states of affairs too, even if your wish has no selfish or self-interested underpinning.

79 There is a pervasive worry that a key asymmetry exists between (D1) and (D2). The connection between overvaluing my projects and being permitted to pursue states of affairs where they obtain appears straightforward, whereas the connection between overvaluing my projects and preventing other people from interfering with those states of affairs seems less direct. I will address this concern below.
(such as Rawls and perhaps Scheffler). I will have more to say on this later. In any case, there is no conventional term for the “device” that correlates with this third exemption. For ease of reference, we might call it an unlimited “other-regarding option,” because the device gives its possessor the special status to call upon unlimited positive (or negative) assistance from everyone else.  

We have defined these dispensations in the strongest possible terms, such that they withstand any amount of agent-neutral pressure, and we have done so in order fully to reflect natural independence. Yet even the strongest version of these exemptions would not be enough to ensure that someone could do anything he or she wished, for reflecting natural independence properly involves giving these dispensations to everyone whose personal point of view is naturally independent from the agent-neutral. In other words, it has been implicit in this discussion that every person has each of these dispensations. That fact alone is enough to introduce limitations on anyone with even the most robust possible agent-relative exemptions. For example, I may have the option (D1) of pursuing states that are far from being objectively optimal, but I will be limited in my pursuit of those states by (at the very least) the protective barriers (D2) around others. These dispensations would only give someone truly unfettered choice about what to do if no one else had his or her own set of dispensations. That way there is no possible set of considerations that could ever outweigh the legitimacy of his preference for a particular state of affairs. (Indeed, he would even have as much assistance as possible in achieving his aims.)

80 People who are liable to be called upon by an other-regarding option might be said to be bound by a “contingent demand,” a decisive request to do or abstain from doing something in accordance with the preferences of someone holding an other-regarding option.
Unfortunately, we can't give all of these dispensations to everybody. Specifically, while (D1) and (D2) are universalizable, (D3) is not. To give every agent (D1) would mean allowing everybody the latitude, or option, to pursue his or her favored states of affairs at agent-neutral expense. If Alpha, Bravo, Charlie and Delta all had that freedom, then there would be some competition among them for the states of affairs that each prefers; but all four would be free, respectively, to fight for the outcomes they desire. Similarly, it is possible to give all four of them (D2). Doing so would mean establishing a protective barrier around each of them (in the form of a constraint on the others) that makes at least some of their projects “unsacrificeable,” even when agent-neutral good stands to be gained by sacrificing them. These barriers would inevitably have to give way in a limited subset of situations—namely, where an actor has no choice but to violate at least one constraint because every possible avenue of action entails such a violation. This range of conflicts aside, however, Alpha, Bravo, Charlie and Delta could each benefit from possession of (D2).

But if we try to universalize (D3), we find that doing so is logically impossible. There are always different states of affairs that are agent-relatively preferable to every single agent, so we simply can't designate everyone as meriting assistance above and beyond what is agent-neutrally required. If Alpha’s calls for assistance count for extra because he overvalues his own projects, and the same is true of Bravo, and of Charlie, and of Delta, then when there are conflicting calls for help, we cannot simply appeal to the fact that each of them overvalues his projects in deciding where to direct our assistance. Other-regarding options only serve an agent-relative function if they are not universally shared.
The fact that (D3) can't be universalized is extremely important, for it means that there is no way to justify giving everyone the level of assistance their agent-relative state-of-affairs valuations demand—at least not on the grounds that we are honoring their agent-relativity. More broadly, the implications of this asymmetry between (D3) on one hand and (D1) and (D2) on the other are particularly significant. Consider a variation on a popular example: Alpha, Bravo, and Charlie have all been stabbed and are in dire need of medical care. Their lives can be saved, but only if we can exchange their damaged organs for undamaged replacements. Delta happens to be a genetic match, and were we to sacrifice him and divide up his organs among his three injured comrades, we could bring about a state of affairs in which only one life is lost rather than three.81

A common but unrefined intuitive response to the prospect of carving up Delta is to appeal to the fact that he overvalues his life and somehow has a right not to be sacrificed. But the consequentialist will remind us that Alpha, Bravo and Charlie also overvalue their own lives; presumably, then, the same right applies to each of them. So why are we not summing up their three overvalued lives against Delta's one? After all, the math is clear. The fact that (D3) cannot be universalized gives us a rejoinder to the consequentialist. We can say that all four individuals have options to pursue certain states of affairs, and that all four have protections around them. But we cannot say that each has a special right to assistance—at least not based on the fact of their respective agent-relative preferences.

81 Typically the dying victims in this example are ill or injured by natural causes, not because of the intentional acts of others. I have modified the case so that we are comparing apples to apples: three murders on one side to one on the other. The reason for making the change is to preclude questions about a confounding factor—namely whether one murder (of Delta) might be objectively worse than the three natural deaths of Alpha, Beta and Charlie.
In the transplant example, we arrive on the scene at a point that evokes this asymmetry. Alpha, Bravo and Charlie have already had the barrier around them violated. They have been wronged, we might say, and they have the option of pursuing the state of affairs in which they receive life-saving care, so long as they do not violate any of the relevant barriers around others. Delta, on the other hand, is fortunate not to have been stabbed. If what I have suggested above is correct, then the choice we are faced with is rather simple. Do we violate the barrier around Delta (albeit for a good cause), or do we leave him be? The fact that Delta overvalues his life in the same way that Alpha, Bravo and Charlie do does not set up an agent-relative stalemate that must be decided by the numbers. We can justify not violating the barrier around Delta on agent-relative grounds, but those grounds are simply unavailable to us in justifying assistance to the other three.\textsuperscript{82}

The only way to justify assisting Alpha, Bravo and Charlie (barring extenuating circumstances, like those in which we have special obligations to do so) is by appeal to the

\textsuperscript{82} The practical effect of the asymmetry between options/constraints and other-regarding options is that we are permitted to allow certain bad things to happen even if we could do something less bad to prevent them. We thus get a result that aligns rather neatly with the distinction between doing and allowing. I want to emphasize, however, that the rationale behind denying agent-relative assistance to Alpha, Bravo and Charlie has nothing to do with the fact (if it is a fact) that it is worse to kill Delta than to allow his three pals to die. I do think it is worse to kill Delta, at least in the sense that there is an agent-relative reason not to (whereas the same is not true of permitting the deaths of the other three). But that is different from relying on the bare intuition that doing bad is worse than allowing it to happen. The overlap here between the prescriptions of the option/constraint system and the doing/allowing distinction appears to be incidental. I should say that I share the intuition that doing evil is generally worse than allowing it, and I think that the reason why can be explained by appeal to the agent-relative point of view. (Agent-neutrally, killing Delta is clearly better than allowing three others to die.) I will address the question of how below. But for now, I simply want to make clear that I do not take myself to be relying on the doing/allowing distinction in order to stand up my verdict in transplant-type cases.
agent-neutral value of a state of affairs in which only one person is killed rather than three. All else equal, the objectively preferable state is quite clearly the one in which there is only one murder rather than three, which suggests that agent-neutral considerations here are weighing against agent-relative ones. The question arises whether we should accept agent-neutral considerations, and if so, how we determine when they outweigh the agent-relative.

Before we inject agent-neutral considerations into the mix, we need to be clear about the relevant agent-relative features of the case at hand. Because we are honoring the respective agent-relativity of our four agents, we have universalized as many of the agent-relative exemptions as possible, giving all of them both the option to pursue preferred states of affairs and barriers around them that constrain the others from interfering. Since we cannot universalize (D3), none of our agents have other-regarding options designating them as disproportionately worthy of assistance.

There are several concerns that may have arisen throughout the preceding discussion. Before moving on to agent-neutrality, I think it would be best to take a moment and address the most significant of these. One might wonder why honoring agent-relativity means providing everyone with moderate versions of (D1) and (D2) (moderated, that is, by the limitations they place on one another) rather than just the strongest version of (D1), or just the strongest version of (D2). That is, I seem to have assumed that giving everyone tempered options and tempered protections is more important than giving them either unlimited options or unbreakable protections. Some

---

83 There is an entire class of concerns about what the moral landscape would look like if we only accepted options or if we only accepted constraints. Shelly Kagan has raised a
might even think it inaccurate to say that we can universalize both (D1) and (D2) and the same time, and that is effectively the same thing as being unable to universalize (D3).

The latter of these concerns is simply misguided. The problem with (D3) is intrinsic; universalizing it is self-defeating such that no broad rationale like the one I use here could possibly justify the dispensation in the first place. Even if we discarded (D1) and (D2) entirely, we would not be able to give everyone (D3) because the common possession of the dispensation would literally cancel out, making the world exactly the same as it was before we enacted (D3). That is why we have no agent-relative reason to help Alpha, Bravo or Charlie when each is crossing his fingers for some of Delta’s organs. We only have an agent-neutral reason (the objective math) because (D3) is useless.

On the other hand, neither (D1) nor (D2) is self-defeating when universalized, so it is not incoherent to talk about everyone having either (or both) of them. At the very least, (D1) and (D2) are individually universalizable while (D3) is not. But it is perfectly true that universalizing both (D1) and (D2) at the same time yields weaker versions of each than using simply one or the other. If everyone has protections, then those protections will close off options possessed by others; and if everyone has options, then those options will expand the number of considerations that will override others’ protections. The question is why balancing both (D1) and (D2) is better than giving everyone a more robust version of just one of these dispensations. The answer is that, on the rationale I have given, it would be inconsistent to exclude one or the other of these. The natural independence of number these specifically with respect to the former, but those are not the sorts of objections I have in mind at this particular point. I will raise some of Kagan’s criticisms below when I take on Samuel Scheffler’s position.
the personal point of view affects our state-of-affairs valuations whether we act in pursuit of an outcome or whether we want others to abstain from interfering with an outcome.\textsuperscript{84} If we are to reflect that fact in a manner that is not arbitrary, we need to include both (D1) and (D2).

If we shape our normative ethical theory in the most natural way, it may perhaps become a bit clearer why we should accept both (D1) and (D2). In building our theory, it is reasonable to account for agent-relativity first, for it captures the psychological limitations we cannot be expected to overcome. Since Alpha has a natural agent-relative preference for (D1), (D2) and (D3), we give him all three. Then we universalize those dispensations, giving them all to everyone. Alpha’s (D1) clashes with Bravo, Charlie and Delta’s (D2), and Alpha’s (D2) clashes with all of their respective (D1)’s. Everybody’s (D3) clashes with everyone else’s (D3), so it annihilates itself. We are left with both a mitigated (D1) and a mitigated (D2) for everyone. Now we are free to impose agent-neutral demands on everyone to the extent we think it appropriate.

I suspect that people would generally prefer to have both mitigated options and mitigated protections rather than a stronger version of merely one or the other. The rationale I am using here is not prudential, however, so our preference does not matter except to the extent that it reflects the agent-relative point of view. I think that natural independence is exactly what explains our preference, but avoiding arbitrariness is a sufficient reason to avoid relying on such speculation.

There remains at least one large concern, specifically with regard to an ostensible

\textsuperscript{84} I will have significantly more to say about this point later as well, as it is the basis for my primary criticism of Samuel Scheffler’s position.
asymmetry between (D1) and (D2). Recall that (D1) gives everyone the freedom to pursue
his preferred states of affairs, while (D2) prevents others from interfering with those
preferred states of affairs. It seems like there is a direct route between Alpha’s
overvaluation of certain suboptimal states of affairs and his (D1) permission to pursue
them. On the other hand, there does not appear to be quite so direct a connection between
Alpha’s overvaluation of certain states and a (D2) prohibition on what others may do to
interrupt them. Perhaps natural independence could ground (D1), but it seems as if it
requires a further claim to establish (D2).

I take this to be among the most prominent concerns about the asymmetry
between options and constraints (at least to the extent that one employs something like a
natural independence justification for them). The first problem with this objection is that
it is not as serious as it seems because the asymmetry it relies on is exaggerated by the
labels we use for each respective normative device. The word “option” (or its more formal
equivalent, “agent-centered prerogative”) explicitly denotes the concordant device’s
benefits, and carries a positive connotation. The word “constraint” (or “agent-centered
restriction”) emphasizes the cost of the device and carries a negative connotation. In part
to address this asymmetry, and in part for clarity, I have introduced the term “agent-
centered protection” to capture the positive side (indeed, the normative purpose of)
constraints. “Protection” (rather than “constraints”) thus serves as the proper counterpart
for “option.” We should also introduce a term that emphasizes the negative side of
options—a counterpart for “constraint”—such as “imposition.”

The objection is overblown because the asymmetry behind it depends, at least in
part, on a tendentious description of the effect of (D1) and (D2). The worry, as framed
above, specifically juxtaposes options with constraints—rather than options with protections, or impositions with constraints—and thus evokes a misleading asymmetry in favor of options and against constraints. We could pull the same trick by comparing protections with impositions. That is, it would be about as misleading to say that Alpha’s overvaluation of his projects might permit him a buffer zone of protection around those projects, but then ask why the natural independence of his personal point of view should permit him to destroy other people’s preferred states of affairs.

What this semantic issue highlights is whether we give Alpha (D1) or (D2), he'll have a significant effect on those around him—whether as someone who has license to tread on their projects or as someone whose preferences impose limitations on their movement. I think this point helps to undermine the power of the objection to some degree because the structural asymmetry does not carry over into such a significant practical asymmetry, but analytically speaking, I have not directly responded to the concern. The worry is that, irrespective of the effects of either dispensation, (D1) links facts about Alpha directly with a freedom of Alpha’s, and only as a secondary matter carries implications for patients around him, while (D2) seems to make a direct connection between facts about Alpha and a duty limiting others.

This difference is undeniable, but I actually see it as a merit of the approach rather than a defect. Not only is it a distinctive element of the justification I have in mind (as against Nagel, for example), but it is also the core feature that makes the constraints we get from this system intuitively appealing. It brings the view into alignment with Frances Kamm’s intuitions in providing a patient-centered justification for constraints (though this
view differs significantly from hers for other reasons that we will discuss below). 

Certain facts about creatures carry normative significance for others around them, determining what they are entitled to and thus how we ought to treat them. Most people accept a sliding scale of moral status for animals of increasing complexity, and a number also accept something similar with respect to humans in various stages of development (as embryos, fetuses, infants, whatever). In every case, we ask what qualities the animal (or future animal) possesses, and we determine how to treat it as a result. Why should natural independence be any different, especially given that it isolates a specific fact that is plausibly of normative significance? Were these constraints not patient-centered, we would object to them on the grounds that they fail to capture the sense that we wrong someone when we treat them a certain way, and that we step over a boundary that emanates from them as a moral being.

Moreover, the purpose of admitting the relevance of natural independence isn’t to provide a rationale for constraints built on the self-interest of those who are constrained. I will discuss below why I do think it is to their advantage, but that is rather a secondary consideration. My understanding of natural independence is that it is simply a basic fact about our psychological constitution, and it is of normative significance precisely because it is so directly relevant to our state-of-affairs valuations. The reason for recognizing the natural independence of others’ points of view is both that it is of normative significance and that we are capable of doing so, much the same reasoning we use to justify the

85 Despite favoring a patient-centered approach to constraints, I think the term “agent-centered” is sufficiently accurate because the patients who generate constraints only do so because they are agents of sorts.
relevance of agent-neutral state-of-affairs valuations. Thus, the fact that Alpha overvalues his own projects is not intended (in the first instance) to give Bravo or Charlie or Delta a prudential reason for respecting Alpha’s preferences—at least, not on this view.

If we were to reflect the fact of agent-relativity fully, we would end the story with options and constraints. There would be no room to include agent-neutral considerations. Instead, Alpha and his friends would all have options that are limited by nothing but the agent-relative protections around each other. Whether Alpha prefers philanthropy or fishing, he is free to pursue whatever he wants. He will never be obligated to abstain from any course of action unless it involves violating someone else’s agent-relative protection. Similarly, each of Alpha and his pals would have inviolable protective constraints around them, meaning that none of their protected projects could be used to rectify a wrong, no matter how much better the world would be (objectively) were we to violate the constraint and make the trade.\(^{86}\)

More importantly: contra Kagan, this is the view we get if we fully reflect the fact of agent-relativity. We have given everyone as many agent-relative dispensations as possible and we have made those dispensations as robust as we can. People are thus provided with as much freedom and protection as can be shared uniformly among a plurality of agents. But this view is implausibly strong. For one thing, there are a number of cases where

---

\(^{86}\) It is conceivable that someone might find himself in a dilemma where either fork requires violating a constraint. In that case, violating a constraint is not a voluntary matter; it is metaphysically impossible to avoid violating a constraint, and since it is literally out of our reach to avoid doing so, we are not necessarily in the wrong. (It would depend on how culpable we are for arriving at the dilemma in the first place.) In cases like that, and assuming the constraints we are to choose between are of equal strength, it seems as if agent-neutral considerations would constitute the salient tiebreakers.
agent-neutral considerations seem quite clearly to outweigh our freedoms to pursue our own projects. Despite being in a rush myself, I might have a moral obligation to stop and help someone who’s been in an accident, even if continuing on my way would violate no protective constraint around the accident victim. Likewise, it may be wrong of people to lie to others (for example), but it is compelling to think that there are cases where deceiving someone could be justified on the grounds that an enormous amount of good stood to be gained by doing so.

The natural independence of the personal point of view notwithstanding, we are capable of taking on “the view from nowhere,” even if only for short periods. We thus have the capacity for taking account of agent-neutral considerations to a limited degree. The entire rationale I have offered for admitting the relevance of agent-relativity is that it is an irrevocable part of how we view the estimated value of various states of affairs. As a result, the extent to which we are able to get away from agent-relativity dictates, to a significant degree, the extent to which we are bound to act from agent-neutral considerations. I am thus happy to agree with Kagan that we need only reflect agent-relativity to some minimal degree rather than fully—so long as we adjust his understanding of what it is to reflect the fact of agent-relativity both minimally and fully.\footnote{87}

Showing that options and constraints are \textit{consistent} with honoring the agent-relative point of view is not the same showing that they are \textit{favorable} from an agent-relative point of view. The general consensus is that this is less an issue for options than constraints, for it is common to regard constraints with a greater degree of skepticism.

\footnote{87 The next question it is natural to ask concerns how to balance agent-neutral and agent-relative considerations. I will have more to say about this below.}
More specifically, there are several asymmetries between options and constraints that make it less obvious how the latter serve to reflect or honor the agent-relative point of view. Even if what I have suggested above is compelling, therefore, doubts may remain about why we would opt for an ethical system that includes constraints even if we are sympathetic to the natural independence of the personal point of view.

The reason for doubt is that universalizing (D1) or (D2)—or both, of course—has both positive and negative implications for everyone. We might wonder whether the balance of those implications counts in favor or against universalizing either exemption. If we give only Alpha (D1), he would have no cause of objection. He would have an option to pursue whatever states he prefers. As soon as Beta, Charlie and Delta get options too, however, they will have the freedom to pursue states of affairs that Alpha may find unappealing or even threatening. We cannot, of course, justify giving options only to Alpha, at least not based on the rationale I have been using. So we are left to choose between giving options to everyone or giving them to no one. Alpha may wonder whether the liberty he gains by having options outweighs the threat posed to him by Beta, Charlie and Delta having options too.

Similar calculations are possible with constraints. We take a step towards honoring the natural independence of Alpha’s point of view if we give him an agent-centered protection, thereby constraining Beta, Charlie and Delta from disrupting or destroying Alpha’s projects. But as soon as we extend those same protections to Beta, Charlie, and Delta, Alpha will have to be cognizant of three sources of constraints that will prevent him from taking certain courses of action or bringing about certain states of affairs. Direct restrictions on Alpha are not necessarily the only costs he must keep in mind; there may
be cases where the presence of protections around the others harms Alpha by imposing indirect restrictions (that is, restrictions on others). For example, there may arise situations in which the protection around (say) Beta prevent Charlie or Delta from taking an action that might benefit Alpha, as in some variation of the transplant case.

It is conventional to regard options as worth the trade, at least for those who think that the natural independence of the personal point of view merits normative accommodation.\(^{88}\) For one, it is quite clear how options help to honor the agent-relative point of view; if we cannot help but overvalue suboptimal states of affairs, allowing us to pursue at least some of those states is a rather obvious way of tipping our hat to our own psychological limitations. Further, without options we would have no license to pursue anything but objectively neutral states of affairs, which would almost certainly be both alienating and miserable in the extreme. Arguably, then, giving us the chance to chase our desired states of affairs is so valuable that it offsets any additional risk to us of giving options to others. Finally, in the world where there are agent-centered protections, others’ options are unlikely to be much of a threat at all; given the intuitive draw of such protections, some might mistakenly think we can count on them in assessing the merits of options.

Worries about constraints seem to run deeper. There have long been concerns about how it is possible that stopping people from bringing about optimal states of affairs could be morally justifiable, even (or perhaps especially) by appeal to agent-relative considerations. Samuel Scheffler offers a particularly clear discussion of this sort of worry

---

\(^{88}\) Cf. Samuel Scheffler, *The Rejection of Consequentialism*. 
in *The Rejection of Consequentialism*. Below I will engage directly with Scheffler’s articulation of this worry. In the preceding pages, I have attempted to approach the agent-relative justification for constraints in a step-wise fashion that makes explicit that constraints are indeed both consistent with and necessary for honoring the agent-relative point of view. Part of the impediment to seeing this in the past has been a matter of semantics, a problem I have tried to deal with above.

If we can agree, based on the preceding, that constraints are at least consistent with honoring agent-relativity, the question becomes whether there is reason to think that they are desirable. In one sense it seems obvious how valuable agent-centered protections would be; constraining others from disrupting or destroying our projects, either for agent-neutral gain or for their own personal gain, gives us the ability to depend on those projects in a way that we are naturally prone to do. The higher the threshold for overriding our projects, the safer it is to invest in them.

There are, however, reasons why we might hesitate to embrace constraints. For one, some may think the limitations they impose on us are too costly. I believe that to be the weakest worry. Having more freedom to do what we wanted because our options were not checked by others’ protections could not plausibly be thought superior to having fewer options paired with agent-centered protections. The particulars of this sort of comparison depend on a more precise description of what options and protections we might have, but it will be close enough if we weigh unlimited options with no protections (something like the state of nature) against something like the Harm Principle. The intuitive support for protections is so pervasive that the latter choice will almost certainly prevail; and the reason for that is probably just that the Harm Principle better reflects the
natural independence of the personal point of view (by giving us both freedom and protection—(D1) and (D2)—rather than just (D1), as the state of nature does).

There may appear to be a bigger worry. It seems as if a world with constraints might result in more of our projects being interrupted than a world without constraints. If that is true, we may face a lower probability of having our projects disrupted in the consequentialist world than we would in the deontological world. The way I have described the agent-relative point of view suggests that there is a strong—perhaps even decisive—agent-relative reason for preferring the state of affairs in which our projects are disrupted the least. Therefore, it appears as if we might have agent-relative reasons for preferring the consequentialist world. (And it may thus appear that while constraints are consistent with the natural independence of the personal point of view, they are not favorable from that point of view.)

I am not convinced that the assumption behind this objection is true. The consequentialist will sometimes interrupt projects to preclude future and more numerous interruptions, and in those cases the deontologist falls behind in the count. But there will also be cases where the consequentialist would advocate numerous (perhaps more minor) interruptions in order to preempt fewer more serious ones. At the very least it is an open question as to which is more advantageous from a strictly statistical standpoint. To the extent that the answer eludes us (which it inevitably will do without more concrete information), this does not amount to a very useful objection.

More significantly, even if the numbers turned out favorable to the consequentialist, the rationale we are using here would not be sympathetic. The loss of a few of our projects is tolerable, if unappealing; but the consequentialist commands what is
literally impossible. If natural independence exists as I’ve described it, agent-neutrality would not win out regardless of the numbers. Our overvaluation of states in which we have invested entails that we perceive our pursuit of those states as morally legitimate, at least within certain bounds. The consequentialist system would never be able to capture that fact. Whenever we would escape sacrifice, it would be a matter of luck rather than a matter of the world recognizing our agent-relative valuation of a certain state of affairs. Similarly, any case in which our projects were violated would be morally appropriate or even required, and the cost to us would be calculated from a strictly agent-neutral standpoint.

Yet the whole point of taking agent-relativity into account—at least on the view that I am defending—is to bring moral demands on agents into alignment with what can reasonably be expected of them given our psychological constitution. An essential element of that constitution is our literal inability to view our projects and goals strictly as beads on the universe’s moral abacus. In any world where we lack agent-centered protections that element will be missing.

In short, we can’t help but want the protection given to us by agent-centered protections, whether it’s ultimately “best” or not. The natural independence of the personal point of view limits our moral capacities in such a significant way that we have to work within those bounds. Nevertheless, I recognize that others may not share that view, and think that constraints can be defended on their merits too, which is the reason for the detailed discussion in the preceding couple of pages.

89 Again, note that I mean this broadly, in a way that might justify (D2) and (D3)—not in the narrow sense we see in Nagel, Rawls and Scheffler.
There are a couple of other objections to this approach that are worth addressing here. One targets the use of the natural independence of the personal point of view as a justification for any agent-relative device, including options. Suppose we choose to honor the fact that agents naturally overvalue their own projects, giving them far more weight than they would by any agent-neutral standard. In order to make practical use of an approach built around that fact, we would need some sense of how much agent-relative overvaluing we are willing to tolerate at agent-neutral expense. One way of striking this balance would be to calculate the strength of our agent-relative perspective (as against the agent-neutral) and then pick the midpoint between the two values in order to find the moral threshold between them.

Let us imagine that agents overvalue their lives by 50 fold—viewing their own lives as worth 50 times more from the agent-relative perspective than the agent-neutral perspective would allow. For simplicity, we can set our agent-neutral value to one. If we want to give equal weight to the two perspectives, we might find the midpoint between these two weightings—one and 50—to come up with 25.5. We might conclude from this that everyone is required to overvalue his or her projects by a factor of 25.5.

I don’t have especially strong feelings about this approach to balancing the two points of view, but there is an enormous difference between saying that we are permitted to overvalue certain states of affairs by such-and-such a factor, and saying that we are required to do so. If we have permission to overvalue, then we have a range of choices; we can be selfish, or at least suboptimal, up to a specific point. If we are required to use the average, then our moral path is set for us as rigidly as it would be under strict, direct consequentialism—in both cases, we would only have a choice about what to do when
choosing among possibilities that would bring about (or would be expected to bring about) equally optimal outcomes. The objection is that the approach I am defending requires picking a rigid point, such as the midpoint between the agent-relative and the agent-neutral valuations, and setting agents to chase that point at any given fork in the road.

Perhaps this objection applies to some understanding of the agent-relative point of view, but I don’t see how it applies to my own. Suppose we accept the description I have given above of the personal point of view and we decide to exclude agent-neutral considerations altogether. If it still remains true that agents overvalue their lives by 50 fold, then this objection would suggest that we ought to pursue states of affairs that line up very specifically with that agent-relative valuations (50 times the agent-neutral value). But that is not at all what would follow; instead, agents would simply be allowed to pursue any states that fall within the given range (one to 50), so long as they work within any other limitations imposed by honoring agent-relativity (such as constraints). This is because the reason for introducing agent-relativity is fundamentally different from the reason for maximizing, and only the latter (so far as we are concerned here) imposes a rigid requirement. Moreover, the only reason for capping this range at 50 would be some sort of empirical fact about how far people actually happen to overvalue various states. If it turned out that they valued states even more, then the range would expand.

By the same token, even if we endorse making use of the midpoint between the agent-neutral valuation of states of affairs and the top range of the agent-relative valuation, all that would follow on the view I have developed is that we are permitted to work within the range set by the midpoint. So instead of working with valuations between
one and 50 times the agent-neutral, we would be permitted to work only with the range of one to 25.5.

A question remains, however, whether I am just assuming options instead of grounding them. The objection starts from the observation that there appears to be an asymmetry between the agent-neutral point of view and the agent-relative. Specifically, optimal agent-neutral states are binding on me (in some sense) and optimal agent-relative ones are not. In other words, objectively superior states of affairs will affect the way I reason about moral situations, assuming I do so properly; but I am not treating the agent-relative point of view in the same way. By admitting options, it may seem as if I’m saying that the agent-relative point of view does not offer a similarly rigid and binding ranking of states of affairs. Otherwise we’d have the situation described above, where the prescribed course of action would simply require that we follow the mean of the agent-neutral and the agent-relative considerations. If the agent-relative point of view truly were analogous to the agent-neutral, then there would be parity here. If not, perhaps I’m just assuming options by stipulating that the agent-relative point of view allows them.

This is an interesting objection, but I think it misses its target. The reason for admitting the relevance of agent-neutrality is just that we are capable of recognizing the moral relevance of objectively optimal or superior states of affairs. In adopting the agent-neutral point of view, we simply accede to the consequentialist’s request and admit that more good is objectively better than less. But the reason for accepting the agent-relative point of view is quite different: agents simply cannot help but create skewed state-of-affairs rankings. Those rankings align with their own personal preferences, so if an agent were to work toward an agent-relatively optimal state of affairs, he would probably satisfy
his strongest preferences at the time. Those preferences do not necessarily have any objective value, and the realization of the concordant state of affairs need not be valuable except to the extent that it is objectively good to have a particular person be especially happy with a given state of affairs.

The difference identified in the objection, in other words, has nothing to do with how the agent-neutral and agent-relative points of view operate, but rather concerns how we treat them. Both provide ordinal rankings of states of affairs. Moreover, our access to those rankings may be limited from either perspective, and each admits of ties in possible outcomes. The two are structurally identical. Yet I think we are justified in treating them differently on the ground that the reason each is morally relevant is enormously different. The purpose of recognizing the agent-neutral point of view is to force agents to consider their behavior from a fairer perspective, while the purpose of recognizing the agent-relative point of view is to control for the psychological fact that we cannot maintain—and cannot reasonably be expected to maintain—fidelity to the agent-neutral rankings.

Though it would establish symmetry between how we treat the two points of view, the value in forcing agents to pursue their agent-relatively optimal states of affairs is unclear. There is no reason to assume objective merit in the states that agents would bring about as a result, and imposing such a requirement cannot be justified by the purpose for which we admitted the agent-relative point of view in the first place. As a result of the justificatory asymmetry, I see no reason to worry about treating the two points of view differently. Additionally, to the extent that people can be said to choose what they regard as the most appealing option at any given moment (all things considered), it might be vacuously true that agents always choose to pursue their agent-relatively optimal states of
affairs. If that turns out to be the case, then the injunction to comply with an agent-relative ranking of states of affairs would be redundant.

Also, keep in mind that this approach does not have to yield an overly relaxed moral theory. There is a principled difference, I think, between an alienating theory of the right and a demanding one. Crude consequentialism is undeniably demanding, but also alienating—its dictates simply contradict essential features of our agency. A better standard for measuring the demands of a theory would be calibrated against what is actually possible for us. The approach I have in mind does just fine on this metric, for even by the standards of commonsense morality, many people fail to act morally, and we are still free to increase moral demands if we wish.

Part of the shape of our theory will hinge on our views about the good, which is a subject I will not be addressing here. But part depends on how much agent-relative indulgence we endorse, as well as on how stringently we define our duties. For example, suppose we accept a duty of beneficence. We can take a hard stance on that duty while at the same time skewing its demands in accordance with the agent-relative point of view (thereby making it much stricter with regard to our family and friends than with regard to strangers). As a result, how demanding we make the theory will depend quite a bit on details that we have not yet stipulated.90

90 We will also want to bear in mind Nagel’s warning that, in using this sort of approach, “we must guard against self-deception and the escalation of personal claims simply to resist burdensome moral demands.” (Nagel, 170)
4.4. Frances Kamm’s “Patient-Centered” Justification for Constraints

At first glance, it may appear that my proposal is nearly indistinguishable from Frances Kamm’s “patient-centered” approach to grounding constraints. At the very least, there are several similarities. The most important similarity is apparent from the name that Kamm gives to her approach. Specifically, Kamm does not think constraints emanate from the constrained party. Instead, she sees every given constraint as reflecting some quality of the person who is protected by it (that is, the “patient”). (Kamm, 272) To use the terminology I have introduced above, the constraint comes from the person who holds the corresponding agent-relative protection. That conclusion seems to me exactly right, even if we use different language to get there.

Our views start to diverge when it comes to explaining the purpose of constraints. For Kamm, constraints are not merely a protective device (though they do benefit us by offering protection); constraints also suggest the high moral status of whatever it is they protect. To be protected by a constraint (as she thinks people generally are) is to be “unsacrificeable” to some degree. According to Kamm, being immune to sacrifice speaks to our moral worth, and highlights our importance—especially relative to those things that are not protected by constraints.

As Shelly Kagan points out, though, this measure of status cuts both ways if we make the natural assumption that all people (or even most people) are protected by the constraints on those around them. (Kagan 1991, 920) Kamm, of course, wants to extend constraint protection to people in general, so this is not a trivial problem. Being unsacrificeable may seem to make me important in some sense, but if everyone has such protection, then sacrificing others for my benefit is, for the most part, impermissible. By
contrast, without constraints, damage done to others for my benefit can be perfectly acceptable so long as the overall balance of harms is appropriate. Giving people constraints can seem to confer on them a higher status from one angle but a lower status from another. (Ibid) (A similar tension arises between the moral status implications for being someone who others are obligated to assist and not being obligated to assist others.)

Given the inherent tension in these opposing propositions about moral status, Kagan points out that we have to pick which status matters more in order to settle the question of whether constraints are preferable. If we cannot find a principled and compelling way to make the choice, we risk begging the question in assuming that the status that comes with constraints is more important than the status we get without them. The view proposed here does not seem vulnerable to Kagan’s critique. His dilemma does not apply to us if we choose a normative arrangement on the basis of something other than status—in this case to reflect a fact about our psychological constitution.

This leads to an even bigger difference between my view and Kamm’s. The justification I’m attempting to use for constraints has a concrete basis. If it’s successful, it grounds our endorsement of anti-consequentialist elements in some empirical (if debatable) fact. Kamm’s approach avoids this sort of step altogether; she appeals to an intuitive notion of status which, quite possibly, in itself reflects the natural independence of the personal point of view. There is thus reason to think that the new approach has the sturdier foundation of the two.

**Conclusion**

In this chapter, I have sketched out the basic elements of a novel justification for agent-centered restrictions. While my approach is inspired by the work Henry Sidgwick,
Thomas Nagel, Samuel Scheffler, and Frances Kamm, it breaks with the positions of each in one significant way or another. Perhaps the most notable distinguishing characteristic of the new approach is that it attempts to justify constraints around a given agent based on psychological facts about other agents. While I have made efforts in this chapter to defend the view, there remain for us two primary tasks. First, we need to consider the objections typically raised against constraints more generally—objections that tend to question the coherence of constraints. If constraints make no sense regardless of where they come from, then no justification for them is worthwhile in the first place. Second, in order to bring out the merits of the new approach, it would be helpful to juxtapose my proposal with the deontologies considered in previous chapters.
The Viability and Merits of Agent-Centered Deontology

Introduction

If nothing else, the last four chapters should have established that agent-centered restrictions are difficult to justify. I hope the proposal I have offered in the preceding chapter amounts to a compelling justification, or at least to a step in the right direction. But quite apart from offering a basis on which we might justify them, there are significant questions about the coherence of the very notion of an agent-centered restriction. There are also questions about how well a set of restrictions derived from a particular source can serve the function we expect them to serve. In this chapter, I will address both of these issues. I will start by considering a range of objections to agent-centered restrictions, including the most prominent ones offered by Samuel Scheffler and Shelly Kagan. Finally, I will conclude by highlighting the advantages of the approach advanced here before revisiting the deontologies from previous chapters and offering closing remarks.

5.1. The Paradox of Deontology

A common worry about constraints concerns their seemingly irrational implications, typically referred to as the “Paradox of Deontology.” It is undeniable that
constraints limit agents, and an ostensibly troubling side-effect arises as a result: “Suppose that if agent $A_1$ fails to violate a restriction $R$ by harming some undeserving person $D_1$, then five other agents, $A_2 \ldots A_6$, will each violate restriction $R$ by identically harming five other persons, $D_2 \ldots P_6$, who are just as undeserving as $D_1$, and whom it would be just as undesirable from an impersonal standpoint to have harmed.” (Scheffler, 84) The problem in this case is that constraints bind $A_1$ so that he may not violate $R$, despite the further (and equally weighty) violations of $R$ that occur as a result of the binding of $A_1$.\footnote{As Scheffler notes, in order to feel the full force of the objection, we must add a couple of qualifiers: “$A_1$ has no way out of this dilemma, and... there are no morally relevant consequences of $A_1$’s action or non-action beyond those already mentioned.” (Scheffler, 84)} If constraints can so blatantly fail to minimize the violations of restrictions that they themselves impose, are they not counterproductive, and thus in some sense irrational?

Many think the answer to this question must be “yes,” but we should note that an affirmative answer here relies on an antecedent assumption that the only purpose of constraints is to prevent constrained behavior. The assumption is unsurprisingly common, but on the view I have developed above, it is mistaken. Now, I have just argued that constraints (as the byproduct of agent-centered protections) are instrumental in honoring the natural independence of the personal point of view, and they do so by generally protecting agents’ most important projects. As such, the assumption that gives rise to the paradox seems implicit in my defense of constraints. But I think a more sophisticated account of how constraints protect our projects will help to dispel this worry. To be clear: the challenge being put to the deontologist, as I understand it, is that constraints undermine the very goal they are intended to achieve. That goal appears to be the

}\footnote{As Scheffler notes, in order to feel the full force of the objection, we must add a couple of qualifiers: “$A_1$ has no way out of this dilemma, and... there are no morally relevant consequences of $A_1$’s action or non-action beyond those already mentioned.” (Scheffler, 84)}
prevention (and barring that, presumably the minimization) of constrained behavior. Constraints are counterproductive, the charge goes, because they disallow sacrificial constraint violations that would minimize overall constraint violations of the same gravity.

It is easy to confuse this challenge with the general consequentialist disdain towards any agent-relative device, based on its interference with the maximization of agent-neutral good. These two observations blend together because the point of view from which the paradox becomes apparent is an agent-neutral one—namely, for example, when we ask why it is better to destroy five agents’ personal projects rather than only one. As a result, in classic examples like that of the organ-harvesting doctor, we simultaneously see how constraints seem to undermine their own reason for existence, while also blatantly failing to maximize agent-neutral good.

It is crucial to keep these two points separate because they constitute two fundamentally different issues. The accusation that constraints are self-defeating (in some sense) is a serious one for any deontologist, while the criticism that constraints fail to maximize agent-neutral good is not. Scheffler himself, for example, brushes off the second criticism as it applies to options; he forgives their trespasses against agent-neutral optimality because he thinks their inclusion in an ethical theory serves an important purpose. That is, Scheffler thinks options account for something that agent-neutral optimality cannot: a particular, rational response to the natural independence of the personal point of view—or in other words, some measure of agent-relativity. (More on Scheffler’s position later.) As a result, we must expect a tension between options and impersonal optimizing because each serves a different (and contradictory) master.  

---

92 If a consequentialist challenged Scheffler by saying options permit bad behavior,
The point of this digression is simply to illustrate that if the paradox of deontology can be solved, we should not expect the solution to transmute constraints into a device compatible with agent-neutral maximization of the good. As obvious as that point may seem in the present context, it is not always so, and a confusion of this sort seems to plague Scheffler’s analysis of constraints. He thoroughly debunks the possibility that some feature of the violation, the victim, or the violator could be weightier in the case of one murder than in the case of five. (Scheffler, 89) But these results are entirely predictable, and they establish very little because Scheffler’s questions evaluate constraints from an agent-neutral perspective. Because constraints are an agent-relative device, we can only expect their value to become apparent from an agent-relative perspective.

So if we solve the paradox, we should still be left with an agent-relative device, which is consistent with the idea that constraints help honor natural independence. But how to solve the paradox? If our intuitive support for constraints fully rests on the (apparently mistaken) idea that they will always minimize constrained behavior (and thereby honor natural independence), then constraints truly are in trouble. But the intuitions underlying constraints are robust, as Scheffler himself acknowledges. To do justice to those intuitions, we must evaluate all avenues—one of which involves rejecting Scheffler’s obvious response would be to deny the relevance of an analysis of options from an impersonal-optimizing point of view. If Scheffler thought nothing mattered except the maximization of the impersonal good, then he would not need options in the first place. He simply does not consider (actively) favoring one’s own projects to necessarily qualify as morally bad behavior by default.

More specifically, I mean to say that constraints will not maximize agent-neutral good in the direct, conventional way. One might argue constraints indirectly maximize agent-neutral good, as when indirect or rule consequentialists argue that constraints may maximize agent-neutral good by cultivating certain valuable dispositions in agents.
the assumption that constraints only serve to prevent constrained behavior. Perhaps constraints provide another source of value, and perhaps that source is only visible from an agent-relative point of view.

I can conceive of at least two related stories to tell about the agent-relative value of constraints, and the argument in the preceding chapter lays the groundwork for the first. Constraints help to honor the agent-relative point of view because of the protections that come with them. Both constraints and protections arise when we universalize (D2), and jointly, they allow us to invest in our projects with some level of confidence. When grounded by appeal to the agent-relative point of view, the purpose of protections is not directly to minimize violations, globally considered. Instead, protections line up moral demands placed on us with our psychological limitations, and give our projects some measure of staying power, allowing us to invest in them with the knowledge that their destruction is illegitimate.\textsuperscript{94} There is nothing paradoxical about accepting protections (and thus constraints) for a reason of this sort.

Of course, people violate constraints sometimes, either when they shouldn’t, or because the consequences of the violation justify doing so.\textsuperscript{95} (The view I have developed

\textsuperscript{94} Again, my understanding of how accounting for agent-relativity accounts for our psychological limitations diverges from the conventional view that we may only infer agent-relative duties on Alpha from psychological facts about Alpha. The justification I prefer for constraints on Alpha derives from psychological facts about (for example) Beta, Charlie and Delta. Nevertheless, it remains appropriate on this view to say that introducing constraints brings moral demands into alignment with our capacities because it is the psychological limitations on the abilities of Beta, Charlie and Delta to rank states of affairs objectively that justifies the constraints on Alpha from harming them. My choice of language should not obscure that essential difference.

\textsuperscript{95} Crucially, as Scheffler himself observes, deontologists disagree about the rigidity of constraints. Scheffler draws a spectrum with “absolutist deontologists” on one end and
here most naturally translates into a non-absolutist deontology that accepts certain
constraint violations when enough good stands to be gained, so we might call these
illegitimate and legitimate violations, respectively.) Thus, the mere fact that there is a
protection around a project does not guarantee its safety, even though it helps. The second
dimension of the value of protections/constraints lies in their ability to explain legitimate
agent-relative violations.96 (Any system will be able to give some explanation of
illegitimate violations, so constraints/protections will not be uniquely poised in that
respect.)97

Consider again Scheffler’s version of the organ-transplantation example. Under a
deontological system (with constraints/protections), the doctor must either allow the five

“non-absolutists” on the other. (Scheffler, 85-86) He rightly points out that while
absolutists refuse to violate constraints under any circumstances, non-absolutists can
occupy any spot on the spectrum, balancing the amount of agent-neutral good necessary
to outweigh a particular constraint violation. As will become apparent, there is an
important difference for my purposes between a non-absolutist deontologist and a hybrid
or consequentialist theorist.

96 There obviously cannot be a constraint violation on a view that has no
constraints/protections, so we will create a new term—“agent-relative violation” for such
systems. As far as we’re concerned, an agent-relative violation can occur under any
system, and involves the execution of an action that would be constrained under the
system if it featured constraints. Constraint violations and agent-relative violations thus
overlap in deontological systems.

97 Not all such explanations are created equal, of course. One virtue of the specific account
of constraints that I offer here is that it locates the rationale for various restrictions in the
would-be victim, and it does so without relying on any circumstantial differential between
the value of the people that we might sacrifice and the value of the people that we might
save. It accomplishes this by neatly tracking sacrificeability, yielding extremely intuitive
results. For example, we often think that if we oughtn’t to kill an innocent man and divide
up his organs to save five others, it’s for reasons pertaining to the potential donor’s status
as a moral agent, which has almost nothing to do with him being more important a person
than any of the others who need his organs (and certainly not more important than all of
them combined). We will revisit this point below.
dying patients to expire, or violate a constraint around the sixth in order to save the first five. But if we adopt either a hybrid or a direct consequentialist position, nothing prevents the doctor from sacrificing the sixth in order to save the other five—and in fact the doctor may have to make the sacrifice.

Aside from the likely disparity in body counts, what is the difference between these two angles on the case? The most apparent discrepancy is that the presence of constraints in the former case allows us coherently to regard all of the murders as wrong, whether or not the doctor prevents the first five by committing the sixth. The five dying patients would be, by definition, victims of serious constraint violations, and the only way the doctor could save them is by committing another violation. In the non-deontological case, however, the doctor’s commission of murder violates no restrictions. As long as it saves lives, this murder becomes unequivocally right—or at least, not wrong in any sense.

At the intuitive level, though, something seems wrong with the doctor’s action, even if we think that action is justifiable because of the harm it prevents. Imagine being one of

---

98 Following Scheffler, I shall call a system that includes options but not protections/constraints a “hybrid.”

99 The following appears to speak to Bernard Williams’ intuitions about his famous “Jim and the Indians” case.

100 Actually, depending on the doctor’s personal preferences, a hybrid system can return an even stranger result: The sacrifice of the innocent individual can be less than ideal (though not completely wrong) if the doctor strongly wishes not to perform it for his own personal reasons, independently of the harm caused to the sacrifice victim.

101 A distinction is in order here. If we read the situation in accordance with Scheffler’s interpretation of Williams’ integrity objection, one problem that could arise here would depend on whether the doctor is forced to commit the sacrifice against his own will. This issue afflicts the consequentialist view, but not the hybrid. Because I am lumping the hybrid and consequentialist views together for the moment, I am overlooking any such
the five dying patients who receives life-saving treatment at the cost of another innocent life. Even knowing that four lives in addition to your own were saved by the doctor’s trade, who would not feel that the donor has borne a price that wasn’t his to bear? Hypothetical as it is, I defy even the most ardent consequentialist to escape such a situation without feeling as if she has benefited illegitimately from the sacrifice of another—even if the sacrifice seemed agent-neutrally justified. This feeling of repugnance must arise, most fundamentally, from the agent-relative violation suffered by the sacrificed victim, and that we feel it at all only supports the phenomenon of the natural independence of the personal point of view.

Of course, all of the victims have suffered agent-relative violations in the case at hand. But the consequentialist and hybrid systems, because they lack constraints, possess no feature that allows them to acknowledge the doctor’s intrusion on the sacrificed man’s integrity in the proper, intuitive sense. These systems do take into account the harm suffered by the sacrificed man, but they do so purely as a matter of mathematics. The doctor faces two future states of affairs, one of which involves a single murder and one of which involves five. Choosing the former of these two is morally obligatory because it yields a surplus of four units of good, or four lives, or whatever. But there is no line crossed by the doctor that leaves a moral residue; neither theory offers a formal reason for thinking that there is something objectionable about sacrificing a man because, in this case, neither theory thinks that there is. On the other hand, if the doctor had to choose between a state in which there is one murder and a state in which there are none, relevant differences between the two theories, and am homing in on a different problem instead.
suddenly it becomes unequivocally wrong to kill.

Yet surely the mere circumstances of the case, where a particular man’s sacrifice maximizes agent-neutral good, do not erase the fact that we commit a wrong against him in murdering him. The sacrifice is a tradeoff, a serious constraint violation (committed against an innocent man) executed with the aim of saving five others (who have already suffered serious constraint violations). Only an ethical system with constraints captures both the agent-neutral and the agent-relative sides of the situation, a prerequisite for coherently speaking of a “tradeoff” at all. We might override the sixth man’s protection, but in having to do so, we retain a lingering moral marker, something reminiscent of Ross’ prima facie duties.

This difference between deontological and non-deontological systems highlights this second potential value of constraints, which stands independent of the agent-neutral desire we might have generally to minimize harms. Systems with constraints allow us to “tag” behavior in a way that accords with the personal point of view. In other words, constraints delimit certain actions that are prima facie wrong for agent-relative reasons. And here is the crucial point, alluded to above: it is consistent with natural independence to think that while certain projects are immune to sacrifice in normal circumstances, there is nevertheless some amount of agent-neutral good that will outweigh those projects. Even with violable constraints of this sort, merely having a technical restriction in the form of a constraint preserves our ambivalence toward particular actions, in accordance with the intuitive judgment that they are both wrong from the agent-relative point of view and right from the agent-neutral point of view.102 Justifiable constraint violations are

102 Note once again the Rossian quality of this feature, as well as its alignment with the
justifiable because, weighing agent-relative against agent-neutral considerations, the
agent-neutral considerations are judged to trump the agent-relative ones. But as long as
we adopt constraints/protections, at least we can properly describe the situation, and
acknowledge that an agent-relative duty is at stake.103

This characterization of constraints meets both of the requirements we settled on
above for sidestepping the paradox. Constraints/protections are still agent-relative
devices, but they serve a purpose—two in fact—other than minimizing constrained
behavior. First, they confer agent-relative value by granting legitimacy to us to overvalue
our projects, thereby preventing, at the very least, discretionary agent-relative violations.
How much constraints protect our projects in this sense will depend on the amount of
agent-neutral weight we build them to withstand. Even weak constraints, however, will
prohibit us from dragging one another’s projects to the altar whenever it is personally
expedient. This attenuated limitation is still consistent with the idea, visible from the
agent-relative point of view, that I should generally, or at least sometimes, have veto
power over proposals offered by other agents to sacrifice my own most valuable projects.

Second, simply by being in force, constraints demarcate the agent-relative

sentiments Williams expressed about trade-offs. Note also that the weakest possible
constraint will still have to prevent the sacrifice of something or someone in cases where
committing the sacrifice would contribute increased agent-neutral good. A constraint set
at the exact threshold of agent-neutral maximization is empty, and provides no constraint
at all. This observation accords with our view of constraints as purely agent-relative
devices.

103 Of course, some agent-neutral good is at stake as well, because the survival of the
sacrifice victim has some agent-neutral value. But the thought is that agent-neutral value
is calculated in such a way as to build the loss of that one life into the recommended
course of action.
perspective—even in cases where we ultimately accept the reasoning offered by agent-neutral considerations (for those of us who do), such as when the agent-neutral good to be gained by the sacrifice crosses a certain preset threshold. This is because the mere fact of officially breaking a constraint reflects the agent-relative violation inherent in any sacrifice—even one that may be appropriately aimed at maximizing agent-neutral good.

If this picture is accurate, then not only have we evaded the paradox, but we have grounded the crucial move in the natural independence of the personal point of view. On this view, the paradox of deontology is not really a paradox at all; instead, it is nothing more than a local manifestation of the tension between agent-relative and agent-neutral reasoning, both lines of which I have argued ought to be incorporated into a normative ethical theory.

5.2. A Reply to Samuel Scheffler

In The Rejection of Consequentialism, Samuel Scheffler claims that appealing to the natural independence of the personal point of view can give us a hybrid system (that is, one with options but not constraints). I believe Scheffler is mistaken, and I will now attempt to illustrate why. Once again, I follow Scheffler (and Nagel) in defining the key terms in the debate. For both of us, an option is a moral permission for agents to pursue agent-neutrally sub-optimal outcomes in at least a range of cases, usually in accordance with their personal preference for those sub-optimal states of affairs. Meanwhile, a constraint is a rule that prohibits agents from performing certain actions even in cases where performing those actions would render a better state of affairs (agent-neutrally considered).
As we have noted previously, both of these features are prevalent in commonsense morality—and Scheffler certainly accepts that much. But Scheffler also thinks there is a key asymmetry between options and constraints: options, he thinks, find a principled rationale in the natural independence of the personal point of view; constraints, on the other hand, seem not to have any principled rationale at all. Scheffler calls this the “asymmetry thesis,” and his book is devoted largely to a defense of that thesis (as well as its corollary, the “independence thesis,” which is that one can accept options without necessarily accepting constraints).

Scheffler’s defense of the asymmetry thesis starts from the following question: if an agent already has an option to act in a way that does not produce the best state of affairs, what additional value can there be in forbidding him from bringing about the best state of affairs? To justify constraints, Scheffler argues, we need an additional rationale for blocking agents from doing what is impersonally best. Despite briefly considering the question, Scheffler is unable to discern how constraints may contribute to honoring an agent’s individual perspective in a way that options somehow omit. True, he says, unlike options, constraints require holding agents to a “standard of accountability which gives the intrinsic qualities of an individual’s acts moral priority over their impersonal optimality.” (Scheffler, 96) Thus, in some sense, they appear to place more emphasis on the individual perspective.

But Scheffler thinks this appearance is an illusion because it depends on the false assumption that agents cannot, “from [their] own point of view, form and act on intentions to do things that will have the best overall outcome, either because they will have the best outcome, or for other reasons.” (Ibid) The reasoning here seems to be that, if constraints
can sometimes clash with the choices agents make from their own perspectives, then such restrictions will begin to “appear... arbitrary,” and so lose their status as champions of the individual point of view. (Scheffler, 97)

Yet Scheffler’s analysis of the asymmetry thesis appears deficient. Conspicuously absent from his considerations are the effects of options and constraints on the people surrounding the agent (who Kamm would call “patients”). Let us adopt Kamm’s terminology. Someone making a choice in a given scenario will be an “agent” and all others (but most importantly, all those who are also present in the same scenario) will be “patients.” In other words, patients are those who stand to be (but are not necessarily) affected by the choices of agents.104 Options and constraints affect both “types” of people, and affect them differently. Options allow agents to deviate from impersonally optimal outcomes, but also allow them to pursue the agent-neutrally best outcome—both of which can involve trampling on the personal projects of patients. Constraints, meanwhile, do more than prevent agents from maximizing agent-neutral good; they protect all patients

104 Everyone is simultaneously an agent and a patient with respect to different decisions. Most of us are actively making decisions frequently throughout our waking hours, and lapse into passivity periodically in between choices (as well as when we are sleeping). In a given scenario, however, I will tend to emphasize one particular decision and its effects. In most examples, then, the relevant status of a person as an agent or a patient will be determined based on the agent’s position vis-à-vis one particular decision. Typically, one is an agent vis-à-vis a decision if one is making that decision and a patient if one is not making it. Consider an example. Suppose A is a pedestrian walking down the sidewalk, trying to decide what to eat for dinner, and B is a driver who makes the choice to swerve onto the same sidewalk in order to avoid a pigeon. Both A and B are agent in relation to a particular decision (they are both engaged in decision-making). B is an agent in relation to a choice that now affects A (because B chose to drive on the sidewalk, thereby threatening A). In relation to B’s choice to drive on the sidewalk, A is a patient. Similarly, B is passive with regard to A’s selection of a meal, but this is less interesting because A’s choice does not impinge on B.
and their respective projects from agents.

This is roughly the sort of point that drives Kamm to adopt a patient-centered approach to grounding constraints as well. She would likely agree that by fixating strictly on agents, Scheffler exposes only the positive aspects of options (that they are liberating) and the negative aspects of constraints (that they are restrictive). Conversely, had he solely considered patients, Scheffler would have revealed only the negative qualities of options (that they threaten the projects of others) and the positive qualities of constraints (that they protect the projects of others). But everyone is either active or passive with regard to any decision made anywhere during her lifetime, for these categories are mutually exclusive and jointly exhaustive. So neither perspective should be omitted because, taken together, acting as an agent and being acted upon as a patient fully capture the natural independence of our point of view.

A couple of cases will highlight the problem with addressing natural independence solely from the agent’s point of view, and will thus help to show why even Scheffler’s intuitions do not respond well to the hybrid theory he proposes. Imagine Ferdinand, a poor widower, dictating a will to Ollie, his attorney. Ferdinand wishes to leave his paltry savings to his daughter, Ferdinancy, even though he knows that it would be (impersonally) best to leave the money to the American Red Cross. Luckily for Ferdinand, Scheffler permits him this option. Unbeknownst to Ferdinand, however, Ollie is easily seduced by impersonally optimal outcomes. Moreover, Ollie is in a position to change Ferdinand’s will without anybody ever finding out. Ferdinand will expire shortly, Ollie

\[\text{105 For Kamm, the merits of constraints might be that they confer some sort of status on the person who holds the correlative protection, but the point surely stands nonetheless.}\]
reasons, and Ferdinancy would just assume that her father chose to donate his meager savings to charity. Poor Ferdinand. In his enfeebled state, he is unaware that Ollie’s freedom to pursue the optimal outcome is unchecked by a constraint.

Now, Ollie is fortunate to have an option too. If his personal projects include upholding his commitments to his client, then Scheffler may allow him to forego agent-neutral gains in the name of honoring Ferdinand’s wishes. But let us suppose that Ollie is drawn the other way, as Scheffler would also permit. Ollie changes the will, though in doing so he feels a small pang at sacrificing his own projects. Intuitively, this result raises a flag: not only does it seem right that Ollie should be able to uphold his commitment to Ferdinand—as Scheffler’s theory allows—but it seems as if Ollie should defer to Ferdinand—as a deontological theory might demand.

Suppose, now, that Ollie leans towards maximizing agent-neutral good because doing so aligns with his personal projects. Perhaps after retiring from the legal profession, Ollie intends to write a book about his experiences maximizing agent-neutral good. This case contradicts the “false assumption” Scheffler thinks we must make in order to believe that constraints help protect the integrity of the individual perspective. But instead of erasing the appeal of constraints, it now seems like Ollie should not only lack the option of maximizing agent-neutral good; he also should be constrained from pursuing his own projects out of respect for Ferdinand’s.

Here we see the agent’s choice carrying significant consequences for the project of a patient. More specifically, we see that both Ollie’s freedom to maximize impersonal

---

106 The crucial decision here is Ollie’s. Thus, per footnote 104, Ferdinand is a patient in the relevant sense (that is, vis-à-vis the alteration of the will).
good and his option to pursue his own goals can threaten to nullify one of Ferdinand’s important goals.⁹⁷ Options sans constraints, it becomes clear, can cancel each other out in a most disturbing way.⁹⁸ And this tension only becomes more pronounced as the stakes increase. The more Ferdinancy needs the money, and the more Ferdinand cares that she receives it, the worse Ollie’s interference becomes, even though that interference stands to maximize agent-neutral good. To take a more radical example, what if Ollie could maximize agent-neutral good by smothering Ferdinand with a pillow, thereby freeing up important medical resources for other patients? Surely Ferdinand’s (even temporary) survival constitutes one of his most important projects; but the absence of a constraint

¹⁰⁷ Ollie’s unchecked option would remain minatory even if Ferdinand aimed to maximize impersonal good.

¹⁰⁸ Tom Hurka points out that accepting constraints without options would yield its own set of problems. This angle on the question is perhaps overlooked because of the general consensus (as a result of Scheffler’s work, I suspect) that options are easier to justify than constraints. Yet it is an interesting point to consider. Imagine a transplant case in a world with constraints but no options. In such a world, the doctor would have a duty to sacrifice the healthy person; and the sacrifice victim would be obligated to permit the doctor to kill him. The doctor would thus be enforcing a duty that the sacrifice victim independently possessed, which seems to undermine the wrongness of killing him. On the other hand, if the victim has an option to resist sacrifice, then killing him seems worse because we are forcing on him to bear a moral burden that he isn’t obligated to bear. We should also note that options balanced with constraints can still cancel each other out, but such cases are much less abrasive to our intuitions. For example, I may have a goal that involves walking down the street without having to adjust my path to accommodate anyone else. No reasonable series of constraints I can think of would protect the integrity of such a silly goal from the innumerable ways in which others could interfere with it. In fact, the reason that searching for reasonable constraints in this case is fruitless is because the project would generally be described as silly. It just is not the type of goal we deem worthy of protection. If we did, that would translate into me having a moral right to walk down the street unimpeded, as well as everyone else having a duty not to get in the way. Instead, we depend on constraints to protect goals that we take especially seriously. A more precise formulation of how constraints function, then, is to protect our most valuable projects and goals from the inherent threat posed to them by other agents.
protecting his life turns Ollie’s freedom to pursue the agent-neutrally optimal outcome into a terrible threat.\textsuperscript{109}

Scheffler is certainly aware of the challenges raised by accepting options but not constraints. After all, Kagan has raised a similar but more powerful version of the present objection in a review of Scheffler’s book: “prerogative[s] will not only permit me to allow harm, (sic) [they] will also permit agents to do harm in the pursuit of their nonoptimal projects.” (Kagan 1984, 251) As an example, Kagan suggests that if we would not be obliged to pay $10,000 to save our uncle, on Scheffler’s view we would also be free to murder our uncle for $10,000. (\textit{Ibid}). Scheffler has a response to this objection in an appendix in the revised edition of his book. Scheffler challenges Kagan’s implicit assumption that “avoiding killing will quite often be just as costly for agents as preventing death. (Scheffler, 178) Scheffler also challenges the notion that the general costs to everyone would be the same if someone were to allow death rather to commit murder. The reply does not strike me as especially compelling, for even if both of these rejoinders are on point, they seem rather narrowly targeted at the example at hand. Moreover, if there is something worse about killing than there is about letting die, it certainly does not seem to be the cost to the person doing the killing or the cost to society at large of being a killer.

Returning to the case of Ferdinand, at any rate, the consequentialist will remind us that those who stand to benefit from the changes to Ferdinand’s will (or even from the smothering of Ferdinand) also overvalue their own projects. We must thus weigh the

\textsuperscript{109} Even though Scheffler will likely want to disallow such sacrifices, the simplest and most plausible way of doing so—invoking a constraint—is unavailable to him.
number of overvalued projects to be protected or satisfied by honoring Ferdinand’s
wishes against the number of overvalued projects to be protected or satisfied by
overriding them. Supposing that more people stand to gain from disregarding Ferdinand’s
preferences than stand to gain by upholding them, the consequentialist might suggest that
we have an agent-relative reason (and certainly an agent-neutral reason) for allowing
Ollie to break out his pen or grab the nearest pillow. How would the possibility of
honoring more projects (overvalued from the respective agent-relative points of view) fail
to constitute an agent-relative reason in favor of doing so?

While it is perfectly true that the would-be beneficiaries of Ollie’s actions also
overvalue their projects, we have seen above that it is literally impossible to develop an
agent-relative reason for assisting them. We cannot simultaneously give everyone
assistance beyond that to which they are objectively entitled, and we are thus unable to
cash out the consequentialist’s concern in agent-relative terms. While there may be some
agent-neutral reason for overruling Ferdinand, there is no applicable agent-relative
reason—except, perhaps, special duties that Ollie might have, or the overwhelming
(though still probably insufficient) strength of Ollie’s personal wishes to maximize.

Equipped with a fuller understanding of the interplay between options and
constraints, we should now reconsider Scheffler’s explanation for accepting the former but
not the latter. Recall that Scheffler is responsive to the idea that each person stands
“naturally independent of the impersonal standpoint.” (Scheffler, 79) The major appeal of
the hybrid view for Scheffler is thus that it allows agents to “devote energy and attention
to their projects and commitments out of proportion to the weight from the impersonal
standpoint of their doing so” and he claims that such a theory consequently “recognize[s]
and mirror[s] the independence of the personal point of view.” (Scheffler, 58)

But is this really true? Do hybrid theories actually recognize and mirror the independence of the personal point of view? They seem to do so only to a limited extent. Where they succeed is in recognizing that we often prefer to favor the pursuit of our own projects at the expense of impersonal optimality. However, what eludes hybrid theories is the fact that our projects matter to us continuously, not just when we are actively engaged in making decisions that affect them. Ferdinand still cares quite a bit that his daughter receives his inheritance, even after he has dictated as much to Ollie. He only becomes passive towards the matter because, once he has conveyed his wish, he has done all he thinks he needs to do to achieve that end. Likewise, Ferdinand’s survival matters to him tremendously, even when he is not actively engaged in fighting for it—as when he is in repose in a hospital bed. Ferdinand is only passive towards his continued existence because he has no reason to assume he is under (immediate) threat.

We can restate the point by saying that the personal point of view shadows every agent all of the time, whether he is active or passive in relation to the decision in question. And in at least one salient sense, it is just as disruptive for an agent morally to be prevented from pursuing his projects as it is to have his projects damaged or destroyed by another agent (acting either out of self-interest or to promote the optimal agent-neutral outcome). This is why merely allowing agents to favor their own projects does not satiate our intuitions. Properly to “recognize and mirror the independence of the personal

110 It is also why Scheffler’s example of a machine called the “Infallible Optimizer” fails fully to assuage those same intuitions. Even if an agent had the option of pursuing his own goals, and he knew those goals would only be sacrificed by others when doing so would truly minimize agent-relative violations of the same gravity, we would still be reserving
point of view,” then, we need to acknowledge both aspects of that point of view: we need options to secure our freedom to pursue our own projects and we need constraints to protect our most important projects from the actions of other agents. The natural independence of the personal point of view, qua principled rationale, thus appears to underlie both options and constraints.

This conclusion has enormous implications for Scheffler’s argument. In the first place, it undermines his “asymmetry thesis.” There is no asymmetry between options and constraints with respect to the natural independence of the personal point of view; instead, both devices draw on that natural independence. It is simply not true that options rest on principled foundations, while constraints merely flap in the wind. Indeed, Scheffler’s own intuition betrays the true asymmetry in play here: that we should accept what is intuitively plausible about options, but not what is intuitively plausible about constraints, even when those intuitions come from the same source. Further, Scheffler’s “independence thesis” relies entirely on the “asymmetry thesis” for support. By toppling the “asymmetry thesis,” we thus collapse the “independence thesis” as well.

My suggestion is that Scheffler’s mistake arises from his insistence on combing the right to interfere with his most precious goals in a way that none of us are psychologically capable of accepting. In so doing, we would simultaneously be reserving the right to override the natural independence of his perspective as a patient without even tipping our hat to that independence.

I should emphasize one more time that this choice of language does not amount to a retreat from the position that the same facts about Alpha can give options to Alpha and constraints to Beta, Charlie and Delta. On my account, those are the two sides of Alpha’s natural independence.

...Barring, of course, the discovery of a different rationale that establishes options but not constraints.
through the possible value of constraints strictly in terms of agency. This approach, which causes him to overlook the relationship between constraints and patience, may arise from his narrow construal of Williams’ integrity objection. In Scheffler’s own words, “a central feature of [Williams’ integrity objection], both as formulated by Williams and as I have been construing it, is precisely its insistence that the theory violates the integrity of agents by virtue of what it directly requires them to do.” (Scheffler, 38, emphasis in original) Such a construal clearly favors agency, and seems to obviate the need to consider the harms a moral system permits us to visit upon one another in the name of minimizing agent-neutral damages. Scheffler admits that such harms, which might often take the form of sacrificing an otherwise safe individual or her projects, “may violate the victim’s integrity.” (Ibid) Further, he concedes that such “victimization” may be considered “objectionable.” But, according to Scheffler, “that is a different objection than the objection to consequentialism dealing with integrity.” (Ibid)

While I am not entirely convinced on this score, I can safely concede the point nonetheless. Shortly after the exposition quoted above, Scheffler abandons Williams’ integrity objection altogether. He expressly cites the objection as the inspiration for his development of the rationale for options. But his strategy is to “argue that whether or not consequentialism violates integrity, there exists a rationale of the appropriate kind for such a prerogative.” (Scheffler, 55) Once again, however, the rationale Scheffler refers to here appeals to the natural independence of the personal point of view. By making this move, Scheffler clears the way for objections claiming the relevance of patience as

113 The actual wording of Williams’ integrity objection is irrelevant for my purposes, as will become clear shortly.
long as we ground that aspect of our agency (broadly considered) in the same natural independence. Thus, the spirit in which Williams put forward his original objection is simply immaterial.

To sum up: the principled rationale Scheffler accepts for options appeals to the natural independence of the personal point of view. His claim is that this rationale establishes options, but does not establish constraints. I have tried to show that Scheffler’s conclusion is mistaken because he only considers the individual point of view for agents engaged in decision-making (that is, for agents). My claim is that the natural independence of individual points of view remains relevant when agents are not engaged in decision-making (that is, when they are passive). Thus, it is inconsistent for Scheffler to account for the active individual standpoint, but not the passive individual standpoint.

Further, I have tried to show that the individual point of view for agents establishes options and the individual point of view for patients establishes constraints (or more precisely, protections, which bring with them constraints). Thus, in order to honor both the active and the passive individual points of view—or in order fully to recognize natural independence—one must embrace both options and constraints. In other words, I have posited that options and constraints rest on the same principled rationale, and thus that both Scheffler’s “asymmetry thesis” and his “independence thesis” fail. Therefore, if my argument is correct, Scheffler has three choices: he must accept both options and constraints; reject both options and constraints; or seek out a new principled rationale for options that legitimately excludes constraints.

The rough idea in using my approach to justify them is that, if everyone is morally permitted to overvalue their projects, then we have a legitimate reason for offering
additional protection to their projects—both in allowing them to pursue those projects at
the expense of agent-neutral good, and in disallowing the sacrifice of those projects for the
purpose of promoting agent-neutral good.

It is worth taking a closer look at this justification for constraints, however, for it is
easy to fall into confusion about how it is meant to work. In particular, it can be tempting
to assume that treating everyone equally will require a consequentialist approach that
does not take account of the agent-relative boosting we all inevitably engage in. After all,
we might think, if everyone overvalues his or her projects, then in any given case where a
tradeoff stands to be made, we are forced to choose between violating one overvalued
project or more. Mathematically, those cases seem just as compelling as any other
consequentialist-friendly case, and therefore it seems unlikely that an approach of the sort
I have in mind stands to succeed.

But this objection hinges on a misunderstanding of what it means to take account
of the personal point of view. Conferring normative significance on agent-relative boosting
does not require treating people unequally. Broadly speaking, there are at least two ways
of treating everyone’s projects equally—one that ignores the natural independence of the
personal point of view and one that takes account of it. Neither of them is unfair on its
face. Instead, both are fair in different ways. But I have argued that failing to account for
the natural independence of the personal point of view is unreasonable, and that any
approach that ignores our natural independence is inherently and fatally flawed. If those
arguments succeed, then we have a reason for selecting the agent-relative-friendly model.

Under the purely agent-neutral approach, we would start by giving everyone equal
weight all of the time. We would stipulate that, all else equal, each life (or, say, unit of
good) counts for the same and must be treated as such, in all cases and at all times.

People’s lives and interests are fungible, to be traded off for the purpose of maximizing the good wherever possible. That would be the crude consequentialist route—an approach that yields so many counterintuitive results that even consequentialists generally temper it. As far as it concerns us here, there are two particularly distinctive features of this method. On one hand, it is phenomenally demanding (or even alienating), in its purest form requiring agents to spend all of their time maximizing agent-neutral good as efficiently as possible. At the same time, this approach also reduces agents to beads on the moral abacus. Each of us is to be treated as a mere vessel for storage of agent-neutral good, to be emptied or filled at the mercy of moral circumstances. This is, in short, the position we seem to be forced into by the misguided objection I have described above.

We are not forced into this situation, however. Alternatively, we could give everyone a moral buffer of equal strength, which carves out a zone around them for the pursuit of their own projects and goals. Within that zone, agents are permitted to pursue agent-neutrally sub-optimal outcomes, as well as finding their projects “unsacrificeable” for agent-neutral gains. This approach is also fair in that everyone receives the same level of protection, even though there are cases where more projects will be destroyed rather than fewer because some minimizing tradeoffs are impermissible. Giving everyone equal treatment in this way has several important advantages over the alternative: most fundamentally, by recognizing the inherently self-referential qualities of our agency, it would incorporate the natural independence of the personal point of view into our theory of the right. In doing so, it would allow us to ground both options and constraints, which are both intuitively essential components of any such theory.
The size of the buffer zone dictates the scope of the options and the rigidity of the constraints, but as long as these agent-relative zones do not extend indefinitely, we can continue to tip our hat to consequentialist pressures. In theory, if we were to adopt such zones without limit, we could derive absolutist deontology; but the approach I have described seems committed to limiting the size of these zones. A major component of the view is that agent-neutral considerations have a role to play in how we ought to rank states of affairs, so it would be difficult to see how we could give agent-relativity lexicographical priority over agent-neutrality.

In any event, the latter of these two approaches above is consistent with preserving the agent-relative point of view in a way that the former is not. The former requires reducing the value of everyone’s projects to nothing more than their agent-neutral value. That is, almost paradoxically, what happens if we try to take account of agent-relativity by giving all projects equal weight all of the time. Natural independence is not reflected in the normative structure of that sort of theory of the right because in order to account for it we need to give projects a fundamentally different kind of value: agent-relative value. Agent-relative value is defined against agent-neutral value, and we can only coherently say that we accept boosting if we allow various projects to stand against maximization. Further, even though more projects overall are preserved on the fully agent-neutral approach, our intuitions reject it because our agency inevitably includes a feature that confers agent-relative value on projects we engage in.

The idea here is that, if we accept the second approach described above, every agent will have a duty not to violate constraints unless he has adequate reason to think that sufficient good will result (assuming that we do not make the constraints inviolable).
And it is worth noting that adopting an approach of this sort will not force us to ignore agent-neutral prescriptions in all cases. Aside from enormous tradeoffs, there may also be cases where we have no choice but to violate constraints no matter what we do. Suppose that Alpha loses control of his car while driving recklessly, and he finds himself careening toward a crowded sidewalk. If the only option available to him is to swerve sharply towards a smaller group of people, then Alpha’s choice truly comprises nothing more than a comparison between violating more constraints and violating fewer. The buffered approach and the agent-neutral approach are likely to recommend the same course of action in such cases because mathematical considerations are all that we have.

Such situations will be unusual, however, and the protections around each of us will come into play in a wide range of cases. Suppose we accept the impermissibility of saving several ailing individuals by cutting up an innocent person and redistributing his organs. On my view, the reason for this is that the buffer around the innocent person prohibits it. The choice we are faced with in this situation is quite different from the choice faced by Alpha above. Here we wrestle with whether to violate one constraint (cut up the man) or violate none (allow the others to expire). Assuming we want constraints strong enough to block one-for-five sacrifices, then the decision is clear—even though we may recognize the objective (agent-neutral) desirability of the state of affairs in which the five survive rather than the one.

5.3. Shelly Kagan’s Skepticism

Shelly Kagan offers perhaps the most prominent rejection of deontology (at least of those articulated in the terms that I have used), and it is important to address exactly why
his arguments do not nullify the result I argue for above. Earlier I noted that, like Scheffler, Kagan also rejects constraints on the grounds that he can locate no sufficiently robust rationale to support them. Of course, Kagan goes further than Scheffler, since he rejects both options and constraints instead of merely the latter. He is less impressed than Scheffler is with natural independence of the personal point of view. He argues—based on what I have claimed above is an inappropriately weak interpretation of natural independence—that such independence only gives us license to explain people’s reluctance to maximize the good, not justify it.

Still, before rejecting anything, Kagan considers how we might justify both devices. More specifically, for our purposes, he canvasses a series of five possible descriptions of how we might attempt to justify constraints.114 All but one of these five draw on intuitions that can be explained by the justification I am defending, although Kagan does not articulate any of them in the way that I would. The only manifestly fruitless approach to justifying constraints that Kagan considers (which I will call Base 1) is an appeal to the (impersonal) badness of committing any constrained act. No matter how bad it is to murder, for example, two murders will be twice as bad (all else equal). On this point, I agree with Kagan and Scheffler; appealing purely to the badness of constrained acts is simply insufficient for grounding a rule that explicitly rules out minimizing such acts. This approach is a nonstarter, most fundamentally, because it leads to the paradox of deontology.

All of the remaining descriptions considered by Kagan have a ring of truth to them

114 I will leave aside Kagan’s rejection of options because we have effectively addressed that above in defending the natural independence of the personal point of view.
that I believe I can account for. It will be easiest to list them here so that I may refer back to them as necessary.

Base 2: “Moderates have often emphasized the quality or character of the acts they wish to constrain.... If I kill my victim, I stand in a particular relationship to him.... Presumably I do not stand in these horrible relationships with all those whom I merely allow harm to befall.” (Kagan 1989, 28)

Base 3: “[Moderates feel] there is something deeply wrong about the extremist's willingness to do evil in order to prevent it.” (Kagan 1989, 29)

Base 4: “[Moderates] may suggest that to deliberately harm another degrades not only the victim of the harm, but the agent of the harm.” (Kagan 1989, 30, emphasis in original)

Base 5. “The moderate... may suggest that if I kill the one to save the two, I am failing to respect the one as a person.” (Kagan 1989, 31)

Now, Kagan develops these bases in more detail than I have given here, but I have reproduced what I take to be the core of each. Kagan rejects all of these bases on structurally similar grounds. That is, in every case given above, Kagan claims that the reasons offered by the deontologist are insufficiently strong to justify rules that (sometimes) inhibit one’s ability to bring about the optimal state of affairs.

I have already presented a positive justification for constraints above. Constraints (of some unspecified strength) are essential for reflecting the natural independence of the
personal point of view because they safeguard that independence with regard to our patientness. That this particular “base” does not find itself on Kagan’s list means only that Kagan has not overtly attempted to refute it. The ostensible failure of the approaches that Kagan considers above is thus no indictment of my own approach.

Nevertheless, I think my approach can help to explain the intuitions behind each of these bases, and it may be instructive to point out how. Base 2 and Base 4 seem to address our own feelings about committing bad acts, even when we are asked to perform such acts for their beneficial effects. I agree with Kagan that these bases fail. The reason that they fail is that they draw on the agentic intuitions that justify options, not constraints. Bernard Williams’ case of Jim and the Indians is perhaps the paradigm example that stresses the sort of point behind these two descriptions. Briefly put, Jim is offered the chance to kill one person; if he declines, a third party will kill 20 others. (Williams, 54) The optimal state of affairs is supposed to be the one in which only a single murder occurs, but Williams stresses that it is legitimate for Jim to feel reluctant to follow the numbers.

On the view I have proposed, we can describe Jim’s dilemma in terms of the tension between his legitimate preference for sub-optimal states of affairs organized around his interests and his reasoned understanding that one murder is objectively less bad than 20 murders. Supposing Jim is generally opposed to committing murder, the natural independence of his point of view may entail that he quite unavoidably attaches much more negative value to states of affairs in which he commits a murder than states of affairs in which someone else commits several. (Whether and why he might have a duty to do so is something I take up below.) So his moral default will be to decline the offer and allow 20 deaths, though he fully recognizes the sense in which it would be better to follow
consequentialist reasoning. After all, 19 lives are objectively worth quite a bit. Of course, even if we accept constraints (for the reasons given above), we might think that the numbers here are great enough that Jim has a moral obligation to kill one person. The point is that, whatever stance we take on that question, we can explain the depth of the dilemma in the terms I have defined above—though again, the dimension of the Jim case evoked by its application to Base 2 and Base 4 concerns Jim’s option to decline his invitation to commit murder. We do not see constraints in this picture because the bases in question fail to bring them out.

Meanwhile, Base 3 and Base 5 both appeal to the intuition that each of us should be protected by the buffer offered by constraints. Returning to the case of Jim and the Indians, each of these bases appeal to the natural independence of the Indian who stands to be sacrificed. We intuitively feel that the Indian’s life is protected by a moral boundary, such that he should not be dragged to the altar of impersonal expediency simply because some lunatic threatens his acquaintances. Even if we think that the Indian should be sacrificed—say, because the impersonal badness of the states of affairs in which he survives is so great—we still recognize that there is something regrettable or unfortunate about the fact that he had to die in order to save the others. That sense of each individual’s inviolability is what drives Kagan’s description of Base 3 (as our reluctance to do evil for the purposes of producing good effects) and Base 5 (as our failure to respect the Indian as a person).

The reason that Kagan seems to have missed the approach I advocate is very

\[115\] Williams seemed open to this possibility as well, although uncertain about it.
similar to the reason for which Scheffler did the same. Kagan seems to share Scheffler’s intuition that there is an asymmetry between options and constraints. In particular, Kagan indicates that he sees how our natural independence cries out for options, but not for constraints. He claims that “...a concern with the moral integrity of the agent cannot provide a justification for having a constraint against doing harm. At the very best it could only help to ground an option to refrain from harming others in pursuit of the good.” (Kagan 1989, 31) It is clear that Kagan and Scheffler share identical concerns here. Put most succinctly, "If I am willing to kill one out of my concern for the two, then the moderate still needs to provide a reason to forbid the killing." (*Ibid*)

Kagan, like Scheffler, insists on considering the value of constraints from the standpoint of agency. As I noted earlier, it is perfectly obvious that a given constraint does not help protect the natural independence of the actor constrained by it; Scheffler and Kagan are both correct to point that out. Rather, that constraint validates the natural independence of the agent whose projects it protects. In other words, if you are to ask how constraints help someone by constraining them, you are simply asking the wrong question.

5.4. The Promise of the New Approach

The approach to deontology that I have advocated here appears to have the resources we intuitively want to see in a theory of the right. I cannot develop each of these in detail here, but a brief sketch may be in order. To begin with, natural independence puts the wrongness of various acts where it seems to belong. To kill Delta generally involves
breaking a constraint imposed on us by some quality of Delta. The reason it is wrong to kill him, if it is, tracks some fact about his psychological constitution that we have a (plausible) objective duty to respect. That does not mean we may never be justified in killing him, even if he’s innocent; but it raises the threshold for such sacrifices while pointing to the spot of the foul.

To take a more complicated example, consider Williams’ famous case of Jim and the Indians. Imagine that I can stand aside and let someone kill 20 people or, at the invitation of the would-be killer, perform one killing myself to get the others off of the hook. One factor counting against killing one person myself is that my duty not to violate constraints is stronger than my duty to minimize overall constraint violations (as evidenced by the asymmetry in universalizing (D2) and (D3)). Further, I may legitimately overvalue the state of affairs in which I am not guilty of a murder, even for a good cause. On the other hand, the consequences of stepping in are nontrivially superior: 20 deaths are, all else equal, much worse than one. However we wish to weigh these considerations, it seems as if accounting for natural independence (and mitigating it with consequentialist considerations) lets us mark the appropriate points in a moral dilemma.

I also think that natural independence gives us the tools to make some reasonable observations about the distinction between doing and allowing. From the agent-neutral point of view, it isn’t worse to do evil than to allow it; but we still regard the distinction as morally meaningful. On the view I have outlined, the explanation would thus have to be agent-relative. I think that the intuition behind the distinction can also be explained by natural independence, though I have no way of proving that claim. We seem to regard our relationship to a state of affairs as relevant in a way that the objective point of view does
not. At the very least, natural independence can give a plausible explanation as to why: acting (as opposed to omitting) is the sort of behavior that plausibly confers additional agent-relative value on a particular state of affairs. That is, we might want it to be the case that \( x \) obtains; but we are likely to place more value on the state in which we are responsible for bringing about \( x \) as opposed to the state in which we are not. If we do not want \( x \) to obtain and it does so despite our wishes, the result will be the same with the valence flipped: being responsible for \( x \) will generally make it worse to us than merely allowing \( x \) to happen.

On the view I have advanced here, we can allow certain things to happen that we cannot ourselves do, but the reason for that difference is a function of what happens when we universalize (D2) and when we try unsuccessfully to universalize (D3). As discussed above, it is logically impossible to give everyone agent-relative assistance, so we can’t justify providing it in the same way that we can justify giving everyone agent-relative protections. As a result, the topography of the normative landscape built around agent-relativity provides for an asymmetry between doing something bad (in the form of constraint violations) and permitting something bad (again, in the form of constraint violations).

In a more developed version of the view I have begun to sketch here, I would attempt to incorporate the claim that our agent-relative relation to the way in which states of affairs came to be helps explain the moral asymmetry between doing and allowing. The primary reason for doing so is that I happen to think it’s true. Note also, though, that the universalization of (D1) and (D2), if carried out the manner advocated above, produces a further asymmetry between doing bad action and allowing it (on one hand) and on doing
good and permitting it (on the other). The structure of the view can prohibit us from violating constraints in a way that it cannot prohibit us from permitting bad things to happen. There is thus a formal difference between doing evil and allowing it—and that difference persists (on most views of the good) irrespective of the personal feelings of agents. (Someone with wacky state-of-affairs valuations would still be subject to the prohibition on breaking constraints.) With respect to the good, however, there is no formal feature of the view (at least as I have developed it so far) that recognizes a difference between doing good and allowing it to come about.

In the first instance, this particular difference counts in favor of a stronger view on the wrongness of doing evil than on the rightness of doing good. I find such a result intuitively palatable though perhaps others will disagree. Nevertheless, it seems as if the view should also be able to distinguish between doing good and allowing it to happen—at least inasmuch as we are to get an intuitively comfortable theory of the right out of natural independence. Here again we see a reason for considering whether natural independence might explain the differential personal valuations of states of affairs based on how they came to pass. To the extent that actively bringing about something requires effort (whereas permitting things to happen does not generally require effort, or at least not require as much) it is plausible to think that one feature of natural independence is that it skews valuations of states of affairs to some degree in accordance with our investment in them.

This strikes me as a promising route because one reason for thinking that agent-relativity matters in the first place is that the projects we invest in come to have personal value to us that far outstrips their agent-neutral worth, notwithstanding our recognition of
the agent-neutral absurdity of such overvaluation. It seems to me equally true that our conception of the place we hold within or without the causal chain leading to various states carries tremendous implications for our agent-relative attitudes towards those states. Accepting the truth of some such claim would go a long way to helping explain another feature of commonsense morality: supererogation. On a view of this sort, supererogation would best be understood as acknowledging, in various respects and above the level that we are expected to given our natural limitations, the relevance both of others’ agent-relative considerations and of agent-neutral considerations.

Additionally, natural independence may be able to give a reasonable explanation of the moral relevance of the salience of various states of affairs. One of the reasons why so many believe consequentialism to be overly demanding is its insistence that, all else equal, there is no moral difference between things that happen in front of our faces and things that happen far away. Agent-neutrally speaking, this is true of course; but if agent-neutral considerations are no longer consistently decisive, we need to reassess our stance towards salience.

If we were to regard evil at a distance as carrying the same weight as proximal evil, then consistency (as Singer, Unger and others remind us) requires us to treat them both equally. Whatever controls we put into place to correct for our doubts about the efficacy of efforts to combat evil at a distance, and even if we screen out all but the most egregious evils, there is little question that the natural independence of the personal point of view would be crushed by our moral demands. To the extent that a moral theory accepts natural independence as a basic fact about our psychology, this result cannot stand. What we need is a way—preferably principled—of limiting the number of claims upon us to a
manageable number.

Salience offers one way of achieving this aim. We can visualize moral salience as a set of concentric circles. If we cut down the moral demands on everyone with each more distant circle, we have a natural way of whittling down the overwhelming burden we start with on the consequentialist view. Still, I do not think that this alone justifies the recognition of salience as morally relevant. The same effect could be achieved by implementing a rule that we will only act on every fifth or 10th or 50th demand, or by any other number of methods.

Yet salience, unlike most other methods for imposing such limitations, dovetails with agent-relativity. One of the key measures of the agent-relative value we see in a given state of affairs seems to be the amount of time or energy we have spent working towards it or even contemplating it. Salient states of affairs, essentially by definition, make a bigger impression on our agent-relative valuations than do unnoticed ones. If we can only reasonably handle so many agent-neutral claims, sorting them out (at least to some extent) by agent-relative valuation seems a good way to go, both from a practical standpoint and a utilitarian one.

Moreover, adopting salience may actually be essential to understanding the nature of at least some special duties. One reason—though probably not the only one—why we may legitimately have expectations of others is that our prominent place in their lives gives them a duty toward us. This might at least help to explain duties to one’s children, or friends, or neighbors. Salience may thus be objectively morally relevant if we accept natural independence as a fact of normative significance.

None of this is to say that we should adopt a lax moral theory. In light of our ability
to recognize considerations that are not to our benefit and to act accordingly, whatever moral considerations follow from natural independence can be justified only within limits. Specifically, there remains the enormously difficult question of figuring out how to balance agent-neutral and agent-relative considerations. All along I have assumed that there is some threshold at which constraints should break down in the face of agent-neutral gains. Wherever we place that threshold, justifiable constraint violations will still leave a residue, much like Ross’ outweighed prima facie duties. A constraint violation committed against Delta, justified on the grounds that it is sufficiently beneficial to Alpha, Beta and Charlie, still constitutes a wrong of sorts. The broken constraint officially marks the need for making amends, even when we have acted appropriately, all things considered.

I don’t actually think that, if value resides in states of affairs and we can recognize it objectively, there is any plausible way of grounding absolutist deontology. But where to draw the line between agent-neutral and agent-relative considerations is a question for another time. How strict a theory we adopt is certainly a matter of legitimate debate, and one that warrants a great deal of attention. But, at the very least, that debate should start from a reasonable premise: a theory of the right that reflects the type of creatures we are.

5.5. Intuitionism, Kantianism and Neo-Thomism Revisited

It may be useful at this point to revisit the other deontologies that we canvassed in earlier chapters to take stock of where we are. In addition to conducting an independent assessment of Intuitionism, Kantianism and Neo-Thomism, I categorized each of those views by their acceptance or rejection of three key claims accepted by the consequentialist: all of our duties are defined in terms of value; value resides in states of
affairs, and our duty with respect to value is always simply to promote it. All deontological approaches, including the one developed here, must in some way reject the third of these claims. But recall that Ross, Kant and Finnis all take on different configurations of the first two.

Ross accepts that value sits in states of affairs but rejects the idea that all of our duties are defined in terms of value. My biggest objection to Ross’ approach with respect to the triad of consequentialist’s assertions is that it is peculiar to accept duties that are not defined by value when you simultaneously accept both that value sits in states of affairs and that it is sometimes dispositive in deciding what you ought to do. As I noted in Chapter 1, if the rock-bottom explanation for why we ought to do x rather than y is that is that x brings about a state of affairs with more value in it, that is not so bad. But if it is just that we ought to do x even though y would be better, then we seem not to have reached the end of our line of inquiry. Unfortunately, there is simply not much more to be said when we reject the most prominent explanation of where our duties come from.

Whatever else we say about the first consequentialist claim, accepting it allows us to avert this worry. The view I have developed here is compatible with the proposition that all duties are value-determined. Sometimes the calculus that determines our duty won’t be simple, for we have to discount the agent-neutral weight of certain states of affairs by accounting for the relevant agent-relative concerns. But everything can be done in terms of value, and that is both simpler and more elegant than what Ross proposes.

Moreover, I think it accurately captures the way we think about moral decision-making. Many of our choices seem to come down to weighing out the value in alternatives. The numbers are rarely clear, and having two scales of value muddies the water, but that
is what makes so many decisions difficult. A lot of choices seem quite explicitly to involve consideration of the two points of view. Should I donate money to help with relief efforts in Haiti? Or should I keep the money so I can go out to eat? The former is quite clearly superior from the agent-neutral point of view, but the latter is disproportionately appealing. Moreover, in cases where the array of options is limited and the outcomes fairly settled, we often aim to minimize damage or maximize gains (all things considered).

There is rather less to say about the comparison of my approach with Kant’s. Kant accepts the notion that our duties are defined in terms of value but he rejects the claim that value sits in states of affairs. In Chapter 2 I suggested that failure to accept that value sits in states of affairs makes it impossible, among other things, to take full account of the intuitively obvious proposition that how much harm an act causes determines to no small degree how wrong it is. The new approach quite plainly locates value in states of affairs and accordingly has no trouble with the counterintuitive implications Kant’s stance.

Aside from the worries we avoid, there is a further benefit to accepting the first two of the consequentialist’s claims. If value determines our duties and all value can be found in states of affairs, the process for discerning duties can follow a predictable (if complicated) formula (contra Ross), and we can plug consequences into that formula (contra Kant). The two propositions function well in tandem, at least with respect to preserving common moral intuitions. Yet accepting both of these claims does not force us to take account of consequences, as we saw with Neo-Thomism. From the consequentialist standpoint, the view I have advocated and the view defended by Finnis are quite similar. Both views accept both claims.

The difference comes down to the way in which we reject the third claim: that our
duty with respect to value is always to promote it. Finnis posits an attitude that we must take toward the good. That attitude resembles respect, and it involves abstaining from committing acts that harm instantiations of the goods, whether with good intentions or bad. Finnis thus stops us from taking account of consequences (for the most part) by stipulating a normatively appropriate attitude toward the good that generally ignores consequences. Finnis also denies that the goods are commensurable, making it doubly wrong to try to maximize good; but the essential neo-Thomistic element of the view is his use of an attitude that would frequently render consequentialist decision-making morally wrong. We should not always aim to maximize the good because we ought, instead, to respect it.

On my view, the reason why we may not always maximize the good flows from the need to take account of a particular objective fact about human agency. I have followed Samuel Scheffler in referring to that fact as “the natural independence of the personal point of view,” and I have argued that taking account of that fact properly both illustrates the impossibility and clearly demonstrates the impropriety of forcing human agents to maximize the good (agent-neutrally calculated) all of the time. It is not that maximizing the good is an incoherent exercise, for we often weigh outcomes in the requisite way; and it is not that maximizing is always wrong, because sometimes it will be permissible, or even required.

It is no surprise that I personally find the latter more appealing, but that is unlikely to be a decisive consideration either way. The crucial difference between the Neo-Thomistic approach and my own (aside from how appealing the implications of each may be) is that Finnis declares his preferred attitude towards the goods by fiat. If we ask why
we ought to maintain such an attitude, he will not have a way of justifying his response except by stipulation. In an echo of the problem I raised with Ross’ view, the Neo-Thomistic view presents us with an unfortunate asymmetry. The consequentialist says we should accept the third claim because more good is better than less. That may be as deep as we can go in search of an answer, but it’s compelling. The Neo-Thomist will say, no, we ought instead to respect the goods—just because that’s what you should do with goods. The consequentialist appeal seems the more powerful of the two, and the only reason it seems not to “win” terribly often is that most people possess deeply entrenched deontological intuitions.

By contrast, I have tried to point to a natural phenomenon that is both likely to be real and plausibly of normative significance. We might have debates about whether natural independence is a fact, or more likely, whether it has normative implications; but at the very least we have a concrete anchor for our reluctance to accept the injunction to maximize. The lowest level of explanation here takes us one level deeper than Finnis can, and it seems to me that a psychological fact like natural independence (if it is a fact) is exactly the sort of explanation where we want the buck to stop. If it is anything like what I have described, natural independence quite may have the power both to explain and to justify many of our moral intuitions, at least to some degree.

In short, more good is surely better than less good; it is implausible in the extreme to deny that. If we want to push back against that consequentialist appeal, we need a legitimate explanation for why that is the wrong moral standard to apply. Unlike the various unjustified attitudes enlisted by neo-Thomists, natural independence appears to
have the resources to provide that explanation.

**Conclusion**

There is a strange disconnect between the towering presence of consequentialism in normative ethical theory and the limited role it plays in the day-to-day ethics of human interaction. There is a similar disconnect between the muddled theoretical foundations for deontology and the widespread institutionalization of deontologically-inspired policies. Fidelity to consistency seems to demand reconciling these two rather striking phenomena, or at least establishing once and for all that they cannot be reconciled. It seems that we must align our understanding of normative ethics with what we practice or do the reverse.

One essential step in this process requires sorting out the status of the agent-centered restriction, a ubiquitous and stubborn moral device that seems to defy the numerous theoretical considerations that militate against it. I have attempted to provide an explanation for why the resilience of such constraints may be more than an artifact of human moral weakness or reflective incapacity. Perhaps this effort is itself misguided, simply further evidence of the human need to rationalize bad behavior. But I don't think so. I have aimed to tell a compelling story as to why that is the case, but if nothing else, I hope to have cast some doubts on the theoretical case against agent-centered restrictions.
Works Cited


