Let's Get Real: Counterfactual Moral Theories in the Actual World

by

Daniel Frederick Webber

BA, Amherst College, 2014

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Daniel Frederick Webber

It was defended on

June 27, 2023

and approved by

Michael Thompson, Professor of Philosophy

Gideon Rosen, Stuart Professor of Philosophy, Princeton University

Committee Co-Chair: Japa Pallikkathayil, Associate Professor of Philosophy

Committee Co-Chair: Jed Lewinsohn, Assistant Professor of Philosophy

Let's Get Real: Counterfactual Moral Theories in the Actual World

Daniel Frederick Webber, PhD

University of Pittsburgh, 2023

Several prominent moral theories (such as contractualism, rule consequentialism, and Kantianism) ground morality in facts about what would happen if we accepted certain rules or principles. I show how this counterfactual approach fails to do justice to many of the ways in which right and wrong depend on facts about what *actually* happens. Our actions can be wrong, for instance, because of the risks they impose on those around us, or because they are unfair to others, but I show that contractualism and rule consequentialism have difficulty making sense of this. I also argue that these theories cannot account for the moral relevance of the social rules we have actually adopted. Finally, I argue that Kantians' explanatory reliance on the counterfactual notion of universalizability leads them to miss the significance of our actual wrongs against others.

Table of Contents

List of Figures

Preface

First and foremost, I would like to thank the members of my committee—Japa Pallikkathayil, Jed Lewinsohn, Michael Thompson, and Gideon Rosen—for all they have done to help me make this dissertation what it is today. This document may not be much, but I shudder to imagine how much worse it would be in the counterfactual world where I did not generally accept the wisdom and advice of these first-rate philosophers and mentors. Michael has helped me not to miss the forest when all I have had on my mind is trees. Gideon has been the ideal external reader: generous with his time and written comments, deeply knowledgeable and insightful, and, perhaps most importantly, a delight to do philosophy with. There are few people I have enjoyed talking contractualism or moral metaphysics with as much as Gideon, and probably fewer I have gained as much from. Without a doubt, though, the lion's share of my thanks belongs to my co-chairs. Japa has been a steady hand on the rudder of my graduate career and intellectual life since she began advising me, guiding me wisely through difficult decisions and treacherous bureaucratic waters. My faith in her philosophical judgment, sound common sense, and logistical competence has been rewarded time and again. I thank her for consistently helping me to find the right way forward, in philosophy and in life. Jed has been a friend since our mutual arrival in Pittsburgh, when he went out of his way early and often to encourage me and help me grow as a philosopher. He has also been my fiercest critic, and while it has not always been easy to be grateful for this in the moment, I have never failed to appreciate just how greatly my work has improved as a result of addressing his penetrating critiques. I thank him for never taking it easy on me, and for always believing that I could do better.

Though the most extensive feedback on these chapters has come from members of my committee, many other philosophers have helped me improve them by generously offering to provide written comments or meet with me to discuss drafts. Though I regret that I am surely forgetting to mention many excellent interlocutors, I am grateful to Samuel Scheffler, Selim Berker, Kyla Ebels Duggan, James Shaw, and Nandi Theunissen for their invaluable feedback on earlier versions of the material presented here. My fellow graduate students at Pitt have also enormously improved this work, whether through formal comments or informal conversations. In particular, I would like to thank participants in Pitt's graduate work-in-progress series and Pitt's dissertation seminar, as well as Stephen Mackereth, Pablo Zendejas Medina, Jack Samuel, Aaron Segal, and especially Aaron Salomon, who has been my trusted consigliere in all matters philosophical and practical since our early days together at Pitt, and whose excellent advice has never led me astray.

I would also like to thank Nishi Shah, without whose encouragement I would never even have considered a career in philosophy. But the greatest thanks in this regard are due to my parents, Chuck and Denise Webber, who raised me to be intellectually curious and to believe that I could follow my passions. I could not have embarked on this journey had I not known that their love and support would always be available to me, in success or in failure.

Julia: I will never know how to thank you for your unwavering faith in me, nor for the way you have cared for me, day in and day out, as I have toiled at this mysterious work of dubious value. The greatest joy of my life has been sharing it with you. Thanks for continuing to share yours with me.

1.0 Introduction

What makes some acts right and others wrong? This is one of the most fundamental questions in philosophy, and also the subject of this dissertation. Sadly, I have not (yet) conclusively answered this age-old question once and for all. What I hope to have done here, though, is to show that we cannot accept the answers proffered by some of today's most prominent moral theories, including contractualism, rule consequentialism, and certain forms of Kantianism. The problem is that these theories all attempt to answer our question by appealing to facts about what *would* happen if certain rules or principles were generally accepted. I argue that this focus on counterfactual situations blinds these theories to the many ways in which right and wrong depend on facts about what *actually* happens, including facts about our actual social practices and the real-world significance of our actions for those actually affected by them.

Before we turn to those arguments, though, I should introduce the *dramatis personae*. First to enter are the contractualist and the rule consequentialist. The rule consequentialist holds that wrong acts are wrong because they are forbidden by the rules whose general acceptance would have the best aggregate consequences. Her theory exhibits what is often called a *two-level* structure. At the first level, we determine what the moral rules are by imagining the consequences that would be likely to eventuate if different (sets of) moral rules were generally accepted in the population. We tally up the goodness and badness of those consequences, and the rules whose general acceptance would result in the greatest net total of good over bad are the genuine rules of morality. Only then do we move to the level of particular acts, which are evaluated not by their consequences but rather by their conformity to the rules selected at the first level. Corresponding to this two-level structure is a two-pronged explanation of the wrongness of particular acts. It was

wrong for Cain to kill Abel, the rule consequentialist will say, because (1) some rule (say, "Don't kill people") forbids Cain's killing Abel, and (2) that rule is one of the rules that it would be best for us to accept.

The contractualist is a similar character. She holds that wrong acts are wrong because they are forbidden by a principle that no one could reasonably reject. "Reasonably reject" is a contractualist term of art; to determine whether a principle is one that no one could reasonably reject, we imagine that the principle is generally accepted in the population, and ask what objections individuals could make to this state of affairs on the basis of how it would affect them personally. We then compare the strength of these objections to the strength of the objections that people would have to the general acceptance of alternative principles. A principle is one that "no one could reasonably reject" if the strongest individual objection to its being generally accepted is weaker than the strongest individual objection to the general acceptance of alternative principles. The number of people who can make each objection is irrelevant; a non-rejectable principle is one that is "least unacceptable to the person to whom it is most unacceptable."[1] In effect, contractualism is just rule consequentialism where the only consequences that count are the consequences for the individuals who stand to be *most* negatively affected by the general acceptance of each rule. Again we have a two-level structure: at the first level, we select moral principles by comparing the objections that could be raised to their general acceptance, and at the second level, we evaluate acts by their conformity to the principles selected at the first level. And again, we have a corresponding two-pronged form of explanation: it was wrong for Cain to kill

[1] Rahul Kumar, "Defending the Moral Moderate: Contractualism and Common Sense," *Philosophy & Public Affairs* 28, no. 4 (Autumn 1999): 294.

Abel, the contractualist will say, because (1) some principle (say, "Don't kill people") forbids Cain's killing Abel, and (2) that principle is one that no one could reasonably reject.

The third character to enter is the Kantian, who features only in the final act. The Kantian is a complex character who appears in two guises. In her first costume, the Kantian bears a striking resemblance to the contractualist and the rule consequentialist. This Kantian, drawing inspiration from Kant's first, "universal law" formulation of the categorical imperative, holds that wrong acts are wrong because their agent's *maxim*—roughly, the principle on which the agent acts—could not be willed as a universal law to be accepted and followed by everyone. There is something like a two-level structure to this view as well: at the first level, the Kantian assesses maxims by their suitability to serve as universal law, and at the second level, the Kantian assesses acts by the maxims on which they are done. One key difference is that at the first level, maxims are not assessed comparatively by how good or bad their general acceptance would be, but rather merely by whether it would be *possible* to will their general acceptance. In her second guise, the Kantian takes up a view decidedly different from the others explored here. Drawing inspiration from Kant's second, "humanity" formulation of the categorical imperative, this Kantian holds that wrong acts are wrong because they show disrespect for the unconditional value of humanity (or rational nature). Here there is no two-level structure, at least as the view is traditionally understood: acts are assessed directly by how they (fail to) show respect for the value of rational nature, rather than by appeal to rules or principles that themselves pass some kind of moral test.

If these are the protagonists in our drama, then I am its antagonist. My purpose will be to make trouble for these theorists—or, more precisely, to show how they make trouble for themselves. Sometimes (like in the first section of Chapter 2, or in Chapter 3), this trouble will be extensional: the problem will be that these theorists give an implausible account of *which* acts are

right or wrong. But other times (the rest of Chapter 2 and Chapter 4), the trouble will not be with these theorists' accounts of which acts are right or wrong but with their explanations of *why* acts are right or wrong. Increasingly, philosophers are recognizing that we expect more from a moral theory than a formula for telling right from wrong—we also expect a moral theory to shed light on the facts that *make* acts right or wrong, the facts *in virtue of which* acts are right or wrong.[2] The theories we will discuss all aspire to do this. The contractualist, for instance, does not just hold that acts are wrong *when* they are forbidden by a principle that no one could reasonably reject, but also that being forbidden by such a principle is what *makes it the case* that they are wrong.

At issue here is a non-causal explanatory relation with a long pedigree in philosophy, going back at least as far as Socrates's question whether the pious is pious because it is loved by the gods, or loved by the gods because it is pious. My own view is that this relation is just the same one that metaphysicians now call *grounding*, but nothing here hangs on this view.[3] What *is* important for our purposes is that this relation is (at least often) transitive: if A makes B the case and B makes C the case, then A makes C the case, *by* making B the case.[4] This transitivity means

---

[2] See, e.g., Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), 368–70; T. M. Scanlon, "Wrongness and Reasons: A Re-examination," in *Oxford Studies in Metaethics*, vol. 2, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2007), 6 and 16ff; and R. Jay Wallace, *The Moral Nexus* (Princeton: Princeton University Press, 2019), 35–6. Of course, attention to this explanatory dimension is not an entirely new phenomenon; cf. W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), 16ff.

[3] Those who share the view that the right- and wrong-making relations in ethics are just the grounding relation include Selim Berker, "The Unity of Grounding," *Mind* 127, no. 507 (July 2018): 729–777; and Gideon Rosen, "Metaphysical Dependence: Grounding and Reduction," in *Modality: Metaphysics, Logic, and Epistemology*, ed. Ben Hale and Aviv Hoffman (Oxford: Oxford University Press, 2010), 110–1.

[4] It was once standard to assume that grounding is transitive; see, e.g., Kit Fine, "Guide to Ground," in *Metaphysical Grounding*, ed. Fabrice Correia and Benjamin Schnieder (Cambridge: Cambridge University Press, 2012), 56. Now it has become fashionable for metaphysicians to deny the transitivity of grounding, citing Jonathan Schaffer, "Grounding, Transitivity, and Contrastivity," in *Metaphysical Grounding*, ed. Fabrice Correia and Benjamin Schnieder (Cambridge: Cambridge University Press, 2012): 122–138, but if I read Schaffer correctly, his conclusion is that his apparent counterexamples to transitivity are merely apparent, and are resolved if one adopts the contrastive treatment of grounding that he defends. At any rate, I will not rely on the assumption that our "makes the case" relation is always transitive; I will merely suppose (plausibly, I think, and out of charity to the theories under consideration) that it sometimes is.

that, for better and for worse, our theorists' high-level explanatory claims commit them to a complete picture of the moral order of explanation. When the contractualist says that wrong acts are wrong only because they are forbidden by a principle that no one could reasonably reject, she is not (necessarily) denying that they are also wrong for more familiar reasons (such as causing pain, or being the breaking of a promise, or what have you), since it may be that these features make acts wrong transitively, *by* making them acts that are forbidden by a principle that no one could reasonably reject.[5] But the other side of the coin is that if these features turn out to play *no* role in making it the case that an act is forbidden by a principle that no one could reasonably reject, they will also play no role, on the contractualist's view, in making the act wrong.

These considerations will first come to a head in Chapter 2, which poses a question for the contractualist and rule consequentialist: when we imagine the general acceptance of a rule or principle, should we imagine that those who accept it always apply it perfectly, or should we imagine a realistic degree of misapplication? I argue that either answer lands these theorists in hot water. If they ignore the possibility of misapplication, they seem to contravene their own commitment to evaluating rules at realistic levels of compliance. These theorists have long recognized that they must evaluate rules by imagining less-than-perfect acceptance of them, lest they end up endorsing rules like "Never use violence," which would have wonderful consequences if everyone complied with them but not-so-wonderful consequences in realistic worlds where some people don't. But if a realistic degree of non-compliance is their aim, it seems these theorists should factor in not only a realistic degree of non-acceptance, but also a realistic degree of misapplication, since non-compliance with a rule can result both from non-acceptance of it and from

[5] Cf. Parfit, *On What Matters*, vol. 1, 368–70, and Rosen, "Metaphysical Dependence: Grounding and Reduction," 110.

misapplication of it. Moreover, if they do not take misapplication into account, it is not clear that these theorists can endorse important moral rules that protect us from others' mistakes, such as rules prohibiting drunk driving or vigilante justice.

If, on the other hand, these theorists take misapplication into account when evaluating rules, they will often end up rejecting rules that appeal to what really matters morally but are difficult to apply (e.g., "Don't act unfairly") in favor of easier-to-apply proxies for these rules (e.g., "Don't cut in line," "Don't hop the subway turnstile," etc.). At first blush, this may seem like a feature rather than a bug. But here is where these theorists' explanatory commitments get them into trouble. For I show how these *proxy rules* would force our theorists to deny that acts can be wrong *because* they possess the morally significant feature—for example, because they are unfair. The problem is that the *actual* unfairness of a particular act—say, my cutting in line—can play no role in explaining why a given rule or principle would be best (or one that no one could reasonably reject), since the facts that *do* explain this are all facts about how things stand in counterfactual worlds where the rule (or alternatives to it) are generally accepted, rather than facts about the actual properties of actual acts occurring in the real world. If the fact that my line-cutting is unfair is to feature in these theorists' explanations of why my line-cutting is wrong, then, it will have to do so at the second level instead, by explaining why the relevant rule or principle forbids my act. This would be possible if the relevant rule were "Don't act unfairly," since my act runs afoul of this rule precisely because it is unfair, but it is not possible if the relevant rule is simply "Don't cut in line," since this rule forbids my act quite independently of its fairness or unfairness. In short, if contractualists and rule consequentialists take misapplication into account, their counterfactual account of right and wrong will lead them to deny that it matters whether we *actually* act unfairly (or realize other morally significant features) in the real world.

In Chapter 3, we will examine how contractualists' and rule consequentialists' focus on counterfactual worlds blinds them to the moral significance of our actual social practices. Consider, for example, our practice in the U.S. of driving on the right. This practice secures an important public good of social coordination, leaving us all better off than if everyone decided for herself which part of the road to drive on. But there's good reason to believe that it's not the *best* practice available: studies have suggested that the practice of driving on the left is somewhat safer, in part because most people are right-handed and right-eye-dominant. Even if that's true, though, it doesn't seem to make a difference to what I'm required to do here and now, where our actual practice is to drive on the right. It would be crazy to think that I am required to drive on the left just because that would be the best practice for us all to adopt. But it seems that the contractualist and rule consequentialist must think exactly this. After all, they hold that we must follow the rules or principles whose general acceptance would be best or least objectionable. And if we generally accepted a rule requiring us to drive on the left, there would be better consequences and less to object to (fewer traffic accidents and fatalities) than if we generally accepted a rule requiring us to drive on the right. So it seems that, by these theorists' lights, I am required to drive on the left here and now—to follow the ideal convention in defiance of our actual convention.

Now, obviously the full picture is a good deal more complicated than this. For one thing, "Drive on the left" and "Drive on the right" are not the only rules to choose from—we could instead accept (as most of us in fact do) more nuanced rules that defer to our actual practices, or at least make exceptions for dangerous situations. We cannot declare "Drive on the left" the best or least objectionable rule until we have evaluated such alternatives. For another thing, we need to consider rules at other levels of abstraction or generality. If, for example, "Follow local customs, provided they're sufficiently good" is one of the best or least objectionable rules, then perhaps our

theorists *can* affirm that I should drive on the right where that is the local practice. As I will argue, though, there is neither a better alternative nor a rule at another level of generality that allows these theorists to fully avoid the trouble. The fundamental problem facing all these workarounds is that a rule instructing us to follow our actual practice could never be better for us all to accept than the rules of the best practice itself. The contractualist and rule consequentialist cannot appreciate the unique benefits of following the rules that others are *actually following*, because they only ever evaluate rules by imagining that others are following them.

Finally, in Chapter 4, we will consider how similar issues arise for the Kantian who grounds morality in facts about which maxims could be willed as universal law. Because the Kantian's evaluation of maxims is not comparative—she does not rely on claims about some maxims being *better* than others—she avoids many of the specific problems raised for contractualism and rule consequentialism in Chapters 2 and 3. There is no worry that the Kantian will be forced to reject important maxims that turn out to be second-best to "proxy maxims" or the maxims of ideal practices (whatever that could mean). But the counterfactual nature of the Kantian's view still keeps her from doing justice to the moral significance of our actual relations to others. In particular, I aim to show that this sort of Kantianism cannot vindicate the role that relational phenomena like wronging and rights play in making acts right or wrong. An act can be wrong, I argue, *because* it wrongs a particular person, but our Kantian must deny this, since facts about who the agent wrongs in the actual world are irrelevant to how things stand in the counterfactual world where the agent's maxim is universal law. For example, the Kantian will say that it's wrong for me to make a lying promise to get a loan because we couldn't all act on a maxim of doing that—if we did, lenders would simply stop lending money on the basis of promises, so it would become impossible to get a loan by making a lying promise. Missing from this explanatory story, though, is the promisee

whom I have actually wronged by actually making the lying promise. One might have thought that my act should be wrong because of its moral significance *for her*, but this is something that the Kantian's counterfactual approach cannot capture. The solution for the Kantian, I argue, is to abandon the counterfactual approach, and instead ground morality directly in respect for the value of humanity. I show how this move would open up two different routes by which the Kantian might make sense of relational morality, and examine the surprising implications of each for first-order issues like emergency rescue, free-riding, and self-wronging. I also suggest how the explanatory role of wronging poses a challenge for moral theories of *all* kinds, one that extends beyond Kantianism and the other counterfactual theories that are our focus here.

2.0 The Misapplication Dilemma

One of the most salient features of public life over the past three years has been the ever-evolving web of rules necessitated by the Covid-19 pandemic. As public health conditions, technology, and our understanding of the virus have changed, so too have manifold regulations and guidelines regarding masks, travel, quarantine, vaccines, and more. The point of all these rules is to prevent us from exposing one another to an unjustifiably high risk of deadly disease. So why so many of them? Why not just one rule: "Don't expose others to an unjustifiably high risk of deadly disease"?

The answer, of course, is that public policy must be sensitive to the reality of human fallibility. This rule might suffice if we all followed it unerringly, but it is inadequate given our actual cognitive limitations, since we would so often misapply it. Absent more specific guidance, most of us would have no idea how to determine for ourselves whether the Covid-19 risk of some everyday activity is justifiable or inordinate. By contrast, it is relatively easy to tell what a rule like "Wear a mask in indoor public spaces" requires in most cases. These realities matter to policymakers, since they care about how successful we will be at applying their rules in practice.

Do they also matter to moral theorists? On views like contractualism and rule consequentialism, moral theory is akin to policymaking: the theorist's aim is to design rules whose adoption by the general public would be in some sense ideal or unobjectionable. As we will see, however, proponents of these views have not given adequate treatment to the question of whether candidate moral rules should be evaluated in light of our propensity to misapply them.

Here, I will argue that the misapplication of rules poses a dilemma for theories like contractualism and rule consequentialism. On the one hand (and as I will argue in Section 1), it is

unrealistic to suppose that the consequences of a rule's general acceptance would not include the consequences of its misapplication, and if these theories ignore these consequences, they will be unable to endorse rules that exist to protect us from others' mistakes (e.g., rules prohibiting drunk driving or vigilante justice). On the other hand, if these theories take the consequences of misapplication into account, they will often endorse rules that forbid acts not for possessing the features that really matter morally (e.g., exposing others to an unjustifiably high risk of deadly disease), but rather for possessing mere proxies for these features (e.g., being an instance of not wearing a mask indoors). As I will argue in Sections 2–4, this would prevent these theories from doing justice to our ordinary understanding of what makes acts right or wrong, and thus to our ordinary understanding of which acts have moral worth and who is wronged by wrong acts. Either way, these theories fail to do what their proponents expect of them, and what we expect of any moral theory. If they ignore the consequences of misapplication, they misidentify *which* acts are wrong; if they don't, they misidentify *why* they are wrong. I will close in Section 5 with some brief remarks on how this dilemma might generalize to other theories and how it relates to the ideal world problem (and to recent rule consequentialist attempts to avoid it).

2.1 Misapplication in Moral Theory

Contractualism says that an act is wrong if it is forbidden by a principle whose general acceptance no one could reasonably reject.[6] Rule consequentialism says that an act is wrong if it

[6] See, e.g., T. M. Scanlon, *What We Owe to Each Other* (Cambridge: Harvard University Press, 1998), 153.

is forbidden by a rule whose general acceptance would have the best consequences.[7] Both theories share a "two-level" structure on which acts are assessed by appeal to rules (or principles), and rules by appeal to the consequences of their general acceptance (for the rule consequentialist, these consequences are directly relevant; for the contractualist, they are relevant as grounds for rejecting principles). Our question is this: Do these (expected) consequences of general acceptance include the consequences of people's (expected) misapplications of the rule?

It might seem obvious that they should. One of the primary consequences of people's acceptance of a rule is that they generally try to follow it. And in the real world, where humans are not angels but rather fallible, cognitively limited creatures, the consequences of people's *attempts* to follow a rule include the consequences of both their successes and their failures. To ignore the failures would be to idealize away a fundamental aspect of human nature. It would be like doing moral theory on the premise that we are not mortal.

Contractualists, however, have been virtually silent on the relevance of misapplication.[8] Rule consequentialists have had more to say, although they have been somewhat equivocal. On the one hand, rule consequentialists often appeal to the consequences of misapplication to rebut the infamous "collapse objection." According to this objection, rule consequentialism "collapses" into act consequentialism because it ends up endorsing exactly one rule, "Do the act with the best

---

[7] See, e.g., Brad Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000), 1–2 and 32. Some rule consequentialists evaluate rules by the consequences of general *compliance* with them rather than general *acceptance* of them, but this view has become relatively unpopular, and I will not consider it here.

[8] There has been some discussion of how contractualism handles non-compliance on the part of those who do not accept the non-rejectable principles (see, e.g., Elizabeth Ashford, "The Demandingness of Scanlon's Contractualism," *Ethics* 113, no. 2 (January 2003): 273–302; and Jussi Suikkanen, *Contractualism* (Cambridge: Cambridge University Press, 2020), 36–7), but hardly any on how contractualism handles non-compliance due to error on the part of those who *do* accept the non-rejectable principles. Scanlon does suggest that we should evaluate principles in light of the fact that "finer-grained principles will create more uncertainty and require [people] to gather more information in order to know what a principle gives to and requires of them" (Scanlon, *What We Owe to Each Other*, 205). But it is not clear that this means principles can be reasonably rejected on the grounds that people are likely to misapply them, rather than merely on the grounds that determining how to apply them correctly would take too much time and effort.

consequences available," since this rule would have the best consequences if generally accepted. As rule consequentialists have pointed out, though, there are several good reasons to think that general acceptance of this rule would not actually make things go best. And perhaps the most cited reason is that this rule is devilishly hard to apply correctly.[9] If we tried to comply with it, we would almost always fail, since we rarely have complete information about the likely consequences of the acts available to us and often misestimate the goodness or badness of these consequences. We would do many suboptimal things, sometimes even horrific things, under the mistaken impression that we were complying with the rule. Things would go better, rule consequentialists say, if we instead accepted familiar moral rules like "Don't hurt people," "Don't lie," "Keep your promises," etc., since we are much more likely to apply such rules correctly. So rule consequentialism will endorse rules such as these rather than "Do the act with the best consequences available."

On the one hand, then, rule consequentialism remains a genuine alternative to act consequentialism in part because it is sensitive to the consequences of misapplication. On the other hand, rule consequentialists have rarely appealed to such consequences outside the context of the collapse objection,[10] and some have even explicitly denied their relevance. Here's Brad Hooker:

> [T]he acceptance of a rule… can have consequences over and above compliance with the rule. One way this can happen is that a given level of internalization of the rules does not result in that level of compliance with them. Why might internalization of rules fail to produce perfect compliance? People make mistakes, give in to temptation, and so on. … I myself believe that rule-consequentialism should take into account many differences between internalization and compliance… but not this one. It seems to me counterintuitive that what is morally right depends on rules designed on the assumption that we will regularly fail to comply with them. If the point of setting a rule one place rather than another is that

[9] See, e.g., Richard Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), 273; Shelly Kagan, *Normative Ethics* (Boulder: Westview Press, 1998), 225–30; Hooker, *Ideal Code, Real World*, 93–4 and 142–3; Michael Ridge, "Introducing Variable-Rate Rule-Utilitarianism," *Philosophical Quarterly* 56, no. 223 (April 2006): 242–3; and Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), 404.

[10] Brandt argues for more concrete rules on the grounds that the application of abstract rules to particular cases "may be too complex for the average person" (*A Theory of the Good and the Right*, 290), but it is not clear whether he is worried about misapplication or simply about the time and effort needed to apply rules correctly. Cf. note 8 above.

> our actions will miss their target to some degree, then a human tendency to make mistakes is shifting the line between the morally allowed and the morally forbidden.[11]

It's not clear how Hooker thinks this position is consistent with the above response to the collapse objection, which he also endorses.[12] Nevertheless, maybe Hooker is right. Wouldn't it be strange for the moral rules to be what they are only because we can't be trusted to follow better ones?

I think not. On the contrary, many sensible moral rules make sense *only* "on the assumption that we will regularly fail to comply with them." A favorite example of rule consequentialists' (Hooker included) illustrates this point perfectly.[13] There would be no reason to accept the rule "Only use violence defensively" (rather than "Never use violence") if we could count on everyone's compliance with it, since universal compliance with this rule would (happily) rob us of occasions for defensive violence. If this is the rule we should accept, it is only because people will all too often fail to comply with it. Contra Hooker, this does not seem problematic.

Rule consequentialists and contractualists alike use this example to illustrate the importance of distinguishing *general* acceptance from *universal* acceptance. When we imagine general acceptance of a rule or principle, they say, we should build in a realistic degree of non-acceptance, and imagine that *most* but not *all* people accept the rule.[14] This move allows these theorists to endorse "Only use violence defensively" over "Never use violence." But it seems to me to fall just short of the heart of the matter. For in the first instance, the problem with "Never use violence" is that not everyone will *comply* with it. To be sure, those who fail to comply with this particular rule will do so because they fail to *accept* it, but the non-acceptance is problematic

---

[11] Hooker, *Ideal Code, Real World*, 76–7; but cf. 81–2, where he appears to take the opposite position.
[12] See ibid., 93–4 and 142–3.
[13] See, e.g., ibid., 80–3, and Parfit, *On What Matters*, vol 1., 312–7.
[14] See, e.g., Hooker, *Ideal Code, Real World*, 80–3, and Suikkanen, *Contractualism*, 36–7.

primarily because of the non-compliance that results from it. So while this particular case can be handled merely by allowing for a realistic degree of non-*acceptance*, it seems to reveal a deeper problem that calls for a stronger solution—namely, allowing for a realistic degree of non-*compliance*, which (as we've seen) can result not only from failure to *accept* the rules but also from failure to *apply* them correctly.

Indeed, we can easily think of analogous cases in which these theories would render the wrong verdicts if they did not assume a realistic degree of non-compliance due to misapplication. Consider two rules we could choose from: "Don't drive drunk," and "Don't drive drunk, unless it will harm no one." How would general acceptance of the latter rule differ from general acceptance of the former? If we assume that everyone who accepts each rule applies it perfectly (even if not everyone accepts it), the latter rule seems to alleviate some of the burdens of the former without adding any of its own: people would only drive drunk when it harms no one, and these harmless drunk drivers would be spared the costs of refraining from drunk driving (the effort and expense of arranging alternate transportation, forgone enjoyment of drinks declined, etc.). So it seems that contractualists and rule consequentialists would have to say that the latter rule is better or less objectionable, and thus that drunk driving is permissible when it harms no one. Intuitively, though, drunk driving is wrong even when the driver manages to harm no one. This is a result that contractualists and rule consequentialists can deliver only if they take into account the intuitively relevant fact that the latter rule would all too often be misapplied by those who accept it. Or take another case: "Punish culpable wrongdoers" would be an unobjectionable rule if all who accepted it applied it correctly (again, even if not everyone accepted it), but it would license vigilante justice that in reality often mistakenly targets innocents. If contractualists and rule consequentialists ignore the consequences of misapplication, they will simply idealize away the circumstance of real

life (namely, human fallibility) that makes "Only punish those who are found guilty in a fair trial" the preferable rule. And if they cannot endorse this latter rule or something like it, these theorists will be left without the resources to say something we want to say about vigilante justice—namely, that it is wrong even when it *does* target culpable wrongdoers.

All told, then, it seems that contractualists and rule consequentialists should factor in the consequences of misapplication when evaluating rules. Ignoring these consequences would not only put these theorists at odds with their own aim of taking a realistic view of general acceptance, but also blind them to the wrongness of conduct whose wrongness depends on our fallibility.

2.2 The Wrong World Problem

I have just argued that contractualism and rule consequentialism must take misapplication into account to avoid counterintuitive results about which acts are wrong. But we have strong pretheoretical views not only about which acts are wrong, but also about what *makes* these acts wrong. As T. M. Scanlon puts it, "we rarely, if ever, 'see' that an action is wrong without having some idea *why* it is wrong."[15] Common sense tells us that it is wrong for you to stomp on my foot because it causes me pain, that it is wrong to cut in line because it is unfair, etc. I will call claims like these *ordinary wrong-making claims*, since they express our ordinary understanding of what makes acts wrong; accordingly, I will call the wrong-making facts that figure in them (e.g., the fact that you caused me pain) *ordinary wrong-makers*.[16]

---

[15] Scanlon, *What We Owe to Each Other*, 198. Emphasis in original.

[16] The ordinary wrong-making claims I will discuss here concern act *tokens* (although we do make similar claims about act *types*). I also take it that our ordinary wrong-making claims do not purport to tell *the whole story* as to why

Just as we would be rightly skeptical of a moral theory that ran roughshod over our ordinary understanding of which acts are wrong, so too we would have reason to doubt a moral theory that contradicted too many of our ordinary wrong-making claims. Again, Scanlon puts the point well:

> If I could easily prevent someone standing nearby from being injured, then I should do so. It would be wrong to just stand there and do nothing. … Any sensible moral view will tell me not to just stand there, but to offer help. And any such view will say that this is so *because* of the injury that would otherwise result….[17]

Pamela Hieronymi takes a similar position:

> It would be wrong, I assume, for you to stomp on my foot for fun. Why is it wrong? One wants to say, with the utilitarian, "Because it causes me pain." And surely, whatever else we say, this must not turn out to be incorrect.[18]

It is no coincidence that these philosophers are both contractualists. For contractualists like Scanlon, Hieronymi, and Derek Parfit argue that contractualism is well-positioned to vindicate our ordinary wrong-making claims, and Parfit argues that the same is true of rule consequentialism.[19]

Over the next three sections, I will argue that these theories are not so well-positioned to do this if they take misapplication into account. Thus, the dilemma: If they ignore misapplication, these theories misidentify *which* acts are wrong; if they don't, they misidentify *why* they are wrong.

It's worth elaborating on what's at stake here. In the first instance, we want our moral theory to vindicate our ordinary wrong-making claims because we want an account of what makes acts right or wrong that strikes us as plausible. But even more than this hangs on a theory's ability

---

our acts are wrong, and thus leave open the possibility that their ordinary wrong-makers make acts wrong *indirectly* (i.e., only by making some other wrong-making fact the case) or *partially* (i.e., only in conjunction with other facts).

[17] T. M. Scanlon, "Wrongness and Reasons: A Re-Examination," in *Oxford Studies in Metaethics*, vol. 2, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2007), 7. Emphasis in original.

[18] Pamela Hieronymi, "Of Metaethics and Motivation: The Appeal of Contractualism," in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. R. Jay Wallace, Rahul Kumar, and Samuel Freeman (Oxford: Oxford University Press, 2011), 106.

[19] See T. M. Scanlon, "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), 118; Hieronymi, "Of Metaethics and Motivation," esp. 106–9; and Parfit, *On What Matters*, vol. 1, 368–70 and 413–5.

to vindicate such claims. For one thing, many philosophers hold that an act has moral worth just in case its agent is motivated by the considerations that make the act right.[20] At least often, though, what makes it right to refrain from a wrong act are the very facts that make the act wrong (e.g., it is right not to stomp on my foot for fun because it would cause me pain). If these philosophers are correct, then, a theory that contradicts our ordinary wrong-making claims will also contradict our ordinary understanding of which acts have moral worth. It may seem that you act with moral worth if you refrain from stomping on my foot because it would cause me pain, but we might have to deny this if our moral theory says that this is not what makes it wrong to stomp on my foot. Moreover, some philosophers hold that a person can only be wronged by an act if she features in the facts that make it wrong.[21] If these philosophers are right, a theory that contradicts our ordinary wrong-making claims might also contradict our ordinary understanding of who is wronged by wrong acts. You wrong me if you stomp on my foot, but our theory will render this mysterious if it denies that stomping on my foot is wrong because of anything having to do with me.

How, then, might contractualism or rule consequentialism accommodate our ordinary wrong-making claims, thereby avoiding such results? According to contractualism, wrong acts are wrong because they are forbidden by a principle that no one could reasonably reject (for short, a *non-rejectable* principle), and all other wrong-making facts make acts wrong only by making them acts that are forbidden by a non-rejectable principle.[22] According to rule consequentialism, wrong

---

[20] See, e.g., Nomy Arpaly, "Moral Worth," *Journal of Philosophy* 99, no. 5 (May 2002), 226; and Julia Markovits, "Acting for the Right Reasons," *Philosophical Review* 119, no. 2 (April 2010), 205.

[21] See, e.g., Philip Stratton-Lake, "Recalcitrant Pluralism," *Ratio* 24, no. 4 (December 2011), 374; and Richard Yetter Chappell, "The Right Wrong-Makers," *Philosophy and Phenomenological Research* 103, no. 2 (September 2021). My argument here mirrors Chappell's argument at 427–30.

[22] See Parfit, *On What Matters*, vol. 1, 369–70. Scanlon now joins Parfit in characterizing contractualism as an account of what makes acts wrong, although he holds that it describes merely one way of being wrong (see "Wrongness and Reasons: A Re-examination," 16). But we can safely ignore such pluralism here, since the acts we will consider are all ones that contractualists should take to be wrong in the way contractualism describes. And the only facts that make acts wrong in *that* way are facts about what non-rejectable principles forbid and the facts that make those the case.

acts are wrong because they are forbidden by a rule whose general acceptance would make things go best (for short, an *optimific* rule), and all other wrong-making facts make acts wrong only by making them acts that are forbidden by an optimific rule.[23] If ordinary wrong-makers are to appear on either view, then, they will have to play a role in making these higher-level wrong-making facts the case. And it seems there are only two ways they could do this, one for each of the two "levels" in these theories' structures: our ordinary wrong-makers could help make it the case that a principle/rule has the *status* of being non-rejectable/optimific, or they could help make it the case that the principle/rule in question *applies* to a certain act. We might illustrate this two-pronged order of explanation thus (where arrows point to a fact from the fact(s) that make it the case):

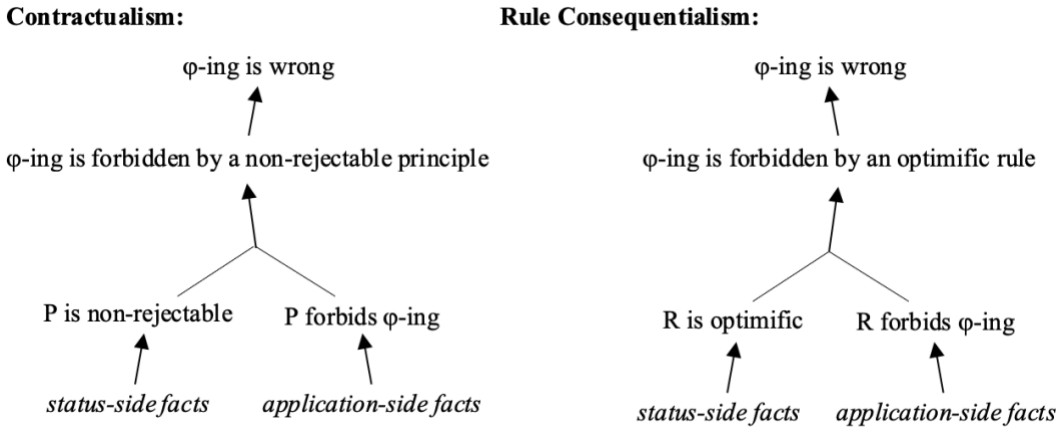

Figure 1. The explanatory structure of contractualism and rule consequentialism.

Given a theory of this shape, we will have to find the ordinary wrong-makers among either the theory's status-side facts or its application-side facts.

---

[23] Most rule consequentialists state their view as a mere biconditional for right or wrong action, rather than as a claim about what makes acts right or wrong (an exception is Derek Parfit, *On What Matters*, vol. 3 (Oxford: Oxford University Press, 2016), 432). But it is clear that they mean this biconditional to be explanatory. Hooker, for instance, states act and rule consequentialism as biconditionals in an attempt to show how the two views disagree about "what makes an act morally permissible, that is, about the criterion for moral rightness" (*Ideal Code, Real World*, 144). I think Hooker is right to identify a theory's criterion of right and wrong with its account of what makes acts right or wrong. Moreover, if the rule consequentialist biconditional is supposed to describe *the* criterion of right and wrong, I think the view can accurately be described as I describe it here: not just as a view about what makes acts wrong, but as a view about (in Parfit's phrase) the sole "*higher-level* wrong-making property or fact, under which all other such properties or facts can be subsumed" (*On What Matters*, vol. 1, 369).

It is natural to suppose that they will show up on the status side. For the contractualist, what makes a principle non-rejectable is that the strongest objection to its being generally accepted is weaker than the strongest objection to every alternative principle[24] (the number of people who can make each objection is irrelevant; a non-rejectable principle is one that is "least unacceptable to the person to whom it is most unacceptable").[25] But these objections tend to cite features that we ordinarily take to be wrong-making. The contractualist will say that it was wrong for you to stomp on my foot because doing so is forbidden by a non-rejectable principle—say, "Don't cause pain for fun." And why is this principle non-rejectable? Because the strongest objection to its general acceptance (namely, that people would miss out on the fun of causing pain) is weaker than the strongest objection to alternative principles that permit causing pain just for fun (namely, that people would cause each other a lot of pain). So the contractualist might seem to vindicate the ordinary wrong-maker in this case, since her theory affirms that facts about the causing of pain play a role in making your foot-stomping wrong (by making "Don't cause pain for fun" non-rejectable). Indeed, this is Hieronymi's strategy for accommodating ordinary wrong-makers—they "will appear as *grounds* for the rejection of principles."[26] A similar move is available to the rule consequentialist. "Don't cause pain for fun" is optimific (if it is) because a world in which this rule is generally accepted would be better than worlds in which it is not, primarily because people in worlds of the latter sort would cause their fellows more pain. So once again, facts about pain help to make your foot-stomping wrong, by making "Don't cause pain for fun" optimific.

---

[24] See Scanlon, *What We Owe to Each Other*, 195 and 205.
[25] Rahul Kumar, "Defending the Moral Moderate: Contractualism and Common Sense," *Philosophy & Public Affairs* 28, no. 4 (Autumn 1999): 294. See also ibid., 229–30.
[26] Hieronymi, "Of Metaethics and Motivation," 109; see also 106–7, which works through the example discussed here.

Here we must tread carefully, though. What we have shown is that these theories can say that it was wrong for you to stomp on my foot because of *something* about the causing of pain—namely, because of various facts about the pain that various people would cause if different rules were generally accepted. To vindicate our ordinary understanding, though, these theories must say that it was wrong for you to stomp on my foot specifically because *you* caused *me* pain in so doing. And we have not yet shown that they can say this.

Or have we? After all, won't one of the "various facts about the pain that various people would cause" be the fact that you would cause me pain if you stomped on my foot? Perhaps.[27] But we must take care to observe the distinction between two different facts: the fact that you *would* cause me pain by stomping on my foot *if* a rule permitting foot-stomping were generally accepted, and the fact that you *actually* caused me pain by *actually* stomping on my foot. The former fact concerns what would happen in an imagined world quite different from our own; the latter fact concerns what has actually happened in the real world. The former fact is the one that might feature in our theorists' story about why rules prohibiting foot-stomping are non-rejectable or optimific. But the ordinary wrong-maker we are looking for is the latter fact: it was wrong for you to stomp on my foot, we ordinarily suppose, because you (actually) caused me pain in (actually) so doing. We have not done justice to this ordinary wrong-making claim if all we can say is that it was wrong for you to stomp on my foot because you would cause me pain by stomping on my foot in a counterfactual world where foot-stomping is generally permitted.[28] More generally, when a

[27] It is actually not clear whether any facts about particular individuals appear on the contractualist's status side, since strictly speaking, objections to principles are grounded in generic reasons belonging to standpoints rather than in ordinary reasons belonging to actual people. See Scanlon, *What We Owe to Each Other*, 202–6. But contractualists often ignore this feature of their own theory and speak as if real individuals were raising the objections. For the sake of argument, I will follow their lead.

[28] Note that I am not saying that this counterfactual fact must be irrelevant; I am only saying that it is not the ordinary wrong-making fact we are seeking.

wrongful act has actually occurred, the facts that we ordinarily take to make it wrong are usually facts about the actual act that was done and the features it actually possesses. It was wrong for me to ride your bike to campus because in so doing I actually used your property without your permission; it was wrong to cut in line because by actually cutting in line I actually treated others unfairly; etc. But there is no room on either theory's status side for facts about what actually happens—there are only facts about what *would* happen if certain rules were generally accepted. On the status side, at least, these theories are simply looking at the wrong world. Call this the *Wrong World Problem*.[29]

A further problem, which would arise even if the status-side facts did concern what actually happens, is that the status side would not give ordinary wrong-makers the special explanatory significance we take them to have. If the fact that you caused me pain made it wrong for you to stomp on my foot only by helping make "Don't cause pain for fun" non-rejectable/optimific, it would play no greater role in making your act wrong than would any fact about anyone's causing pain to anyone, since all such facts help make "Don't cause pain for fun" non-rejectable/optimific. But we ordinarily suppose that *your* causing *me* pain is relevant to the wrongness of *your* stomping on *my* foot in a way that other people's pain-causing (or your causing other people pain) is not. Perhaps, as our theorists claim, the wrongness of your act really does depend on the totality of facts about pain-causing, but the fact that you caused me pain cannot be relevant *solely* as a

[29] Some ordinary wrong-making claims are themselves counterfactual (e.g., "it *would* be wrong for you to stomp on my foot because you *would* cause me pain"). But even these claims cannot be accommodated via the status side. For again we must carefully distinguish two different facts: the fact that you would cause me pain if you stomped on my foot, and the fact that you would cause me pain if you stomped on my foot *in a world where foot-stomping is generally permitted*. The former fact concerns what would happen in the counterfactual world most like our own in which you stomp on my foot; the latter fact concerns what would happen in a much more distant world. It is the former fact that is supposed to explain why it would be wrong for you to stomp on my foot, but there is no room for this fact on the status side—at best, we will find only the latter fact. Strictly speaking, then, the reason we cannot find ordinary wrong-makers on the status side is not just that it contains no facts about what *actually* happens, but that it contains no facts about what happens *in the world where the act in question occurs*.

member of this totality—it must also play some privileged role in the explanation. So even if the fact that you caused me pain *did* appear on the status side, this would not exhaust the explanatory role that we ordinarily take it to play. Call this the *Privileged Role Problem*.

Contra Hieronymi, then, we will not find our ordinary wrong-makers among the facts that make rules non-rejectable (or optimific). The remaining option is that they appear on the application side, among the facts that make it the case that a given rule forbids a certain act. What makes it the case that, e.g., "Don't cause pain" forbids your stomping on my foot? Presumably, the fact that your stomping on my foot causes me pain. The same can be said about "Don't cause pain for fun" or "Don't cause pain on Tuesdays": acts that run afoul of these rules also do so (partly) in virtue of causing pain. Generalizing, it seems a rule will forbid acts at least partly in virtue of their being *p* just in case it can be put in the form "Don't do acts that are *p**," where *p** is either *p* or a property that acts have in virtue of being *p* (e.g., the property of being *p* and/or *q*). For example, "Don't (do acts that) treat people unfairly" will forbid acts in virtue of their being unfair, "Don't (do acts that) use others' property without permission" will forbid acts partly in virtue of their being uses of others' property, etc. The fact that φ-ing is *p* will be a wrong-maker on the contractualist's application side, then, just in case a rule of the form "Don't do acts that are *p**" is non-rejectable, since then the fact that φ-ing is *p* will make it the case that φ-ing is forbidden by a non-rejectable principle. And similarly for rule consequentialism and optimific rules:

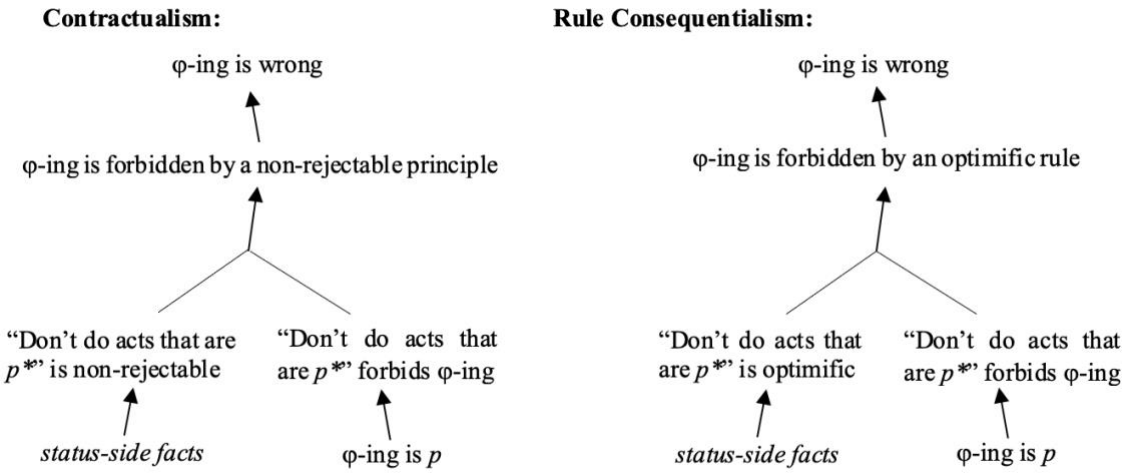

Figure 2. The structure of contractualism and rule consequentialism, elaborated.

Thus, contractualists can accommodate an ordinary wrong-making claim of the form "φ-ing is wrong because φ-ing is *p*" just in case a rule of the form "Don't do acts that are *p**" is non-rejectable, and rule consequentialists can accommodate it just in case a rule of the form "Don't do acts that are *p**" is optimific.

Plausibly, many ordinary wrong-making claims meet these conditions. Some rule of the form "Don't do acts that cause pain (unless…)" is likely non-rejectable or optimific; if so, our theorists can vindicate our ordinary wrong-making claim that it is wrong for you to stomp on my foot because you cause me pain. As I will argue in the next section, though, this will not always work if our theorists take misapplication into account: there will be ordinary wrong-making claims of the form "φ-ing is wrong because φ-ing is *p*" where no rule of the form "Don't do acts that are *p**" is non-rejectable (or optimific). Since these ordinary wrong-makers would appear on neither the status side nor the application side, our theorists could not accommodate them.

2.3 Proxies for What Really Matters

Consider the following case:

> *Barhopping*: It is September 2021, and the Delta variant of Covid-19 is spreading like wildfire in the United States. Although safe and effective vaccines are widely available to American adults, Joe has declined to be vaccinated, reasoning that he personally is unlikely to become seriously ill since he is young and in good shape. Last night, Joe went barhopping with his friends, spending several hours maskless in poorly ventilated rooms packed with strangers.

It was wrong, I think, for Joe to go barhopping. Supposing I am right about this, *why* was it wrong? The answer, at least in rough form, seems obvious: it was wrong because Joe exposed others to an unjustifiably high risk of deadly disease. Indeed, I suspect that even those who doubt whether Joe acted wrongly will agree that *if* he did, it's because of the risk he imposed on others.

Can rule consequentialism accommodate this ordinary wrong-making claim? The rule consequentialist will say that Joe acted wrongly because his barhopping was forbidden by an optimific rule. Which optimific rule, exactly? Here's one possible answer: "Don't expose others to an unjustifiably high risk of deadly disease." If this is the optimific rule that forbids Joe's barhopping, then the rule consequentialist can vindicate the claim that it was wrong for Joe to go barhopping because he exposed others to an unjustifiably high risk of deadly disease, since this fact will make it the case that Joe's behavior is forbidden by an optimific rule.

But *is* this rule optimific? Not if the rule consequentialist factors in the consequences of misapplication. The rule would have wonderful consequences if it were generally *complied with*: if most people did what this rule requires, there would be relatively little risk of people getting deadly diseases, and disruption only of those everyday activities that are unjustifiably risky. But it would have less wonderful consequences if it were generally *accepted*—that is, if people generally reasoned in accordance with it when deciding what to do—since even people making a good-faith

effort to follow this rule would be prone to misapply it. Real life can serve as an example here, since most of us, I think, actually accept something like this rule, although our attempts to apply it are all over the map. Many people earnestly believe that this rule requires almost nothing of us during the Covid-19 pandemic;[30] others believe it requires us never to leave the house; even those in the middle disagree often about what it requires. Whatever the truth is, it is clear that we have all gotten it wrong at least sometimes, and some of us have gotten it disastrously wrong. Those who underestimate the threat posed by Covid-19 have likely spread much avoidable disease and death; many who overestimate it have lost out on (and deprived others of) the many great goods of social interaction. These *misapplication costs* are among the consequences we should expect if "Don't expose others to an unjustifiably high risk of deadly disease" were generally accepted.

Contrast a relatively small set of less open-ended rules, such as "Stay at home if you've tested positive for Covid-19," "Don't go to indoor restaurants or bars unvaccinated," etc. Perfect compliance with these *proxy rules* would not be quite as good as perfect compliance with the *base rule* they're designed to track ("Don't expose others to an unjustifiably high risk of deadly disease"), since the tracking will be imperfect—even the best possible set of proxy rules will likely permit a few overly risky acts and prohibit a few sufficiently safe ones. But if our set of proxy rules is well constructed, so that these *tracking errors* are minimal, general acceptance of it will be better than general acceptance of the base rule. This is because the best set of proxy rules would be harder to misapply than the base rule: each individual proxy rule would be easier to apply (it is easy to tell what, say, "Stay at home if you've tested positive for Covid-19" requires in most cases), and it is plausible that the best possible set of such rules would not be unmanageably long or hard

[30] Most of those who take a lax attitude toward Covid-19 still accept this rule, I think. No one thinks that we *should* expose others to an unjustifiably high risk of deadly disease. The disagreement is over what counts as doing this.

to remember. In short, the tracking error and misapplication costs of the best set of proxy rules would be minor compared to the hefty misapplication costs of the base rule.

But if there is a set of proxy rules that would be better than "Don't expose others to an unjustifiably high risk of deadly disease," then the latter rule isn't optimific after all, and so can't be the optimific rule that forbids Joe's barhopping. The optimific rule that forbids Joe's barhopping will instead be one of the proxy rules—something like "Don't go to indoor restaurants or bars unvaccinated."[31] But then the rule consequentialist's explanatory story will look like this:

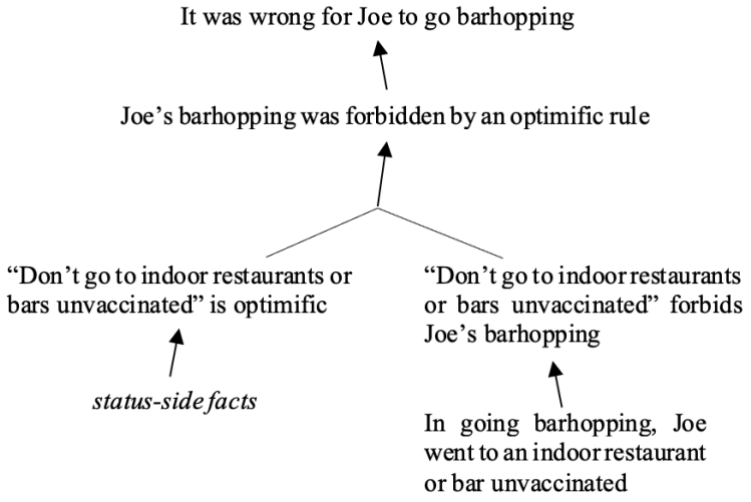

Figure 3. The rule consequentialist's account of why Joe's barhopping was wrong.

[31] In short, I am arguing that the rule consequentialist must endorse the proxy rules over the base rule because it would be better for us to apply the proxy rules when deliberating about what to do. This might raise worries that I am ignoring the time-honored distinction between a theory's *criterion* of right and wrong (i.e., its account of what makes acts right or wrong) and the *decision procedures* (i.e., rules to apply in deliberation) it recommends. It is well-known that these two things can come apart; see, e.g., Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13, no. 2 (Spring 1984): 134–71.

I am not, however, ignoring this important distinction. I have been clear from the start that I am discussing the rule consequentialist's account of what makes acts right or wrong, that is, its criterion of right and wrong. But rule consequentialism, as I understand it, builds its criterion of right and wrong *out of* the decision procedures that it would be best for us all to adopt: it says that what *makes* acts right or wrong is their (non)conformity to the rules that it would be best for us all to *accept and (try to) follow*. So it is quite proper to assess the base and proxy rules as decision procedures, even though it is the rule consequentialist's criterion of right and wrong that is at issue.

Note that this does not mean that rule consequentialism necessarily collapses the distinction between criterion and decision procedure. For it may be that rule consequentialism recommends we adopt decision procedures in the actual world that differ from those that would be best in the world where we *all* adopt them.

There is no room in this picture for the fact that Joe exposed others to an unjustifiably high risk of deadly disease. This fact does not feature among the application-side facts that explain why "Don't go to indoor restaurants or bars unvaccinated" forbids Joe's barhopping, since it neither is nor explains the fact that Joe's barhopping was an instance of going to an indoor restaurant or bar unvaccinated. Nor does it feature among the status-side facts that make "Don't go to indoor restaurants or bars unvaccinated" optimific. As we saw in our discussion of the Wrong World Problem, these status-side facts *do* include facts about the risks unvaccinated people *would* pose to others if rules permitting unvaccinated bar- and restaurant-going were generally accepted, but they *don't* include the fact that Joe *actually* exposed others to risk of disease. So the rule consequentialist must deny, against common sense, that Joe acted wrongly because he actually imposed risk on others.

A similar argument applies to contractualism. Which is non-rejectable: the base rule "Don't expose others to an unjustifiably high risk of deadly disease," or the best set of proxy rules for it?[32] To answer this question, the contractualist compares the strongest individual objections to each alternative. And if objections can be grounded in the consequences of misapplication, the strongest objection to the base rule is presumably that its general acceptance would put everyone at greater risk of disease and death than would general acceptance of the proxy rules. Plausibly, you are likelier to get sick and die as a result of someone misapplying "Don't expose others to an unjustifiably high risk of deadly disease" than as a result of edge-case risky behavior that falls through the cracks of well-constructed proxy rules. By contrast, it is almost difficult to come up with an objection to the proxy rules, since in addition to being less life-threatening, they are also easier to apply. Perhaps the strongest objection to them is that people would have to remember a

[32] Assuming that these are the two least objectionable alternatives.

greater number of rules. But this objection pales in comparison to the strongest objection to the base rule, especially since the *best* set of proxy rules would be one that is easy enough to remember. So the base rule can be reasonably rejected, and the proxy rules cannot.[33] The non-rejectable principle that forbids Joe's barhopping must therefore be one of the proxy rules (e.g., "Don't go to indoor restaurants or bars unvaccinated"), and not "Don't expose others to an unjustifiably high risk of deadly disease":[34]

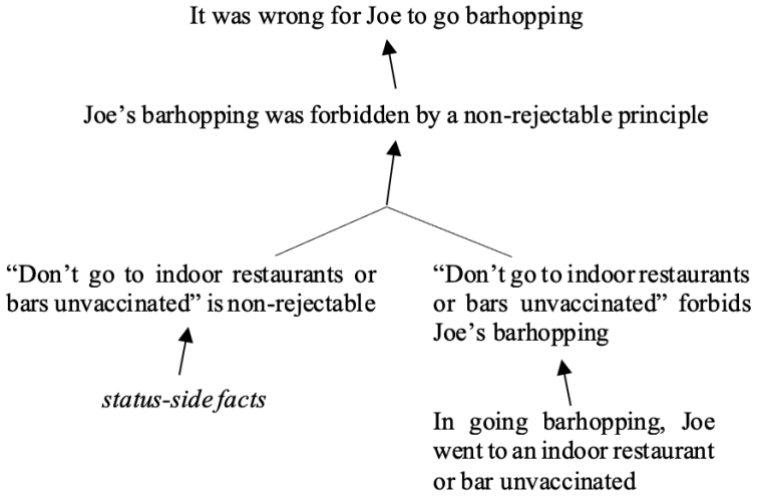

Figure 4. The contractualist's account of why Joe's barhopping was wrong.

---

[33] This argument assumes *ex ante* contractualism, on which objections appeal to the objector's *ex ante* likelihood of being affected in a certain way by general acceptance of the principle. But we would likely reach the same conclusion if we assumed *ex post* contractualism, on which objections appeal to the objector's *ex post* outcomes under general acceptance. The strongest *ex post* objections to each alternative are of equal strength: if the base rule were accepted, someone would die due to someone else's misapplication of it, and if the proxy rules were accepted, someone would die due to someone else's risky behavior falling through the cracks. The question of how to break ties in *ex post* contractualism is very much unresolved in the literature. The most popular suggestion is that in the case of a tie, the objection that can be made by a greater number of people should be considered stronger. Plausibly, more people would die due to others' misapplication of the base rule than due to overly risky behavior permitted by the best set of proxy rules. At least on this tie-breaking procedure, then, it is plausible that the strongest *ex post* objection will be to the base rule, and thus that the base rule will be reasonably rejected in favor of the proxy rules even on *ex post* contractualism.

On the distinction between *ex ante* and *ex post* contractualism, see, e.g., Sophia Reibetanz Moreau, "Contractualism and Aggregation," *Ethics* 108, no. 2 (January 1998): 296–311; Ashford, "The Demandingness of Scanlon's Contractualism"; Barbara Fried, "Can Contractualism Save Us from Aggregation?" *Journal of Ethics* 16, no. 1 (March 2012): 39–66; Rahul Kumar, "Risking and Wronging," *Philosophy & Public Affairs* 43, no. 1 (Winter 2015): 27–51; and Johann Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs* 43, no. 3 (Summer 2015): 175–223.

[34] As before with rule consequentialism, I am arguing that the contractualist must endorse the proxy rules over the base rule because the proxy rules would be less objectionable as principles to guide our deliberation. Also as before, I do not think that this argument erroneously ignores the distinction between criteria of right and wrong and decision procedures. Cf. note 31 above. I believe what I say there can also be said, *mutatis mutandis*, about contractualism.

So the fact that Joe exposed others to an unjustifiably high risk of deadly disease is also absent from the contractualist's story about why Joe's barhopping was wrong. It cannot be found on the application side, since it neither is nor explains the fact that Joe's barhopping was an instance of going to an indoor restaurant or bar unvaccinated, nor on the status side, due to the Wrong World Problem.

The reason these theories cannot vindicate the ordinary wrong-maker in Barhopping is that the relevant non-rejectable/optimific rule forbids the wrongful conduct under a description other than the one that seems to matter morally. But if these theories take misapplication into account, this will happen often, whenever rules based on the morally significant description are sufficiently difficult to apply. Consider, for example, the ordinary wrong-making claim that it's wrong for me to cut in line because it's unfair. The contractualist or rule consequentialist can vindicate this claim just in case some rule of the rough form "Don't act unfairly" is non-rejectable or optimific, but given how easy it is to misapply the notion of unfairness, it's hard to believe that any such rule could have either of these statuses once misapplication costs are factored in. Far better (or less objectionable) would be a set of proxy rules that forbids unfair acts under descriptions that are easier to apply: "Don't cut in line," "Don't hop the subway turnstile," etc. But if the moral principle that forbids my cutting in line is simply "Don't cut in line," neither contractualism nor rule consequentialism will deliver the result that it's wrong for me to cut in line because it's unfair. Or consider another case: It's wrong to serve alcohol to minors in part because they're too immature to make safe choices about drinking, but the contractualist and rule consequentialist can't say this if (as seems likely) the non-rejectable/optimific rule is something like "Don't serve anyone under the legal drinking age" rather than the easier-to-misapply "Don't serve anyone who is too

immature to make safe choices about drinking." We could go on, but the extent of the problem should already be clear.[35]

As I suggested earlier, this problem might beget two more. If morally worthy acts must be motivated by the considerations that make them right (or make acting otherwise wrong), then in denying that, say, cutting in line is wrong because it's unfair, our theorists will be forced to deny that we act with moral worth when we refrain from cutting in line because of its unfairness. And if the person wronged by an act must be singled out by the facts that make it wrong, our theorists will be forced to deny that, e.g., Joe wrongs his fellow patrons with his unvaccinated barhopping, since Joe's fellow patrons are not singled out by any of the facts in the explanatory structures just depicted—not by the application-side fact that Joe went to an indoor restaurant or bar unvaccinated, not by the status-side facts, and not by any of the facts above or below these in the order of explanation. This latter problem will not arise with every proxy rule, but it will arise with any proxy rule that forbids acts under a description that does not pick out the wronged party.

Now, our theorists might quibble about whether the balance of consequences or objections really supports the proxy rules in the cases I've described, or indeed in any case. But what is objectionable is the mere possibility that the balance of consequences or objections might support

---

[35] Could our theorists somehow avoid these results by insisting that these sorts of proxy rules are binding only because they are enshrined in law or social convention? Even if this move would help (and it's not clear how it would), our theorists are not in a position to make it. I have argued that general acceptance of these proxy rules would be better or less objectionable than general acceptance of their base rules. If this is right, then contractualism and rule consequentialism are committed to saying that these rules have moral force for us even when they are not enshrined in law or convention. Perhaps our theorists should not *want* to be committed to this, but they are. Cf. Liam Murphy, "Nonlegislative Justification," in *Principles and Persons: The Legacy of Derek Parfit*, ed. Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (Oxford: Oxford University Press, 2021): 247–276.

A different appeal to the conventional nature of these rules would be similarly unhelpful. It might be objected that the best set of proxy rules in these cases is not a laundry list of specific prescriptions, but rather a set of rules commanding us to obey the relevant authorities or conventions in our community (e.g., "Follow the guidance of local public health officials"). Cf. Scanlon's "Principle of Established Practices" (*What We Owe to Each Other*, 339). I am happy to concede that this may be the case in some of the examples I have discussed. But it is no help to our theorists if a *different* set of proxy rules is non-rejectable or optimific, since this still means that the base rule is not.

such rules. Even if, e.g., "Don't act unfairly" is in fact superior to the best set of proxy rules for it, it seems that our theorists must admit (if they take misapplication into account) that the proxy rules *would* be superior if we were bad enough at determining which acts are unfair. In that case, they would have to say, line-cutting would not be wrong because it is unfair. But it seems to me that our pretheoretical understanding is not just that line-cutting is wrong because it is unfair, but also that it would be wrong because it is unfair *even if we were terrible at telling which acts are unfair*. Moreover, our theorists would be committed to similar counterintuitive counterfactuals about virtually every ordinary wrong-maker, even those they seemed to accommodate more easily. Although they can plausibly say that it is in fact wrong to stomp on my foot because it causes me pain, for example, they will have to say that it would not be wrong for this reason if there were a proxy for pain that we were better at deliberating about. But this seems absurd. If we were terrible at telling which acts cause pain, we might less often be blameworthy for causing pain, but our wrongful acts of pain-causing would still be wrong because of the pain they caused. The problem with taking misapplication into account, then, is not just that these theories would be committed to denying so many ordinary wrong-making claims, but that they would be committed to denying so many more if convenient proxies were even easier to come by.

2.4 Why Not Both?

At this point, the contractualist or rule consequentialist might try a belt-and-suspenders approach by insisting that the non-rejectable/optimific set contains *both* the proxy rules (e.g., "Don't go to indoor restaurants or bars unvaccinated") *and* the base rules for which they are proxies (e.g., "Don't expose others to an unjustifiably high risk of deadly disease"). After all, general

acceptance of such a set might seem to be the best of both worlds. Even if people misinterpret, e.g., "Don't expose others to an unjustifiably high risk of deadly disease" as permitting unvaccinated barhopping, they are very unlikely to engage in unvaccinated barhopping if they *also* accept "Don't go to indoor restaurants or bars unvaccinated," since this rule clearly prohibits such behavior. So the misapplication costs of the base rule have been mitigated. And so has the tracking error of the proxy rules, since the base rule is around to serve as a backstop: any overly risky behavior that falls through the cracks of the proxy rules will be forbidden by the base rule. So it may seem that the base and proxy rules together are superior to the proxy rules alone. And if this is so, our two theories can accommodate the ordinary wrong-makers discussed above, since the acts in question will be forbidden by (among others) the base rule, which forbids acts under the morally significant description. Joe's barhopping will be wrong because of the risk he imposes on others, since one of the non-rejectable/optimific rules that forbids it is "Don't expose others to an unjustifiably high risk of deadly disease."

This analysis does not hold up to scrutiny, though, as it fails to factor in the misapplication costs of the base rule in cases of permissible action. Consider acts that are permissible by the lights of both the proxy rules and the base rule properly understood—say, taking a walk around the block, or having a friend over if you're both vaccinated. If the proxy rules alone were generally accepted, few would mistakenly conclude that such acts are forbidden, since the best set of proxy rules would leave little doubt as to what it required. By contrast, if we accepted both the proxy rules and the base rule, we would have to apply the base rule after determining that the proxy rules permit such acts. And here the door would be wide open for misapplication, since the proxy rules would do nothing to prevent us from mistakenly concluding that the base rule forbids something the proxy rules permit. Overly cautious people might misinterpret "Don't expose others to an unjustifiably

high risk of deadly disease" as forbidding such acts as taking a walk around the block or having a vaccinated friend over, even if they correctly concluded that the proxy rules permit such acts. The base rule would have a *chilling effect* on the behavior of such people, who would forgo many valuable (and permissible!) goods of human life by erring on the side of caution with respect to it.

Once this chilling effect is taken into account, it no longer seems that the base and proxy rules together are superior to the proxy rules alone—at least not in this case. The rule consequentialist has to weigh the badness of the base rule's chilling effect against the badness of the proxy rules' tracking error. Plausibly, though, the tracking error is the lesser evil here. The best possible set of proxy rules would deviate from the base rule in only a handful of cases, so the bad consequences of this deviation would be limited even if occasionally deadly. By contrast, the chilling effect would be widespread, and would often be deadly serious. For all the underreaction that has rightly worried us, there are probably millions who have *over*reacted to the threat of Covid-19. It is not unreasonable to suppose that thousands of them have suffered grave consequences as a result: deadly ailments left untreated for fear of going to the hospital, death from increased substance abuse in self-imposed isolation, etc.[36] Plausibly, more people would die of believing that "Don't expose others to an unjustifiably high risk of deadly disease" requires them to stay home at all times than would die of risky behavior forbidden by this rule but permitted by the best proxy rules for it. Even setting aside the less-than-deadly costs of the chilling effect, then, it seems likely that the proxy rules alone would have better consequences than the base and proxy rules together. So the rule consequentialist would be hard-pressed to defend the claim that the optimific set includes both the base and proxy rules.

---

[36] See, e.g., Sunita Puri, "My Patient Didn't Die From Covid. He Died Because of It," *The New York Times*, May 21, 2022, https://www.nytimes.com/2022/05/21/opinion/covid-deaths-million.html (accessed January 10, 2023).

For the contractualist, the question is whether the strongest objection to the proxy rules alone is stronger than the strongest objection to the base and proxy rules together. The strongest objection to the base and proxy rules together is presumably from those extremely cautious people who would deprive themselves (in some cases fatally) of the great goods of health and social life due to the base rule's chilling effect (or, if such an objection is ruled out, from the dependent children who would be unwillingly deprived of such things by their overcautious parents). The strongest objection to the proxy rules alone, on the other hand, is presumably from everyone else, who could object that the proxy rules alone would slightly increase their chances of death from disease, since the proxy rules alone would inevitably allow a few overly risky acts that the base rule would forbid. But so few overly risky acts would fall through the cracks of a well-constructed set of proxy rules that any individual's increased chance of death due to them is tiny. Plausibly, then, the objection of the extremely cautious is stronger: (the child of) an extremely cautious person is quite likely to see their physical and mental health deteriorate as a result of never leaving the house, which is surely a weightier burden than a tiny increase to an already relatively small chance of death. So it seems that it is the proxy rules alone, and not the base and proxy rules together, that are non-rejectable.[37]

Admittedly, this reasoning will not apply in every case, because not every base rule is liable to have a substantial chilling effect. Since people are not generally prone to *over*estimate their degree of drunkenness, for example, they are unlikely to think that their driving is forbidden by

[37] Again, this argument assumes *ex ante* contractualism (see note 33 above), but again, we would likely reach the same conclusion if we assumed *ex post* contractualism. The strongest *ex post* objections to each alternative are equally strong: if the proxy rules alone were accepted, someone would die due to someone else's risky behavior falling through the cracks, and if the base and proxy rules were accepted, someone would die of the chilling effect. So we have to look to the numbers to break the tie. As I argue above, though, it is plausible that more people would die of the chilling effect than of risky behavior forbidden by the base rule but permitted by the best proxy rules for it. If this is right, the stronger *ex post* objection is to the base and proxy rules together, so it is the proxy rules alone that are non-rejectable.

"Don't drive drunk" unless it actually is. So there would be hardly any chilling effect if this rule were accepted alongside its proxy rule "Don't drive if your blood alcohol content is above the legal limit," and whatever effect there was would not be grave (a few unimpaired people needlessly calling cabs, etc.). But the above argument will work in cases where erring on the side of caution is both tempting and sufficiently pernicious, and unfortunately for the theories under discussion, these are not particularly rare.

Indeed, there are even cases of this sort where the *correct* application of the base rule counts against it. For example, it is wrong to serve alcohol to minors in part because they are too immature to make safe choices about drinking. But the proxy rule "Don't serve alcohol to anyone under the legal drinking age" would be preferable not only to the base rule "Don't serve alcohol to anyone who is too immature to make safe choices about drinking," but also to both rules taken together. If both rules were generally accepted, servers would have to apply the base rule whenever they determined that the proxy rule permitted serving someone—that is, whenever they determined that a patron was of legal age. Relative to general acceptance of the proxy rule alone, this would result in some reduction of potentially dangerous service to immature but of-age patrons, since servers would sometimes assess their patrons' maturity correctly. Plausibly, though, they would more often get it wrong, refusing to sell to younger (but of-age) patrons who are in fact mature enough but whom they consider less intelligent or whose lifestyles they disapprove of. This would especially be the case with patrons whose race or gender makes them especially likely to be incorrectly regarded as immature. There would even be weighty costs when servers got it right. For one thing, servers would face high *decision costs*, since they would often have to make difficult judgments about the psychological maturity of a total stranger in order to decide whether to serve someone. Perhaps more significantly, of-age patrons would often have to submit themselves for

appraisal by a total stranger just because they wanted a drink, which could be humiliating even when the result of the appraisal is positive. It is hard to believe that these widespread costs would be outweighed by the benefits in the rule consequentialist's calculus. Nor does it seem that the contractualist could endorse the base and proxy rules together, since the objection to the two rules together from mature drinkers (namely, that they would be profiled and potentially discriminated against when they wanted a drink) seems stronger than the objection to the proxy rule alone from immature but of-age patrons (namely, that their potentially unwise drink orders would more often be fulfilled). Neither contractualists nor rule consequentialists, then, can plausibly avoid the problem simply by endorsing more rules.

2.5 Conclusion

When contractualists and rule consequentialists evaluate a rule, should they take the consequences of the rule's misapplication into account? I have shown that these theorists face a dilemma. If they ignore the consequences of misapplication, these theorists will not only end up with an unrealistic interpretation of general acceptance, but also fail to endorse important moral rules that protect us from others' errors. On the other hand, if they take misapplication costs into account, they will be forced to deny many of our ordinary wrong-making claims, and thus many of our ordinary claims about moral worth and wronging as well.

I have framed this *misapplication dilemma* as a problem for standard versions of contractualism and rule consequentialism, but I believe it generalizes beyond these views. For example, it seems to generalize to Caleb Perl's recent version of rule consequentialism, which is designed to avoid the "ideal world problem" faced by other rule consequentialist and contractualist

theories. I will have more to say about this problem in the next chapter. Briefly, though, the problem is this: since these theories evaluate rules by their consequences in imagined worlds that differ greatly from our own, they are liable to endorse rules that are excellent in the imagined worlds but terrible for our actual one.[38] Perl attempts to avoid this problem by dispensing with appeal to imagined worlds. On his version of rule consequentialism, rules are evaluated non-counterfactually by the actual goodness and badness produced by the actually occurring acts they classify as right.[39]

As stated, this view finds itself squarely on the first horn of the misapplication dilemma. "Punish wrongdoers" would be an excellent rule on this view, since it would be credited with the goodness of all actual punishment of wrongdoers, but none of the badness of actual mistaken punishment of innocents (since the rule does not classify punishing innocents as right). To avoid this horn of the dilemma, the theory could be modified to take misapplication (and acceptance more generally) into account, perhaps by evaluating rules by the actual consequences of their actual acceptance, where this includes the consequences of people's actual attempts to follow the rule. But this modified view would land on the dilemma's second horn. Although on this view, rules would be optimific in virtue of facts about what actually happens, these facts would still not single out our ordinary wrong-makers as having special explanatory significance. So we could not locate our ordinary wrong-makers on this theory's status side, since the Privileged Role Problem

---

[38] The classic example is one we encountered earlier. Although "Never use violence" would be a great rule in worlds where everyone followed it, it is less great in the real world, where innocent people often need defending. But the objection goes deeper than this. See Parfit, *On What Matters*, vol. 1, 312–20; Gideon Rosen, "Might Kantian Contractualism Be the Supreme Principle of Morality?," *Ratio* 22, no. 1 (March 2009): 78–97; and Abelard Podgorski, "Wouldn't It Be Nice? Moral Rules and Distant Worlds," *Noûs* 52, no. 2 (June 2018): 279–294.

[39] See Caleb Perl, "Solving the Ideal Worlds Problem," *Ethics* 132, no. 1 (October 2021): 89–126. Note that on Perl's view, a rule need not actually be accepted by anyone in order to "classify an act as right." "Ruin as many parties as you can" classifies all actual party-ruinings as right, and is therefore credited with all the goodness and badness produced by all actual party-ruinings, even if no one has ever accepted this rule.

would remain even if the Wrong World Problem does not. Nor could we always find them on its application side, since it seems likely that some proxy rules will outperform their base rules in terms of the actual consequences of their actual acceptance. So Perl's theory does not avoid the misapplication dilemma, even if it does avoid the ideal world problem.

I cannot offer a general formula specifying which kinds of theory are susceptible to the misapplication dilemma. The devil is always in the details. But it is worth noting two features that seem to contribute to a theory's susceptibility. The first is a two-level structure in which acts are evaluated by appeal to rules and rules by appeal to the consequences of their acceptance. It is the commitment to evaluating rules by the consequences of their acceptance that forces a theory to choose between ignoring and acknowledging the consequences of their misapplication. The second feature is an optimizing approach to rule selection that selects only the rules that perform best in the theory's evaluation. Contractualism and rule consequentialism (Perl's version included) get stuck on the second horn of the dilemma because they cannot endorse a rule if there is an alternative (like a set of proxy rules for it) that is even slightly better (or less objectionable). But a satisficing version of either view might escape the dilemma more easily, since it would be under no pressure to choose between a base rule and its proxy rules when both are sufficiently good.[40]

This points to a way in which the misapplication dilemma is more robust than the ideal world problem. The ideal world problem and the dilemma's first horn are structurally similar: both show how two-level theories get into trouble for being *unrealistic*, either in appealing to distant possible worlds or in ignoring human fallibility. But the dilemma's second horn goes beyond this—it shows how even theories that are realistic in these respects will get into trouble for refusing

[40] Jussi Suikkanen and Shelly Kagan briefly consider satisficing versions of contractualism and rule consequentialism, respectively. See Suikkanen, *Contractualism*, 39–40; and Kagan, *Normative Ethics*, 224.

to settle for rules that are second best. The problem with theories like contractualism and rule consequentialism, then, is not just that they appeal to ideal *worlds*, but, more fundamentally, that they appeal to ideal (i.e., optimal) *rules*. As we've now seen, ideal rules—even realistically conceived—cannot account for morality as we ordinarily understand it.

3.0 Ideal Code, Real Conventions

In the United States, we drive on the right; in some other countries, they drive on the left. It has sometimes been suggested that the latter practice results in fewer traffic accidents and fatalities, since (among other reasons) most people are right-handed and right-eye dominant, and so drive more safely with oncoming traffic on their right and their right hand on the wheel in a right-hand-drive vehicle.[41] Suppose that's true. That *might* give us reason to consider switching to left-hand traffic in the U.S., although the costs of doing so would almost surely be prohibitive. But what it certainly would *not* do is give me reason to start driving on the left (even in a right-hand-drive vehicle) despite our current practice. Given that we in fact drive on the right around here, it would be lunacy for me to drive on the left, even if driving on the left would be the better practice for us all to adopt.

In this chapter, I will argue that theories like contractualism and rule consequentialism commit us to at least some degree of this lunacy: if what I've said about traffic fatalities is true, these theories will in some cases require us to drive on the left, whatever our local custom may be. And of course, the rules of the road are just one example—the broader point is that these theories leave insufficient space for our actual practices to play a role in determining what we're morally required to do. The problem, in brief, is that these theories subscribe to a kind of *ideal conventionalism*, on which right and wrong are determined by the hypothetical social rules that would be best or least objectionable for us to adopt. As the case of driving on the right illustrates,

[41] See, e.g., Prashant Poddar and Vijaya Singh, "When Left Is 'Right'! The Impact of Driving-Side Practice on Road Fatalities in Africa," *Transport Policy* 114 (2021): 225–32.

though, right and wrong can depend on the social rules we have *actually* adopted even when those rules are less than perfect.

I will begin, in Section 1, by setting up the problem in greater detail. In brief (and in terms of the driving-side example that will occupy us throughout) the issue is that "Drive on the left" seems a better or less objectionable policy than "Drive on the right," which seems to force our ideal conventionalists to say that we must drive on the left here and now, regardless of any local custom to the contrary. I will then explore in turn two broad classes of strategy for addressing this problem. In Section 2, I will consider whether there might be an even better or less objectionable alternative to "Drive on the left" that allows our actual practices to exert some influence over its requirements—for instance, a rule like "Drive on the left, except when doing so would be dangerous" or "Drive on the customary side." What we will find is that "Drive on the left" can be improved upon in ways that avoid some of the problematic cases, but not all of them; rules that would be sufficiently deferential to our actual practices, I will argue, run into an underexplored version of the ideal world problem. In Section 3, I will consider whether the ideal conventionalist can account for the moral significance of our practices by appealing to other rules at her disposal—for example, rules requiring us to treat others fairly or to follow established practices in our community. This is a strategy that contractualists and rule consequentialists have actually attempted, but I will argue that, at best, this approach succeeds only in generating conflict cases where the requirements of these other rules are at odds with the requirements of the best rule that directly pertains to the activity in question (e.g., cases where "Follow good enough local practices" conflicts with "Drive on the left"). As I will suggest, this result mischaracterizes our actual moral situation even if the conflicts are always resolved in favor of our actual practices. With neither of these strategies fully successful, I conclude that contractualism and rule consequentialism cannot

in general make sense of the moral significance of our actual practices. I remark briefly in Section 4 on what this might mean for these theories' accounts of property rights, as well as on how these theories might be modified to avoid the issues I raise.

3.1 The Problem

Our social practices undoubtedly play a role in determining right and wrong. Given that our convention in the United States is to drive on the right, for example, it is right for me to drive on the right and wrong for me to drive on the left. Can contractualism and rule consequentialism vindicate these seemingly obvious moral facts?

Rule consequentialism says that an act is wrong just in case it is forbidden by one of the rules whose general acceptance would have the best consequences (for short, the *optimific* rules).[42] So the rule consequentialist's ability to vindicate common sense here turns on the content of the optimific rules that bear on driving side. Now, there may be many such rules, and some of them may not be specific to driving; indeed, most of this chapter will be devoted to canvassing the many relevant possibilities. But just to see the *prima facie* problem facing the rule consequentialist, let's start by assuming that there are only two candidates for the optimific rule that determines which side I should drive on: "Drive on the left" (DL) and "Drive on the right" (DR). Which of these rules would have the better consequences if generally accepted? As I suggested earlier, there is reason to think that traffic accidents and fatalities would be slightly reduced if we drove on the left

[42] See, e.g., Brad Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000), 1–2 and 32. There are versions of rule consequentialism that depart from this formulation in certain respects, but for the sake of concreteness I will focus primarily on this version of the view.

rather than the right. If this is so, then it seems that it would be slightly better for us all to accept DL rather than DR. So if those were the only relevant alternatives, DL would be optimific, and therefore rule consequentialism would require that I follow it, even where the local custom is to drive on the right. Accordingly, the rule consequentialist would have to say not only that it is *not* wrong for me to drive on the left in the United States, but that it *is* wrong for me to drive on the right. The challenge for the rule consequentialist is to show how a more nuanced picture of the available rules avoids these highly implausible results.

Before we consider how that challenge might be answered, though, note that a similar challenge arises for contractualism. Contractualism says that an act is wrong just in case it is forbidden by a principle that no one could reasonably reject (for short, a *non-rejectable* principle), where a principle is non-rejectable when the strongest individual objection to its being generally accepted is weaker than the strongest individual objection to every alternative principle.[43] In effect, contractualism is just rule consequentialism where the only consequences that count are the consequences for the individuals who stand to be *most* negatively affected by the general acceptance of each rule. So again, suppose for now that DL and DR are the only candidate principles. Which of these principles would be the one that no one could reasonably reject? Those who stand to be most negatively affected by the general acceptance of either principle are presumably those who might be killed in traffic accidents. And such people could object that, compared to DL, the general acceptance of DR would increase their risk of untimely demise from traffic accidents. There does not appear to be a similarly weighty objection that could be leveled

[43] See, e.g., T. M. Scanlon, *What We Owe to Each Other* (Cambridge: Harvard University Press, 1998), 153, 195, and 205.

against DL.[44] So if these principles were the only relevant alternatives, it would be DL, not DR, that is non-rejectable, and so contractualism would require that I follow DL, even where it is contrary to local custom, with all the same implausible implications as before. The contractualist must therefore show that we have glossed over candidate principles that would help her avoid these counterintuitive results.

It might be thought that there is not even a *prima facie* problem here—that DR is just superior to DL in communities that already drive on the right. For one thing, the vast majority of vehicles in such communities are left-hand-drive, but the benefits of driving on the left are chiefly benefits of driving on the left *in right-hand-drive vehicles*. American drivers would not drive more safely if they all drove on the left in their existing left-hand-drive cars. But we should not let this quirk of the example trouble us. Imagine that we all drove cars with a steering wheel and pedals on both sides, and a switch that easily toggles between the left- and right-side controls. DL would certainly be superior to DR in such a world, since our vehicles would all be set up to reap the full benefits of left-hand traffic, but even in such a world it would be a grave mistake for these theories to recommend that I adhere to DL in communities that drive on the right. So the more general problem cannot be solved simply by pointing out that we have designed our existing tools around

---

[44] Here I am assuming the view called *ex ante* contractualism, on which objections appeal to the objector's *ex ante* likelihood of being affected in a certain way by general acceptance of the principle. But it seems we would reach the same conclusion if we assumed *ex post* contractualism, on which objections appeal to the objector's *ex post* outcomes under general acceptance. The strongest *ex post* objections to each alternative seem to be of equal strength: the general acceptance of either DR or DL would presumably result in at least one fatal accident that would have been avoided had the other principle been generally accepted. The question of how to break ties in *ex post* contractualism is very much unresolved in the literature. The most popular suggestion is that in the case of a tie, the objection that can be made by a greater number of people should be considered stronger. By hypothesis, though, right-side driving results in more traffic fatalities than left-side driving. At least on this tie-breaking procedure, then, the strongest *ex post* objection is to DR, and thus DR will be reasonably rejected in favor of DL even on *ex post* contractualism.

On the distinction between *ex ante* and *ex post* contractualism, see, e.g., Sophia Reibetanz Moreau, "Contractualism and Aggregation," *Ethics* 108, no. 2 (January 1998): 296–311; Ashford, "The Demandingness of Scanlon's Contractualism"; Barbara Fried, "Can Contractualism Save Us from Aggregation?" *Journal of Ethics* 16, no. 1 (March 2012): 39–66; Rahul Kumar, "Risking and Wronging," *Philosophy & Public Affairs* 43, no. 1 (Winter 2015): 27–51; and Johann Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs* 43, no. 3 (Summer 2015): 175–223.

our existing practices. Nor can the problem be solved by appeal to the fact that our existing *competencies* have developed around our existing practices. Those in communities that drive on the right, it might be argued, would not in fact drive more safely under general acceptance of DL, since their preexisting practice has made them more skilled at driving on the right than on the left. But rule consequentialists have taken pains to clarify that we are to evaluate rules by the consequences of their internalization by each new generation, not by existing generations whose lives have already been shaped by the rules they accept. Brad Hooker, for instance, argues that this is necessary to avoid counting such things as the "costs of getting a non-racist and non-sexist code accepted by people who have already internalized racist and sexist rules."[45] Presumably contractualists also want to avoid counting costs like these. But once general acceptance is understood as general internalization by those with no previously internalized rules, the appeal to practice-dependent skill development falls flat. New generations of Americans would have no more difficulty learning to drive on the left than on the right.

This points to a broader way in which the notion of general acceptance might be misunderstood. One might think that, contrary to my earlier assertions, not much would change if DL were generally accepted as a moral rule in places like the U.S. For in the U.S., we drive on the right, which usually makes driving on the left extremely dangerous. Moreover, driving on the left is (typically) illegal, and there is no shortage of traffic cops or highway patrol. These considerations give Americans strong prudential and legal reasons to drive on the right. We should not, therefore, assume that Americans would generally drive on the left if DL were generally accepted as a moral rule. We should instead imagine that American drivers would feel profoundly conflicted, taking themselves to be morally required to drive on the left but also to have strong

[45] Hooker, *Ideal Code, Real World*, 80.

non-moral reasons to drive on the right. And, realistically, we should imagine that most drivers would resolve this conflict in favor of their own safety—even if we genuinely accepted DL, few of us would follow it when it meant risking our lives. So if DL were generally accepted, Americans and others who presently drive on the right would generally continue to do so, although some people would dangerously drive on the left, prioritizing compliance with DL over safety, and everyone else would have to live with the psychological distress that comes with regularly prioritizing one's own well-being over what one takes to be one's moral duty. General acceptance of DL would therefore not be as different (or as appealing) as I've made it out to be.

Now, part of what this reasoning brings out is the artificiality of our initial assumption that DL and DR are the only available rules. Our actual behavior is not captured by such simple rules, but is instead sensitive to important conditions like safety and what those around us are doing, and so our survey of the available rules is not complete until we consider rules that are themselves conditional on these things. We will consider such rules in the next section. Before we do, though, it is important to stress that the above is simply the wrong way of imagining a rule's general acceptance. We are not to imagine general acceptance of a rule by holding fixed all our existing norms, behaviors, and laws. Rather, we are to imagine the social world as it would have developed had we generally accepted the rule in question. If we generally accepted that we were morally required to drive on the left, we would not have developed a practice of driving on the right, nor enshrined this practice in law, and therefore would not have the reasons that our existing practices and laws give us. To object that DL would sit uneasily with our existing driving-side practices is to put the cart before the horse.

Of course, even if it turned out that DR is superior to DL, the more general problem that this example is supposed to illustrate would remain. Rule consequentialism and contractualism are

both species of a genus we might call *ideal conventionalism*, the view that actual right and wrong are determined by the hypothetical social rules that would be ideal for us to adopt in some sense of ideal (best aggregate consequences, least weighty individual objections, etc.). The general problem I am raising for ideal conventionalism arises whenever the *ideal* rule or set of rules for governing a given sort of behavior (call this ideal rule or set *A*) conflicts with a sufficiently good (but less than ideal) rule (or set of rules) for governing that behavior that is *actually accepted* in a certain community (call this rule or set *B*). When B is good enough (in a sense that I will leave open),[46] it seems obvious that those in the B-accepting community should continue to follow B, even when it conflicts with the demands of A. "Drive on the right" is a good enough rule that I should continue to follow it in my community where it is accepted, even though its demands conflict with those of the better rule "Drive on the left." The problem for the ideal conventionalist is how to make sense of this, given that their view seems to require adherence to A by virtue of its ideality. Where, on such a view, is there room for the requirements of B to get a grip?

Broadly speaking, it seems there are two strategies that the ideal conventionalist might pursue to solve this problem. Given a seeming instance of the problem—that is, an (A, B) pair that seems to fit the description above—the ideal conventionalist can either (1) argue that there is an alternative to A and B that is superior to both and whose demands are not inconsistent with those of B, or (2) argue that there is another ideal rule, one that is *not* an alternative to A and B, that requires acting in accordance with B. An example of the first strategy would be to argue that "Drive on the left" is not in fact an ideal rule, since "Drive on the customary side" is superior to it; an example of the second strategy would be to argue that the ideal rule "Follow the rules of

[46] Though I will not completely specify the relevant sense of "good enough," I should emphasize that it will almost certainly involve more than B's being sufficiently beneficial. For instance, it will probably also have to be the case that B is sufficiently just.

sufficiently good established practices" requires obedience to B quite independently of whether A or B is ideal. The two strategies are potentially complementary: the ideal conventionalist might address different putative instances of the problem by invoking different strategies. And, as we'll see, each strategy admits of many varieties, which similarly may also complement one another.

Over the next two sections, I will examine varieties of these two strategies in turn. For ease of exposition, I will continue to focus on our driving-side example, and one of the aims of my argument will be to suggest that there is no plausible version of either strategy that can treat this particular case in a satisfactory manner. There really are cases in which these theories say we must drive on the left when it's clear that we needn't. But behind this point will always be a more general one: that although some versions of these strategies will work in some cases, there are instances of the problem that are not solved by any of them. Ideal conventionalism cannot in general admit of the moral significance of practices that are merely good enough.

3.2 Better Alternatives

There are many ways in which the ideal conventionalist might pursue the first strategy of arguing that the putatively ideal rule A is not ideal after all. Indeed, it would be impossible to catalogue them all here, since in principle any alternative to A could be *argued* (although not necessarily convincingly) to be superior to A. Instead, I will focus on more general varieties of the strategy—ones that could be pursued for most choices of A—using our driving-side example and the rule DL as a guide. What we will find is that none of these versions of the strategy are completely successful: they either fail to avoid all the troublesome implications of DL, or else fail to produce a rule that outperforms DL in the ideal conventionalist's evaluation. As we shall see,

the root cause in the latter case is what is sometimes called the *ideal world problem*: the ideal conventionalist cannot take into account certain real-world considerations that tell against rules like DL, since these considerations do not arise in the counterfactual worlds in which these rules are generally accepted.[47] This will become clearer if we turn to some concrete examples.

Perhaps the most obvious alternative to DL is simply a qualified version of it. The problem with DL, one might think, is just that it paints with too broad a brush, making no exceptions for cases where driving on the left is downright dangerous. One such case is of course when driving on the right is customary, but even where driving on the left is the norm, exceptional circumstances sometimes arise—trees fall in the road, lanes are closed for construction, etc. Better or less objectionable than DL, it seems, would be a qualified rule like "Drive on the left, except when doing so would be dangerous" (DLQ). Now, one might wonder whether general acceptance of DLQ really would be better than general acceptance of DL. After all, if DL were generally accepted, it's not as if people would drive on the left *no matter what*. We would all still have strong prudential reason to avoid dangerous car accidents, and realistically, most of us would probably side with prudence when it conflicted with the moral demands of DL. The advantage of DLQ, then, may be less about traffic accidents prevented and more about the psychological unity that would be enjoyed by those who accepted it rather than DL. For those who accepted DL would see every tree in the road as a tragic conflict between the demands of morality ("Drive on the left!") and those of prudence ("Don't hit the tree!"). That would be a distressing way to live, even if one always sided with prudence. By contrast, those who accepted DLQ would see no conflict in cases like these—they could swerve to avoid the tree without regret. Even if our behavior would

[47] See, e.g., Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), 312–20; Gideon Rosen, "Might Kantian Contractualism Be the Supreme Principle of Morality?," *Ratio* 22, no. 1 (March 2009): 78–97; and Abelard Podgorski, "Wouldn't It Be Nice? Moral Rules and Distant Worlds," *Noûs* 52, no. 2 (June 2018): 279–294.

ultimately be the same under general acceptance of DL or DLQ, our inner lives as agents would be easier and more pleasant if we generally accepted DLQ. So it seems that general acceptance of DLQ would indeed be better or less objectionable than general acceptance of DL, and thus that ideal conventionalism requires us to follow DLQ rather than DL. But DLQ would not share DL's counterintuitive implications in the real world, since driving on the left in places where driving on the right is customary is manifestly dangerous. So it may seem that we have easily identified a superior alternative to DL that avoids its problems.

Here some degree of concession is in order. I think this case for the superiority of DLQ over DL is strong. And I agree that DLQ avoids most of the problematic cases for DL, since it is indeed usually dangerous to drive on the wrong side of the road. But consider the following case:

> *Desert Highway*: I am driving on a rarely-used two-lane highway through a remote area of the Nevada desert. It is a clear day with perfect visibility, and the road ahead of me is a long straightaway at a slight downgrade, so I can see for miles. I would easily spot any oncoming car ahead of me long before we would meet. But, as I can clearly see, there are no oncoming cars; I am very obviously alone on this stretch of highway.

In Desert Highway, it would be no more dangerous for me to drive on the left (at least for the short stretch of highway described) than to drive on the right. *Perhaps* that makes it *permissible* for me to switch into the left lane, although even this is doubtful. But I am certainly not *required* to switch into the left lane, though this is exactly what the ideal conventionalist must say if DLQ is ideal, since DLQ requires me to drive on the left when doing so is not dangerous. So while DLQ is a major improvement over DL, it does not entirely avoid the weirdness: it seems to require that I drive on the wrong side of the road whenever I can get away with it.

Is there a better qualified rule available? Maybe something like "Drive on the left, *provided that others are doing so*, except when doing so would be dangerous" (DLQ2). This rule directly addresses the gap we just noted in DLQ: the trouble is not just that driving on the left is sometimes

dangerous, but that it is sometimes contrary to local practice. But whether or not the ideal conventionalist can endorse DLQ2 depends on how things would stand if DLQ2 were generally accepted. And this does not seem to be a difficult state of affairs to imagine, since it seems that DLQ2 (or at least, a rule very close to it) is *actually* generally accepted worldwide. Surely most drivers, regardless of their local driving-side custom, accept that you must drive on the left *provided* that others are doing so and that it's safe to do so, just as they accept that you must drive on the *right* provided that others are doing so and that it's safe to do so. Most Americans, for example, take themselves to be required to drive on the left when in the U.K., and part of the explanation for this is, I think, that they take themselves to be required to drive on the left when this is what others are doing, although that proviso is rarely satisfied in their home country. But if our actual world is one of the worlds in which DLQ2 is generally accepted, then we already know how its general acceptance would shape our social lives. In some places, for various reasons practical or arbitrary, a convention of driving on the left would emerge, and in others, for different reasons practical or arbitrary, a convention of driving on the right would emerge. General acceptance of DLQ2 would not (did not) inhibit the development of practices of driving on the right, since the rule only requires driving on the left when others are already doing so.

By contrast, general acceptance of DLQ *would* inhibit the development of practices of driving on the right. It is hard to imagine a practice of driving on the right getting off the ground among people who take themselves to be morally required to drive on the left whenever doing so is not dangerous. In the world where DLQ is generally accepted, then, we should expect driving on the left to be the custom everywhere, whereas in the world where DLQ2 is generally accepted, we should expect diverse driving-side customs as we have now. Given that this would be the only difference we should expect, and given that driving on the left is marginally safer than driving on

the right, it seems that general acceptance of DLQ would have better consequences (fewer traffic accidents) or be less objectionable (less risk of traffic fatality for individuals in communities that drive on the right) than general acceptance of DLQ2. The ideal conventionalist thus cannot endorse DQL2 over DLQ.

To solve this problem, we might consider a rule with an exception for other practices rather than one that is conditional on a left-side practice: perhaps "Drive on the left, unless doing so would be dangerous *or contrary to local custom*" (DLQ3). Like DLQ, DLQ3 would inhibit the rise of practices of driving on the right, since those who accepted it would prefer to drive on the left even at the start, before any local convention was up and running. But this very fact leads to a different problem for DLQ3, which is that if it were generally accepted, its "contrary to local custom" clause would be otiose, since driving on the left would be the custom everywhere. DLQ3 would be more complicated than DLQ despite always requiring (and resulting in) the same behavior. This needless complication would make DLQ3 worse or more objectionable than DLQ, since simpler rules are easier to learn and apply.[48]

Both these problems are instances of the ideal world problem mentioned above. The real-world circumstance that makes DLQ2 or DLQ3 seem superior to DLQ—namely, the diversity of driving-side practices—is simply idealized away in the world where DLQ is generally accepted. In that world, every community has a practice of driving on the left, because everyone (or nearly everyone) accepts DLQ. And this ideal world problem seems to generalize to other rules that are conditional on or make exceptions for our actual practices. If rule A is superior to other alternatives in the ideal conventionalist's evaluation, it will generally also be superior to "Follow A provided

[48] On "internalization costs" of complexity, see Hooker, *Ideal Code, Real World*, 78–80. On what we might call "application costs" of complexity, see Richard Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), 290, and cf. Scanlon, *What We Owe to Each Other*, 205.

that others are" or "Follow A unless others aren't," since the world where these qualified rules are generally accepted will either be one in which every community follows A (in which case the qualifications are otiose) or else one in which some communities *don't* follow A (in which case they follow an inferior rule).

It may seem that if we're looking to top DLQ, different qualified versions of DL are the wrong rules to be looking at. Intuitively, the best rule is not one that requires everyone everywhere to drive on the left provided that (or except when) certain circumstances obtain; rather, it's one that requires everyone everywhere to drive on the side that others are driving on. There are many different rules that would do this—for example, "Drive on the side others are driving on," "Drive on the customary side," or "Drive on the side established by law." Rules like these all straightforwardly require me to drive on the right in places like the U.S., and not just as an exception to a general policy of driving on the left. But as we've just seen, the fact that a rule sometimes requires (and therefore permits) driving on the right is its downfall in the ideal conventionalist's book. If any of these rules were generally accepted, people in communities that actually drive on the right would still do so. But this would be worse or more objectionable than if people in those communities instead drove on the left, as they would if they accepted DLQ. So it seems that none of these rules could be ideal, given that DLQ is an option. Once again, we have encountered an ideal world problem: the real-life diversity in conventions that makes these rules seem superior to DLQ simply disappears in the worlds that are relevant to the ideal conventionalist's calculus. And again, this problem seems to generalize beyond this particular case. If general acceptance of A would be superior to general acceptance of B, then it will generally also be superior to general acceptance of any rule that boils down to "Follow whichever of A or B

others around you are following," since rules like these would at least sometimes permit people to follow B, even though their following A would be better or less objectionable.

Is this too quick? Perhaps it will be thought that only rules like these—ones that defer to our actual behavior or practices—can engender valuable social coordination by getting people to act on the same rule as others. But here again we face an ideal world problem: in the world where DLQ is generally accepted, nearly everyone *does* act on the same rule as others—namely, DLQ. So it seems that general acceptance of DLQ (or indeed, of any rule) would result in the same valuable social coordination as general acceptance of one of these "deferring rules". Or perhaps one might think that deferring rules are superior to rules like DLQ because their contextual nature allows the moral requirements to change as our practices do, whereas rules like DLQ fix the moral requirements once and for all. If our community switches from left-hand traffic to right-hand traffic, a deferring rule will accordingly switch from requiring left-side driving to requiring right-side driving, whereas DLQ will persist in requiring left-side driving. But when we imagine the world in which DLQ is generally accepted, we are presumably supposed to imagine a world where it is accepted indefinitely—that is, a world in which our practice is forever to drive on the left. In the world relevant to our evaluation of DLQ, then, there is no possibility that our practice might change, since a world where our practice changed would cease to be a world in which DLQ is generally accepted. So it seems that the possibility of conventional change is also idealized away in the ideal conventionalist's calculus. (And even setting this aside, driving on the left is the best practice available. It does not seem to be an advantage of deferring rules that they would let the moral rules keep up with changes in our practice if DLQ would simply endow us with the best practice to begin with.)

Could it be that the best rule is not one that defers to our actual behavior or practices, but rather one that requires us to do whatever would in fact best address the concerns that motivate having a rule in the first place? In our driving-side example, this would be a rule like "Drive on whichever side is actually safest in the circumstances" (DS). The now-familiar trouble, of course, is that DS is actually *worse* than DLQ at addressing the problem of safety when we compare worlds where each rule is generally accepted. In a world where it is generally accepted, DLQ always requires driving on the safest side, since it is generally safest if everyone drives on the left, and precisely what DLQ requires is driving on the left except when this would not be safest. The only difference between the world in which DLQ is generally accepted and the one in which DS is generally accepted is that in the latter world, some communities might drive on the right (or adopt other inferior driving-side practices), which is in fact *less* safe than if they drove on the left instead, as they would if they accepted DLQ. So we again have not found a superior alternative to DLQ.

The problem we keep bumping up against is this: DLQ seems problematic because it sometimes requires us to deviate from our actual norms in cases where it seems wrong to do so, but the ideal conventionalist's method for assessing rules erases the very possibility of deviating from the norm by following a rule, since each rule is evaluated in a world where it itself is the norm. As I have indicated throughout, this is just the well-known ideal world problem, but it is the ideal world problem in a new or at least underexplored guise. Early discussions of the problem focused on how it arises for rules that are blind to the possibility of error or intentional wrongdoing.[49] The now-classic example is "Never use violence." It seems like it should count against this rule that it would never allow us to use even limited violence to defend ourselves or others against violent attack, but this is a consideration that cannot be taken into account if the rule

[49] See, e.g., Parfit, *On What Matters*, vol 1., 312–20.

is evaluated in a world where it is universally accepted, since in such a world, there would (happily) be no violent attacks to defend against. Early examples like this one have the virtue of simplicity: the problem with the rule is an intuitive one, and it is easy to see how it makes trouble for ideal conventionalist theories. But simple examples admit of simple solutions. In this case, the solution is as simple as evaluating rules in worlds where they are *generally* but not *universally* accepted, which allows the ideal conventionalist to factor a realistic degree of violence into her assessment.

More recent work on the ideal world problem has aimed to show that the problem generalizes beyond cases that admit of easy fixes like this one. As it turns out, we can construct instances of the ideal world problem that can't be solved just by adjusting the degree of acceptance at which rules are assessed. But these instances end up looking pretty esoteric, relying on recherche devices like evil gremlins or "utility landmines."[50] Indeed, a leading response to these cases in the literature is just to insist that the worlds in which rules are evaluated be "normal," in the *ad hoc* sense of excluding such fantasy elements.[51] I agree with other commentators that this response is inadequate, and that for several reasons, the silliness of such cases does not tell against their potency as counterexamples to these theories (not least because these theories are attempting to state necessary moral truths that would hold even in worlds with evil gremlins). Still, it is not hard to imagine ideal conventionalists taking some degree of comfort in the current state of play around the ideal world problem. They could be forgiven for thinking that every instance of the problem is either easily solved or else hopelessly unrealistic.

---

[50] See, e.g., Rosen, "Might Kantian Contractualism Be the Supreme Principle of Morality?"; and Podgorski, "Wouldn't It Be Nice?"

[51] This response is considered, although not endorsed, by Rosen ("Might Kantian Contractualism Be the Supreme Principle of Morality?," 88ff), who attributes the response to Parfit. See also Podgorski, "Wouldn't It Be Nice?," 289.

But the ideal world problems we have encountered here are neither of these things. The rules we have considered share the simplicity of "Never use violence": DLQ and its alternatives are perfectly mundane rules governing an ordinary activity, and they generate ideal world problems without the need to resort to fanciful devices. There is nothing "abnormal" about the worlds that make DLQ come out superior to its alternatives (except, of course, that these worlds involve general acceptance of DLQ or its alternatives). At the same time, the problems we have encountered are more robust than the one generated by "Never use violence." At no point in setting up these problems did we rely on the assumption that general acceptance would be universal, and so nothing about our analysis changes if we assume a realistic degree of non-acceptance or non-compliance in the worlds of general acceptance. The non-acceptors will not tip the scales, since their behavior will be equally bad or objectionable no matter which of DLQ or its alternatives is generally accepted: whichever rule it is they don't accept, they will be driving on the wrong side of the road, with all the same consequences that entails. So DLQ will still outperform its rivals when evaluated at any level of acceptance sufficient to sustain a practice of driving on the left (since it is DLQ's ability to effect such a practice that makes it superior to rules that allow the practice of driving on the right to develop).[52] The problems we have seen here thus cannot be

[52] Of course, some ideal conventionalists advocate assessing rules at multiple levels of acceptance, including ones far too low to sustain a social practice (see, e.g., Michael Ridge, "Introducing Variable-Rate Rule-Utilitarianism," *Philosophical Quarterly* 56, no. 223 (April 2006): 242–53; and Parfit, *On What Matters*, vol 1., 317). But it seems to me unlikely that even these views could secure the superiority of any alternative to DLQ. For just one example, compare DLQ to "Drive on the side established by law" (DLaw). I have already argued that at levels of acceptance sufficient to sustain a practice of driving on the left, DLQ would be superior to DLaw, since the former would require driving on the safer left side, whereas the latter would permit driving on the less safe right side where that is what the law presently demands. At low enough levels of acceptance, DLaw would outperform DLQ for similar reasons: if very few people accepted DLQ, the norm everywhere would be to drive on the inferior right, whereas if few people accepted DLaw, the norm in the many places that currently drive on the right would be to drive on the superior left. At mid-range levels of acceptance, both rules, it seems, would be similarly awful: chaos would reign, with people everywhere closely split between a tendency to drive on the left or right. (Local equilibria might eventually be reached by safety-minded folks looking to avoid the chaos, but there is no reason to think these would be any better or worse under DLQ or DLaw). In short, DLQ would outperform DLaw at high levels of acceptance, underperform it at low ones, and be equally bad as it in the middle. Though I will not dig deeper into this hole here, this does not seem like

avoided simply by factoring in a realistic degree of non-acceptance. In short, these are instances of the ideal world problem that the ideal conventionalist could not even hope to dismiss as either trivial or fantastical.

More to the present point, we have stumbled upon a quite general class of rules for which the ideal world problem arises—namely, rules that defer to our actual practices. When there is a better alternative to our actual practice (or indeed, even where there is a possibility that we may *someday* adopt a less-than-ideal practice), rules that defer to our actual practice (or index to its associated behaviors or results) will generally not be superior in the ideal conventionalist's evaluation to the rules of the better practice themselves, because the better practice *becomes* our practice in worlds where its rules are generally accepted. Just as simple versions of ideal conventionalism are blind to the significance of actual violence because such violence is idealized away in worlds of universal pacifism, even more nuanced versions are blind to the significance of our actual social practices because these practices are replaced by better ones in worlds where better rules are generally accepted. This is a version of the problem that is interesting in its own right and has not received sufficient attention.

As we've already seen, the upshot of all this for the case at hand is that general acceptance of DLQ would be better or less objectionable than general acceptance of alternative rules requiring us to drive as others actually drive. These alternative rules thus cannot save the ideal conventionalist from DLQ's counterintuitive implications in cases like Desert Highway. If the ideal conventionalist wishes to avoid results like that I am sometimes required to drive on the

---

the beginning of a promising case for the superiority of DLaw. What reason could we possibly have to prefer the rule that is only better when few people accept it?

wrong side of the road even when there is nothing to be gained from it, she will have to try a different tack.

3.3 Other Ideal Rules

There is a different tack available for avoiding the counterintuitive implications of a seemingly ideal rule. Rather than trying to find a superior *alternative* to the rule in question, the ideal conventionalist might instead try to get the hoped-for results out of an ideal rule that concerns a different or more general sphere of activity than the rule in question, and which is therefore *not* an alternative to it. For example, the ideal conventionalist might try to accommodate the moral significance of practices like driving on the right by appealing to something like T. M. Scanlon's Principle of Established Practices, which he claims is non-rejectable:

> *Principle of Established Practices* (PEP): When "there is a need for some principle to govern a particular kind of activity, but there are a number of different principles that would do this in a way that no one could reasonably reject… if one of these (nonrejectable) principles is generally (it need not be unanimously) accepted in a given community, then it is wrong to violate it simply because this suits one's convenience."[53]

Now of course PEP as stated will be of no help to the ideal conventionalist in this particular case, since it only applies when several rules or principles are tied for non-rejectability (or, more generally, ideality), and DL and DR (or rather, DLQ and its right-side analogue, DRQ) are not tied for ideality—driving on the left is a superior policy to driving on the right, even if only marginally so. But there are similar rules in the vicinity that cast a wider net. Consider

[53] Scanlon, *What We Owe to Each Other*, 339.

> *Satisficing Conventionalism* (SC): When a rule is generally accepted in your community, follow it just in case it is good enough.[54]

I will leave open exactly how to cash out "good enough";[55] in the interest of charity, I will assume that the best or least objectionable reading of "good enough" is one that succeeds in picking out just those practices that we intuitively think are worth following. So SC seems to be just the kind of rule the ideal conventionalist needs: since "Drive on the right" is a good enough rule if anything is, SC will require me to follow it in places where it is accepted.

Of course, it only matters what SC requires if it is one of the ideal rules. But is it? Hooker argues that we should not accept it, although his argument for this conclusion is not that SC fails to be optimific but rather that it has counterintuitive implications.[56] Whether or not SC is ideal turns out to be a surprisingly difficult question, one that might depend on whether we assess rules for ideality one at a time or as a set. One way of understanding the ideal conventionalist's method of rule evaluation is this: we evaluate a rule by imagining the nearest possible world in which that rule is generally accepted, and comparing it to the nearest possible worlds in which alternative rules are accepted. On this approach, we determine whether SC is ideal by comparing worlds where people accept the rules they *actually* accept, with just one addition: either SC or an alternative to SC. If *those* are the relevant worlds, then the case for SC being ideal is straightforward: it would be better for people to follow all and only the socially accepted rules that are good enough, rather than ignoring some of the sufficiently good rules or also following some of the insufficiently good ones, as alternatives to SC would require. But rule consequentialists like Hooker (and some contractualists, like Jussi Suikkanen) characterize their view as one that compares *whole moral*

---

[54] Cf. Hooker, *Ideal Code, Real World*, 118.
[55] Cf. note 46, above
[56] See Hooker, *Ideal Code, Real World*, 118–21.

codes at a time.[57] On this "whole code" version of ideal conventionalism, we evaluate a whole moral code by imagining the nearest possible world in which that *entire set* of rules is generally accepted, and comparing it to the nearest possible worlds in which alternative sets of rules are accepted; an individual rule is one of the ideal rules if it belongs to the best-performing moral code. And it does not seem that SC could be ideal on this version of the view. The problem, again, is an ideal world problem. In the world where we generally accept the whole ideal moral code, there are no less-than-ideal rules that are generally accepted, and so SC would be otiose: it would just tell us to follow the rules that we already accept and would follow anyway (since the rules we accept are all ideal and therefore all good enough).[58] And if SC would be otiose as a member of the ideal set, it could not even belong to that set to begin with, since the costs of learning it would not be offset by any benefits.

So it seems that SC is one of the ideal rules only if the ideal conventionalist evaluates individual rules one at a time and not (as Hooker and Suikkanen advocate) whole moral codes at once. But suppose that we grant that method of evaluation, and therefore that SC is ideal. Still it is not clear that the ideal conventionalist is out of the woods. True enough, SC will require me to drive on the right wherever it is customary to do so. But DLQ will still require me to drive on the left in cases like Desert Highway. And isn't DLQ still one of the ideal rules, even if SC is too? If so, the ideal conventionalist has only set herself up to render conflicting verdicts in such cases: one ideal rule will require me to drive on the right while another requires me to drive on the left.

---

[57] See ibid., 32, and Jussi Suikkanen, *Contractualism* (Cambridge: Cambridge University Press, 2020), 34–5.

[58] Here I have assumed that if the whole ideal code were generally accepted, we would not generally accept any other social rules. This might be doubted: perhaps some rules not in the ideal code would survive acceptance of it because the ideal code contains no alternative to them—it is simply silent on the matters that they cover. But in this case, SC would be worse than otiose as a member of the ideal code, since it would require us to follow rules that by hypothesis are not members of the ideal set, and that it would therefore be better if we did not follow.

One option available to the ideal conventionalist is simply to accept this result. I will consider that response in a moment. First, though, it is worth examining the premise that leads to this result in the first place: *is* DLQ still one of the ideal rules if SC is? More generally, if SC is included in the ideal code, what are we to make of the ideal alternatives to the (merely) good-enough rules that SC requires us to follow? A natural answer is that the ideal alternatives to the good-enough rules we actually accept are still ideal rules, and so still members of the ideal set irrespective of SC's membership. Rules like DLQ are not rendered suboptimal by the existence of SC because they are not *alternatives* to SC: they concern different spheres of human life, address different questions about how we should behave in our shared world (e.g., which side of the road to drive on vs. how we should act in light of our community's conventions). But the ideal conventionalist might resist this line of thinking. The ideal code, she might argue, does not contain a rule like DLQ that directly addresses which side of the road to drive on, because that's just the wrong level at which to look for the ideal rules. The ideal rules are basic principles of morality, and we should not expect to find basic principles of morality for every level at which it is possible to pose questions about human action. We need a basic principle of morality that addresses sufficiently basic or high-level concerns like how we should act in light of our community's conventions, but questions like which side of the road to drive on are too applied or low-level to have their own rule in the ideal code—these matters are properly settled by conventional rules, which are given moral force exclusively via an ideal rule like SC. So there is no conflict between SC and the ideal driving-side rule, because the ideal code *has no* driving-side rule.

Now, in some sense there is nothing stopping the ideal conventionalist from making this move: it is her view, after all, and if she wants to amend it to stipulate that right and wrong are determined by only the *sufficiently basic or high-level* rules whose general acceptance would be

ideal, that's her prerogative. But many ideal conventionalists have stuck to their guns on this point, and held that the question of whether there should be a rule at a given level of abstraction is to be answered only by considering whether general acceptance of a rule at that level would be ideal.[59] On this approach, it may well turn out that there is no ideal rule that speaks directly to (say) which side of the road to drive on, but if this is so, it will be because it would be worse to have a rule at that level than to have matters at that level be governed only by higher-level rules.

Again because of the ideal world problem, though, it would be better to have a rule at the level of DLQ than to let the question of driving side be governed entirely by SC. If we generally accepted SC but not DLQ, some communities would continue to drive on the right, whereas if we generally accepted DLQ, the superior practice of driving of the left would everywhere replace the inferior practice of driving on the right. This would be so even if we accepted both DLQ and SC, since in a world where DLQ is generally accepted, SC would just redundantly require us to follow DLQ. Moreover, acceptance of DLQ would be costless—even the tiny cost of having to learn and remember one more rule would wash out, since people would presumably have to learn and remember a rule about which side to drive on even if that rule *weren't* one of the rules in the ideal code. So although we can imagine a variant of ideal conventionalism with the resources to reject DLQ as too "low-level" a rule, a more thoroughgoing commitment to embracing the rules whose general acceptance would be ideal seems to favor a rule at DLQ's level. And the reasoning for this conclusion seems to generalize to any "low-level" rule: for any activity A, it would be better for us all to adopt the best low-level rule concerning A than to let A be governed only by a higher-level rule like SC that might require us to follow a worse rule concerning A.

---

[59] See, e.g., Brandt, *A Theory of the Good and the Right*, 290.

But perhaps the ideal conventionalist is willing to accept all this. Yes, she might say, SC and DLQ are both rules in the ideal code, and yes, they will make conflicting demands on us in cases like Desert Highway. But that, she might insist, is just life. Sometimes the ideal rules conflict with one another, and any plausible ideal conventionalist view will have a method for resolving such conflicts. As long as that method always privileges the requirements of SC over those of DLQ, the conflict needn't worry us—at the end of the day, ideal conventionalism will still hold, correctly, that I am required to drive on the right when doing so is customary. And similarly for any low-level rule that SC might conflict with.

I will not take a stand here on what the correct method is for resolving conflicts between ideal rules. For all I will say, perhaps the correct method really would privilege SC whenever it conflicts with rules like DLQ. But it seems to me that ideal conventionalism has gone badly wrong even if it *does* consistently resolve such conflicts in SC's favor. As Hooker points out, "to accept a code of rules is just to have *a moral conscience of a certain shape.*"[60] Those who accept the ideal rules will feel, among other things, an aversion towards doing the things those rules forbid, and "[w]hen rules conflict, so do the aversions that are attached to them."[61] This seems to me to be a plausible bit of ideal conventionalist moral psychology. But it fails to do justice to the phenomenology of cases like Desert Highway if such cases are understood to involve a conflict between SC and DLQ, even if that conflict is ultimately resolved in favor of SC. When I imagine being in Desert Highway, I don't imagine feeling conflicted, even to the smallest degree. It is not as if I am between a rock and a hard place, that I feel averse to driving on the right yet compelled to do so by an overriding aversion to driving on the left. What I imagine feeling is just an aversion

[60] Hooker, *Ideal Code, Real World*, 91.
[61] Ibid., 90.

to driving on the left, with no conflicting aversion, however weak, to driving on the right. Moreover, this seems like the right thing to feel—it's not just some deficiency in me as a moral agent. I feel unconflicted about such a case, I want to say, because it involves no conflict. In short, even if the ideal conventionalist can achieve extensional adequacy by arguing that SC always overrides or outweighs rules like DLQ, it seems she would be mistaken to even characterize such cases as involving a conflict of rules in the first place.

Even granting that SC is ideal, then, the ideal conventionalist cannot always do justice to the moral significance of our actual practices by appealing to it, since at least in some cases the good enough rules that SC requires us to follow will conflict with the ideal alternatives to those rules that ideal conventionalism also requires us to follow. As I've noted, this is a problem that generalizes beyond DLQ to other "low-level" rules. But it also generalizes beyond SC, to other rules that might be employed to capture the moral significance of our actual practices in the face of ideal alternatives to those practices. For instance, Hooker, who rejects SC, proposes to capture the moral force of our merely good-enough practices by appeal to fairness. Surely some rule of the rough form "Don't treat others unfairly" will be ideal. Hooker's suggestion is that we sometimes treat others unfairly by failing to follow the practices actually accepted in our community (for example, because we free-ride on their sacrifices or upset their legitimate expectations formed on the basis of those practices), such that the ideal rule requiring us not to treat others unfairly would sometimes require us to follow our actual practices, even when these are less than ideal.[62]

Now as it happens, I do not think this strategy will actually capture all the cases in which we should follow our merely good-enough practices (for example, it is not clear to me that I treat others *unfairly* by driving on the wrong side of the road, in general or especially in cases like

[62] See ibid., 121ff.

Desert Highway). But even if every case where we should follow our merely good-enough practices were one in which it would be unfair not to, the ideal conventionalist would not be home free. For in just the sort of case that the ideal conventionalist is trying to accommodate—namely, cases where the ideal rules conflict with those of our actual practices—the ideal conventionalist will now find the ideal rules offering conflicting verdicts: for example, the rule forbidding unfairness will (arguably) require driving on the right, while DLQ requires driving on the left. For the reasons just discussed, this seems to me an implausible account of our moral situation in such cases even if the rule forbidding unfairness ends up always taking priority over rules like DLQ. Nor is there an ideal conventionalist case for excluding DLQ from the ideal code in light of this fairness rule, since it would be better for us all to accept DLQ and drive on the left irrespective of whether we also accept a rule forbidding unfairness.

The strategy of trying to get around problematic ideal rules like DLQ by appealing to other ideal rules thus does not seem promising. At best, this approach allows the ideal conventionalist to say that we should follow the rules of our merely good-enough practices while continuing to accept the conflicting rules of the ideal alternatives to those practices. And that does not seem to be enough.

3.4 Conclusion

I have argued that moral theories like contractualism and rule consequentialism are committed to a sort of ideal conventionalism that prevents them from ascribing the right kind of moral significance to our actual practices. The fact that we actually drive on the right in the United States seems to settle the question of which side I ought to drive on, yet on these theories, the

question seems to be settled by the fact that it would be best or least objectionable for us to drive on the left. We have now seen that this problem is not as easy to escape as one might suppose. Rules that defer to our actual practices or their resulting behaviors face an ideal world problem that makes them come out inferior to the rules of the ideal practice itself. And more general rules requiring us to abide by good-enough practices or treat others fairly will not *replace* the rules of the ideal practice on the ideal conventionalist's picture, but at best *join* them in the ideal set, leading to implausible diagnoses of conflict in cases where we should obviously follow our actual practices without regret.

As I have suggested throughout, this is a quite general problem that extends beyond our simple example of driving on the left. By way of closing, let me briefly sketch just one of the broader implications it might have. Consider the set of property rules actually accepted in some community—the U.S., say, or perhaps a smaller subdivision of it. These rules define our conventional property rights and associated obligations, but at least where these rules are good enough, we tend to think that they play a role in determining our moral rights and duties as well. But can the ideal conventionalist account for this? Ideal conventionalism requires us to follow the property rules whose general acceptance would be ideal. Whatever the ideal property rules are, though, surely they differ in at least *some* respect from the ones we actually accept. No matter how highly we think of our existing property conventions, it seems naïve to think that they are *perfect*. So the rules of our actual property conventions will not (all) be ideal rules in their own right. As we have now seen, though, there are issues facing other ideal conventionalist strategies for giving moral force to our actual rules. Deferring rules like "Follow the good enough property rules actually accepted (or given legal force) in your community" would face the same ideal world problem faced by rules like "Drive on the customary side": if the deferring rule were generally

accepted, we would continue to have property practices that are worse or more objectionable than the ideal property convention, which is what we would have if we instead accepted the ideal property rules themselves. And rules like SC would at best allow the ideal conventionalist to give moral force to our actual property rules *in addition to* the ideal property rules, with all the counterintuitive implications that brings. The upshot seems to be that ideal conventionalists are stuck with a natural rights view on which the rules of the ideal property convention morally bind us irrespective of whether we have actually adopted it.

There is of course much more that could be said about this case; I don't mean to suggest that this brief sketch of an argument is conclusive. What I do mean to suggest, though, is just how wide the scope of this problem is. It is one that will crop up constantly—one that the ideal conventionalist will have to wrestle with time and again, even if it turns out to occasionally be solvable in some particular context.

In the previous chapter, we noted two features of theories that seem to make them susceptible to the misapplication dilemma: a two-level structure in which acts are evaluated by appeal to rules and rules by appeal to the consequences of their acceptance, and an optimizing approach to rule selection that selects only the rules that perform best in the theory's evaluation. These two features also seem to be the culprits in this case, at least when "acceptance" is read as "general acceptance"; indeed, taken together, they seem to constitute the ideal conventionalism that has proved so troublesome for these theories. For it is the counterfactual appeal to general acceptance (the "conventionalism" of the "ideal conventionalism") that leads these theories into the ideal world problem; as Abelard Podgorski puts it, this problem "faces any view that determines what we as individuals ought to do in this world by evaluating worlds that differ from

the actual world in more than what is up to us."[63] If these theories ceased to compare rules by imagining their general acceptance, they might avoid idealizing away the diversity of our local customs, and thus be in a position to appeal to that diversity to justify rules that defer to our actual practices. Alternatively (or perhaps additionally), a satisficing version of these theories (one that abandoned the "ideal" aspect of "ideal conventionalism") could easily accommodate the notion that we set out to vindicate—namely, that many of the rules we actually accept are good enough to follow even though they are less than ideal. Of course, it is not immediately clear whether an appealing view would remain if these theories were divorced from their ideal conventionalism. Given where ideal conventionalism leads, though, it seems well worth these theorists' time to find out.

[63] Podgorski, "Wouldn't It Be Nice?," 279.

4.0 Putting Wronging First

Often, when what we've done is wrong, it's because we've wronged a particular person. If I've promised you that I'll drive you to the airport but go to the movies instead, what I've done is wrong because I've wronged *you*. The same goes if I pick your pocket or kick you in the shin: in doing such things, I wrong you, and this is why it's wrong for me to do them.

These commonsensical claims posit an explanatory relation between wronging and wrongness: they represent certain acts as being wrong *in virtue of* wronging a person. Although there has been a recent surge of philosophical interest both in wronging and in explanatory relations in ethics, the explanatory relationship between wrongness and wronging has stayed just below the surface, remaining virtually unexplored. My aim here is to explore this relationship and its surprising consequences for moral theory.

In Section 1, I will motivate the claim with which we began: an act can be wrong *because* it wrongs a person. This thesis, which I will call *Wronging First*, is supported not only by our ordinary explanations of moral requirements, but also by its ability to help us solve three philosophical puzzles in one fell swoop. In Section 2, I will show how Wronging First operates as a constraint on moral theories, requiring them to make room for wronging in their accounts of the wrong-making features of acts. I will then sketch two Kantian accounts of right and wrong, which will serve as case studies in how moral theories might (fail to) meet this constraint. In Section 3, I will argue that a Kantian theory that grounds right and wrong in the universalizability of maxims is inconsistent with Wronging First, and must therefore be abandoned. In Section 4, I will show how a Kantian theory that grounds right and wrong directly in the value of humanity can be made consistent with Wronging First, but only if the Kantian is willing to endorse an account of

wronging that might have surprising implications for (among other things) promising, free-riding, emergency rescue, and doxastic wronging. A key takeaway for theorists of all stripes, which I will discuss briefly in Section 5, is that a moral theory will be hard-pressed to accommodate Wronging First if it cannot give its own account of what it is to wrong someone. Wronging First thus suggests a new approach to moral theory, one that illuminates wronging rather than obscuring it.

4.1 Wronging First

Central to our investigation is the distinction between (moral) wrongness and (moral) wronging. Whether it is *wrong* for X to φ is a matter of whether morality forbids X to φ—whether, in φ-ing, X runs afoul of a moral requirement. The claim that it is wrong for X to φ applies a monadic predicate to X's φ-ing. By contrast, the claim that X *wrongs Y* by φ-ing applies a dyadic predicate that relates X's φ-ing to a victim, Y. Whether X wrongs Y is not in the first instance a matter of whether X violates a moral requirement; instead, it is a matter of whether X in some sense violates *Y herself*, whether by violating Y's person, trust, privacy, rights, etc. This is reflected in the interpersonal phenomena that are commonly taken to be hallmarks of wronging: when X wrongs Y, Y (and Y alone) is warranted in resenting X, is owed an apology or restitution from X, has the power to forgive X, etc.[64] Other ways of saying that X would wrong Y by φ-ing include that X owes it to Y not to φ, that X has a (directed) duty to Y not to φ, and that Y has a (claim-

[64] See, e.g., Stephen Darwall, *Morality, Authority, and Law* (Oxford: Oxford University Press, 2013), 30-4.

)right against X that X not φ.[65] Other ways of saying that it is wrong for X to φ include that it is impermissible for X to φ, that X is (morally) required not to φ, and that X (morally) must not φ.

A perennial question about wrongness and wronging is when, if ever, these phenomena come apart. Can we act wrongly while wronging no one, or wrong someone without acting wrongly? This is an important question, but it is not my topic here. What I am interested in is a question of explanatory priority that arises in the many cases where these phenomena *don't* come apart: When an act both is wrong and wrongs someone, is the act wrong because it wrongs that person, or does it wrong that person in part because it is wrong, or neither?

I think the first option is correct. More precisely, I will defend

> *Wronging First*: At least other things equal, when X wrongs Y by φ-ing, this makes it the case that X's φ-ing is wrong.[66]

Another way of putting Wronging First would be to say that for all persons X, *wronging X* is a wrong-making feature of acts. Like the "other things equal" clause, this formulation leaves open the possibility that the wrong-making force of wronging may be outweighed or defeated in certain exceptional circumstances, as is the case with many other wrong-making features (e.g., causing harm, being the breaking of a promise, etc.). Wronging First is thus neutral on whether there are *permissible wrongings*—that is, acts that wrong someone but are not wrong. It is also neutral on

---

[65] See, e.g., Michael Thompson, "What is It to Wrong Someone? A Puzzle About Justice," in *Reasons and Values*, ed. R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith (Oxford: Clarendon Press, 2004), 334; Margaret Gilbert, *Rights and Demands* (Oxford: Oxford University Press, 2018), 47–8 and 65–71); F. M. Kamm, *Intricate Ethics* (Oxford: Oxford University Press, 2007), 239; Simon Căbulea May, "Directed Duties," *Philosophy Compass* 10, no. 8 (August 2015): 523; Joel Feinberg, "The Nature and Value of Rights," *Journal of Value Inquiry* 4, no. 4 (December 1970), 249–50; and H. L. A. Hart, "Are There Any Natural Rights?," *Philosophical Review* 64, no. 2 (April 1955), 180.

[66] I call this view "Wronging First" because it asserts that wronging is first relative to wrongness in the order of explanation. It does not assert that wronging is first in the sense of being ungrounded or normatively basic. That is, it is not an alternative to the "Reasons First" program in metaethics.

the possibility of *victimless wrongdoing*—that is, wrong acts that wrong no one—since it does not say that *every* wrong act is wrong because it wrongs someone.

Note too that, although Wronging First says that wrongings are (usually) wrong because they are wrongings, it doesn't deny that these wrongings are *also* wrong because of more ordinary wrong-making features (again: causing harm, being the breaking of a promise, etc.), since these more ordinary features might make acts wrong *by* making them wrongings. This sort of transitivity will be important later, when we explore the implications of Wronging First for moral theory.

But first, why think that Wronging First is true? I suggested at the outset that the view is commonsensical, and it is only more so when translated into some of the terms mentioned above: Wronging First says that we can be required to do something because we owe it to someone to do it; that we can act wrongly in virtue of violating someone's right against us (that is, in virtue of violating a duty that we owe to her). We might also state Wronging First in virtue-ethical terms by saying that we can act wrongly in virtue of treating someone unjustly,[67] a claim that is hard to deny even if one is not a virtue ethicist. Wronging First sounds plausible, I think, because it accords with a familiar way of explaining why certain acts are wrong. To explain why I mustn't occupy this parcel of land, for example, you might cite your right to it; in order to explain why it would be wrong for us to surrender our post to the enemy, I might cite the fact that we owe it to our fellow soldiers not to. We frequently account for moral requirements in this way, by appealing to rights or what we owe to others. We could concoct various revisionary stories about why these

[67] Cf., e.g., Philippa Foot, "Euthanasia," *Philosophy & Public Affairs* 6, no. 2 (Winter 1977): 97; Judith Jarvis Thomson, "A Defense of Abortion," *Philosophy & Public Affairs* 1, no. 1 (Autumn 1971): 56ff; and Joel Feinberg, "Voluntary Euthanasia and the Inalienable Right to Life," *Philosophy & Public Affairs* 7, no. 2 (Winter 1978): 119–20.

explanations merely *seem* to work, but the simplest story is that they actually work because Wronging First is true.

I am not the first to find Wronging First intuitive. G. E. M. Anscombe, using "a wrong" to mean "a wronging," says that "What is wrong about an act that is wrong may be just this, that it is *a* wrong."[68] Stephen Darwall includes among the "wrong-making features of wrongful actions" the fact "that it would violate someone's rights and so wrong that person."[69] And Joel Feinberg maintains that

> If Nip has a claim-right against Tuck, it is because of this fact that Tuck has a duty…. It is only because something from Tuck is *due* Nip (directional element) that there is something Tuck *must do* (modal element).[70]

R. Jay Wallace concurs, writing that "An action can be 'to-be-done' or 'not-to-be-done' just insofar as and just because it is something that we owe it to another party to do or to refrain from doing."[71] In their various idioms, these philosophers all affirm that an act can be wrong because it wrongs a particular person.

Indeed, philosophers have hung real argumentative weight on Wronging First, and it could bear more—there are at least three philosophical puzzles that Wronging First would help solve. First, many philosophers presuppose Wronging First as part of their solution to the "paradox of deontology." Many moral prohibitions are such that it is wrong to violate them even to prevent a greater number of violations of them. It is wrong for me to kill you by throwing you in front of an oncoming trolley, say, even if doing so would prevent the murderous trolley-driver from killing five others who are tied to the tracks. The challenge is to explain how there could be "constraints"

---

[68] G. E. M. Anscombe, "On the Source of the Authority of the State," in *Ethics, Religion, and Politics* (Oxford: Basil Blackwell, 1981), 138.
[69] Darwall, *Morality, Authority, and Law*, 67–8.
[70] Feinberg, "The Nature and Value of Rights," 250.
[71] R. Jay Wallace, *The Moral Nexus* (Princeton: Princeton University Press, 2019), 49.

(or "restrictions") with this seemingly paradoxical character. And as it happens, many philosophers think this challenge can be met by appeal to wronging or the associated notion of rights. F. M. Kamm, for instance, suggests that constraints must "be understood as victim- rather than agent-focused—that is, the agent acts wrongly if he violates the constraint because some right of the victim's is being transgressed."[72] Daniel Muñoz agrees:

> We already know the source of restrictions: they are based in rights. I may not throw you onto the tracks because you have rights against harm… you have a right against me that I not throw you in front of trolleys.[73]

Rahul Kumar puts the same point in terms of wronging, writing that "the violation of a constraint, or interference with a permission, is intuitively thought to be wrong because doing so amounts to *wronging* another person."[74] Again in their various idioms, these philosophers suggest that the first step in dissolving the air of paradox around constraints is to recognize that it is wrong to violate constraints because in so doing, we wrong someone (or violate someone's rights, etc.). This is only a first step; the harder step is defending the relevant rights. But it is a step that can only be taken if Wronging First is true, that is, if acts can be wrong in virtue of wronging someone.

Next, consider the claim that rights are prior to, or ground, their correlative duties. This has struck many philosophers as intuitive,[75] but has also seemed to be inconsistent with the Hohfeldian doctrine that X's having a right against Y just *is* Y's having a duty to X.[76] How could rights be

---

[72] F. M. Kamm, "Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status," *Philosophy and Public Affairs* 21, no. 4 (Autumn 1992): 355.

[73] Daniel Muñoz, "From Rights to Prerogatives," *Philosophy and Phenomenological Research* 102, no. 3 (May 2021): 609. At 608, Muñoz puts this "because" claim in other explanatory terms, writing that "you have rights against harm, which restrict my choices by making it wrong for me to save the five at your expense."

[74] Rahul Kumar, "Defending the Moral Moderate," *Philosophy & Public Affairs* 28, no. 4 (Autumn 1999): 280.

[75] See, e.g., Feinberg, "The Nature and Value of Rights," 250; Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1986), 171; and Ariel Zylberman, "Relational Primitivism," *Philosophy and Phenomenological Research* 102, no. 2 (March 2021): 407.

[76] The seeming inconsistency is noted in Feinberg, "The Nature and Value of Rights," 249–50, and Kamm, *Intricate Ethics*, 241. On the Hohfeldian doctrine, see Wesley Newcomb Hohfeld, *Fundamental Legal Conceptions* (New Haven: Yale University Press, 1923), 38 and 73; Luís Duarte d'Almeida, "Fundamental Legal Concepts: The Hohfeldian Framework," *Philosophy Compass* 11, no. 10 (October 2016): 555-6; Judith Jarvis Thomson, *The Realm*

prior to their correlative duties if they just *are* their correlative duties? Wronging First supplies a solution, although we will see it only if we resist the urge to reify rights and duties and instead state these priority and identity claims in terms of the *facts* they relate. The Hohfeldian identity is between the fact that X has a right against Y that Y φ and the fact that Y *has a duty to X* to φ, where the latter fact, as we've seen, is to be cashed out in terms of Y's owing X or standing to wrong X. So we could restate the identity thus: the fact that X has a right against Y that Y φ just is the fact that Y owes it to X to φ, that is, just is the fact that Y would wrong X by not φ-ing. This identity claim would indeed be inconsistent with the claim that the fact that X has a right against Y that Y φ is prior to the fact that Y *has a duty to X* to φ (that is, owes it to X to φ). But the priority claim is more charitably interpreted as follows: the fact that X has a right against Y that Y φ is prior to the fact that Y *has a duty* to φ, that is, the fact that Y is *required* to φ, or that it would be *wrong* for Y not to φ. And this claim is consistent with the identity claim. Both can be true after all. But notice that both can be true only if Wronging First is true: if facts about rights just are facts about wronging (i.e., directed duty), then they can be prior to facts about duty *simpliciter* (i.e., wrongness) only if wronging is prior to wrongness.

Finally, consider the claim that practice-based accounts of the wrong of promise-breaking (such as Rawls's)[77] are inadequate because they cannot account for the fact that promise-breakers wrong their promisees in particular. This claim has become gospel truth in the promising literature.[78] Yet, intuitive as it is, it is puzzling why it should be true. To be sure, the fact that

---

of Rights (Cambridge: Harvard University Press, 1990), 39–42; Gilbert, *Rights and Demands*, 47–8; and Feinberg, "The Nature and Value of Rights," 249–50.

[77] See John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971), 342-7, and cf. 111-2.

[78] See, e.g., T. M. Scanlon, *What We Owe to Each Other* (Cambridge: Harvard University Press, 1998), 316; Niko Kolodny and R. Jay Wallace, "Promises and Practices Revisited," *Philosophy & Public Affairs* 31, no. 2 (Spring 2003): 126; and Stephen Darwall, "Demystifying Promises," in *Promises and Agreements: Philosophical Essays*, ed. Hanoch Sheinman (Oxford: Oxford University Press, 2011), 263–4.

explains the wrongness of promise-breaking on Rawls's account—namely, the fact that promise-breaking is forbidden by the rules of a just practice whose benefits promisors voluntarily accept—does not suffice to explain why the promise-breaker wrongs her promisee in particular. But why *should* it suffice to explain this further fact? Why couldn't Rawls hold that an additional fact must be added to explain why the promisee in particular is wronged—for example, the fact that the promisee alone has the power to waive the promissory obligation?[79] Since this move would allow Rawls to account for the fact that promise-breaking *wrongs* the promisee without forcing him to abandon his original account of why promise-breaking is *wrong*, the intuitive objection against Rawls is only good if this move is unavailable. And one straightforward way of denying this move is to appeal to Wronging First. If promise-breaking is wrong *because* it wrongs someone, then the facts that explain why promise-breaking wrongs the promisee must be the same ones that explain why promise-breaking is wrong, since they will explain why promise-breaking is wrong *by* explaining why they wrong the promisee. If Wronging First is true, the objection against Rawls is as compelling as it seems.[80]

I do not mean to suggest that the *only* way to solve these puzzles is to appeal to Wronging First. Just as there might be other ways of accounting for our propensity to explain what is required by citing what is owed, each of these puzzles individually might admit of alternative solutions. But

---

[79] As a matter of fact, Rawls does hold that obligations springing from the principle of fairness (and thus promissory obligations on his account) are owed to all who cooperate to maintain the relevant practice. See Rawls, *A Theory of Justice*, 113. But we could imagine a Rawlsian who denounces this view; the objection against Rawls is not merely that he *endorses* this view of who is owed, but that his position on the wrongness of promise-breaking *implies* it.

[80] Some have suggested that a similar objection can be leveled against moral theories: a moral theory is inadequate if its wrong-makers cannot account for who is wronged. See, e.g., Richard Yetter Chappell, "The Right Wrong-Makers," *Philosophy and Phenomenological Research* 103, no. 2 (September 2021), esp. 429–30 and 438–9; and Aleksy Tarasenko-Struc, "Kantian Constructivism and the Authority of Others," *European Journal of Philosophy* 28, no. 1 (March 2020): 77–92. As we are about to see, I think this suggestion is right, although existing arguments for it are few and uncompelling. Once again, Wronging First provides a ground for the objection: the facts that explain wronging must be the ones that explain wrongness because they will explain wrongness *by* explaining wronging.

it is surely a point in favor of Wronging First that it helps us solve all dthree of these puzzles (and account for our ordinary explanations) in one fell swoop. Several otherwise perplexing things fall into place if Wronging First is true. At the very least, that's reason to see where Wronging First would take us.

4.2 Moral Theory

Although philosophers have, as I've noted, touched upon Wronging First before, they have rarely appreciated its significance for moral theory.[81] One of the central tasks for moral theory is to say what makes right acts right and wrong acts wrong. Wronging First says that one thing that can make a wrong act wrong is that the act wrongs a particular person. It is thus a claim that could potentially conflict with a moral theory's account of right and wrong. If the account of right and wrong says that only such-and-such facts make acts wrong, and the fact that an act wrongs a particular person is not among them, the account in question will be inconsistent with Wronging First. Wronging First thus serves as a constraint on moral theories: theories must include wronging among their wrong-makers, on pain of contradicting the intuitive and powerful thesis that we can act wrongly in virtue of wronging a person.

Not every theory meets this constraint. According to act-utilitarianism, for instance, the only facts that play a role in making an act wrong are (1) facts about the utility or disutility that would result from the various acts available to the agent and (2) the facts that make these facts the

[81] An exception is Wallace, *The Moral Nexus*, which explores the implications of a stronger claim: that an act's being wrong *just is* its wronging someone, such that that *every* wrong act is wrong because it wrongs a particular person.

case. Plausibly, though, the fact that some particular person is wronged by the act falls into neither of these categories. If that's right, then Wronging First rules out act-utilitarianism.

That was just a quick example, and no doubt an oversimplified one. But of course, it would be no great surprise if Wronging First were inconsistent with act-utilitarianism, a moral theory which famously eschews individual rights. What would be more interesting is if Wronging First made trouble for a view that *tries* to allow for duties to particular persons.

One such view is Kantianism, whose proponents have often emphasized the relational aspects of morality. Christine Korsgaard, for instance, argues that "The subject matter of morality is not what we should bring about, but how we should relate to one another," and suggests that constraints against such acts as hurting people, lying, and breaking promises "do not spring from the consequences of those actions, but rather from the claims of those with whom we interact to be treated by us in certain ways."[82] Here, Korsgaard seems to be agreeing with Kamm, Muñoz, and Kumar that these acts are wrong because of others' claims against us—that is, because of the fact that we wrong others by performing them. We should therefore think of Wronging First not as imposing a constraint on Kantian ethics from the outside, but rather as articulating a constraint imposed on Kantian ethics by its own relational ambitions.

Over the next two sections, we will take Kantianism as our case study in how Wronging First operates as a constraint on moral theories. Since there is disagreement among Kantians over what makes acts wrong, I will examine how two different Kantian views might meet this constraint.[83] First, we will consider the popular constructivist reading of Kant on which the

[82] Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 275 and 291.
[83] I don't mean to suggest that these two views exhaust the Kantian's options; they're just two views that find support in the literature.

universalizability test[84] or "CI procedure" does not track independently existing moral facts, but rather generates moral facts that obtain only because they result from the procedure. On this reading, an act is wrong not just *when*, but also *because* (and only because) the maxim on which it is done cannot be willed as universal law. [85, 86] As constructivist Kantians have been careful to note, this view is consistent with there being other, more "substantive" wrong-making facts, so long as these other wrong-making facts contribute to making an act wrong only by playing a role in the CI procedure—that is, only by helping to make it the case that the agent's maxim is non-universalizable.[87] More precisely, then, we can state the view thus:

> *Universal Law Constructivism*: X's φ-ing is wrong when and because the maxim on which X φ-s is non-universalizable, and every other fact that makes X's φ-ing wrong does so only by making it the case that X's maxim is non-universalizable.

On this account, the fact that an agent's maxim is non-universalizable is not the only wrong-making fact, but it is, in Parfit's phrase, the only "*higher-level* wrong-making property or fact, under which all other such properties or facts can be subsumed."[88]

---

[84] Cf. Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor and Jens Timmerman (Cambridge: Cambridge University Press, 2012), 4:421–3.

[85] See, e.g., Andrews Reath, *Agency and Autonomy in Kant's Moral Theory* (Oxford: Clarendon Press, 2006), 103, 143 and 169, and cf. 118–9, 164, and 222. Reath states the view as a view about what makes maxims impermissible, rather than as a view about what makes acts wrong, but I don't think anything hangs on this, especially since a maxim's being impermissible seems to be nothing other than its being a maxim that it's wrong to adopt or act on. On the order of explanation in constructivism more generally, see John Rawls, *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge: Harvard University Press, 2000), 242–3; and Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 36–7.

[86] This sort of constructivism is distinct from Kantian constructivism as a metaethical view. Metaethical constructivism is the view that all normative truths are true only in virtue of resulting from a procedure of construction. By contrast, what we might call *moral* constructivism is the view that truths about moral right and wrong are true only in virtue of resulting from a procedure of construction. It is possible to be a moral constructivist without being a metaethical constructivist; arguably, Scanlon's view is one on which facts about right and wrong are constructed, whereas some other normative facts (e.g., those about reasons) are not. See Scanlon, *What We Owe to Each Other*, and T. M. Scanlon, *Being Realistic About Reasons* (Oxford: Oxford University Press, 2014).

[87] See Reath, *Agency and Autonomy in Kant's Moral Theory*, 103 and 118–9.

[88] Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), 369.

Some Kantians, however, have resisted the notion that the universalizability test does explanatory work. As an alternative, they propose an account of right and wrong that appeals directly to the value of humanity (or rational nature).[89] As Allen Wood puts it, the view is that "the moral wrongness of… actions always consists fundamentally in the way they show disrespect for the objective value of rational nature."[90] I think we can gloss this view as follows:

> *Humanity-Grounded Kantianism*: X's φ-ing is wrong when and because it shows disrespect for the value of humanity, and every other fact that makes X's φ-ing wrong does so only by making it the case that X's φ-ing shows disrespect for the value of humanity.

On this account, the higher-level wrong-making fact under which all others are subsumed is the fact that X's φ-ing shows disrespect for the value of humanity. All sorts of facts might make an act wrong, but if Humanity-Grounded Kantianism is correct, they only ever do so by making the act one that shows disrespect for humanity, since this is what wrongness always fundamentally consists in.

Note that these two views are not disagreeing about which of Kant's formulations of the categorical imperative is the "right" one, or even the most important one. They are only disagreeing about which formulation *explains* the wrongness of acts. Universal Law Constructivists needn't (and don't) deny that the value of humanity is an indispensable element of Kantian ethics, and something that we must respect—they merely deny that it is (in the first instance, anyway) *in virtue of* disrespecting humanity that wrong acts are wrong. Similarly, proponents of Humanity-Grounded Kantianism needn't (and don't) deny that universalizability is an indispensable element

---

[89] Cf. Kant, *Groundwork of the Metaphysics of Morals*, 4:428–9.
[90] Allen Wood, "Humanity as End in Itself," in Derek Parfit, *On What Matters*, vol. 2, ed. Samuel Scheffler (Oxford: Oxford University Press, 2011), 63. Cf. Allen Wood, *Kantian Ethics* (Cambridge: Cambridge University Press, 2008), 82. See also Barbara Herman, *The Practice of Moral Judgment* (Cambridge: Harvard University Press, 1993), 124 and 127, and cf. 226–30.

of Kantian ethics—they merely deny that it is (in the first instance, anyway) *in virtue of* their maxims' being non-universalizable that wrong acts are wrong. For all either view says, the two views may even be extensionally equivalent: it may be that an act shows disrespect for the value of humanity just in case the maxim on which it is done is non-universalizable.[91] Even if the two views always agreed about *which* acts are wrong, though, they would still disagree about *why* those acts are wrong. This is an important question in its own right, and Kantians should (and do) care about where they come down on it.

I will now turn to the question of whether these two views can accommodate Wronging First. What we will find is that one of them can do so only with the help of certain controversial theses about wronging, and the other cannot do so at all.

4.3 Universal Law Constructivism

Is Universal Law Constructivism consistent with Wronging First? Wronging First says that the fact that X wrongs Y by φ-ing can make X's φ-ing wrong; Universal Law Constructivism says that only two sorts of facts make X's φ-ing wrong: (1) the fact that X's maxim in φ-ing is non-universalizable, and (2) the facts that make this fact the case. For the two to be consistent, then, the fact that X wrongs Y by φ-ing would have to be either (1) the fact that X's maxim in φ-ing is non-universalizable or (2) a fact that helps make X's maxim non-universalizable.

[91] For doubts about their extensional equivalence, see Korsgaard, *Creating the Kingdom of Ends*, 135–44, and Wood, *Kantian Ethics*, 81–2.

We can dismiss the first option quite quickly. If the fact that X wrongs Y by φ-ing were identical to the fact that X's maxim in φ-ing is non-universalizable, it would be impossible for one to obtain without the other. But there are many choices of X, Y, and φ for which X does not wrong Y by φ-ing, even though X's maxim in φ-ing is non-universalizable. Allen doesn't wrong Barbara by killing Christine, even if Allen's maxim in killing Christine is non-universalizable.

Of course, this is too quick, for the Kantian might hold that the fact that X's maxim in φ-ing is non-universalizable is identical to the fact that X wrongs Y by φ-ing only in certain instances: for example, when Y features in the act-description φ, or in the content of X's maxim. But this strategy doesn't work either. If Allen promises Barbara not to dance with Christine, his maxim in *dancing with Christine* will (presumably) be non-universalizable, but he doesn't wrong *Christine* by dancing with her. And if Allen blows up the Chrysler Building while Barbara is inside, acting on the maxim, "Destroy the most aesthetic skyscraper in New York," he wrongs Barbara, even though she doesn't feature in his maxim.

So much for the first option. What remains is the second option: that the fact that X wrongs Y by φ-ing helps *make* X's maxim non-universalizable. Suppose that Allen wrongs Barbara by falsely promising to repay her loan next week, acting on the maxim "To get a loan, make a lying promise to repay." If the current proposal is correct, the fact that Allen wrongs Barbara by making the lying promise helps make Allen's maxim non-universalizable. But does it?

Well, what *does* make Allen's maxim non-universalizable? This is a vexed question in Kant interpretation. Some interpreters support the *practical interpretation* of universalizability, according to which what makes a maxim non-universalizable is that its agent's purposes would be thwarted if she acted on it in the world in which the maxim is universal law. Other interpreters back the *logical interpretation* of universalizability, according to which what makes a maxim non-

universalizable is that its being willed as universal law is somehow logically impossible or inconceivable. Still other interpreters go in for other interpretations.[92] I will not attempt to settle this dispute. Instead, I will suggest that the current proposal is doomed on any plausible interpretation of universalizability. Some reflection on how the proposal fares on the practical interpretation should suffice to show the difficulties facing it on any interpretation.

On the practical interpretation, Allen's maxim is non-universalizable because Allen couldn't achieve the maxim's purpose of getting a loan by acting on it in a world where everyone did. It's tempting to conclude that this on its own defeats the current proposal—we have found the fact that makes Allen's maxim non-universalizable, and it is not the fact that Allen wrongs Barbara. But this would be too hasty. For all we have yet shown, it could be that the fact that Allen wrongs Barbara makes Allen's maxim non-universalizable *by* making it the case that Allen couldn't achieve the maxim's purpose by acting on it in a world where everyone did (or by making some other fact that case that in turn makes it the case that Allen couldn't do this, etc.). So we must ask: what makes it the case that Allen couldn't achieve the maxim's purpose by acting on it in a world where everyone did? This question is perhaps most naturally answered in causal terms—we want to say that Allen couldn't do this because a long history of everyone making lying promises to secure loans would cause potential lenders to distrust promises, and therefore refrain from making loans on their basis. Given that the explanatory relation that interests us here is non-causal, however, it might be better to say that Allen couldn't achieve his maxim's purpose by acting on it in such a world because his purpose is to get a loan, and in a world where everyone acted on his maxim, no one would do the things that constitute making Allen a loan (such as putting cash in his

[92] For a survey, see Korsgaard, *Creating the Kingdom of Ends*, 77–105. It might be better to think of these interpretations as rival views of what it *is* for a maxim to be (non-)universalizable, rather than of what *makes* a maxim (non-)universalizable, but the distinction makes no difference for our purposes.

hand, making out checks to him, etc.), as a result of his promising to repay. And this, in turn, is so in virtue of the fact that in such a world, none of the social facts and arrangements of particles that constitute Allen's having money in his hand, a check made out to him, etc. would obtain as a causal result of the social and physical facts that constitute Allen's having promised to repay, etc.

We could keep digging deeper into this order of explanation, but I think it is already clear that no matter how deep we go, we will never encounter the fact that Allen wrongs Barbara by making the false promise. This fact simply plays no role in making Allen's maxim non-universalizable. At least on the practical interpretation, then, we should dismiss the proposal that the fact that X wrongs Y by φ-ing helps make X's maxim non-universalizable.

But the trouble this proposal faces on the practical interpretation seems to generalize to any plausible interpretation of universalizability. On the practical interpretation, facts about universalizability obtain in virtue of how things stand in the world where everyone adopts the agent's maxim, and the difficulty for the proposal is that it is hard to see how facts about who the agent wrongs in the actual world could have any bearing on how things stand in the counterfactual world where everyone adopts the agent's maxim. But it seems that on any plausible interpretation, facts about universalizability will obtain in virtue of how things stand in the world where everyone adopts the agent's maxim. So it is hard to see how the proposal could succeed on any interpretation.

It seems safe to conclude, then, that the fact that X wrongs Y by φ-ing does not make it the case that X's maxim in φ-ing is non-universalizable. Nor, as we saw earlier, are the two facts identical. But if the fact that X wrongs Y by φ-ing is neither the fact that X's maxim is non-universalizable nor a fact that can make it the case, then Universal Law Constructivism is

inconsistent with Wronging First. Unless the Kantian has an argument against Wronging First, she will have to abandon Universal Law Constructivism.[93]

4.4 Humanity-Grounded Kantianism

Does Humanity-Grounded Kantianism fare any better with respect to Wronging First? Wronging First says that the fact that X wrongs Y by φ-ing can make X's φ-ing wrong; Humanity-Grounded Kantianism says are only two sorts of facts make X's φ-ing wrong: (1) the fact that X's φ-ing shows disrespect for the value of humanity, and (2) the facts that make this fact the case. For the two to be consistent, then, the fact that X wrongs Y by φ-ing would have to be either (1) the fact that X's φ-ing shows disrespect for the value of humanity or (2) a fact that can make it the case that X's φ-ing shows disrespect for the value of humanity.

These options are structurally identical to those available to the Universal Law Constructivist, and the same argument that ruled out the first option in that case can be deployed to rule out the first option here: There are many choices of X, Y, and φ such that X shows disrespect for humanity by φ-ing without wronging Y by φ-ing, so the two facts cannot be identical. Let's turn, then, to the second option (and return to our example of Allen's lying promise to Barbara). According to Humanity-Grounded Kantianism, it is wrong for Allen to make his lying promise because in so doing, Allen shows disrespect for the value of humanity. And what, in turn, makes

[93] Note that we have not concluded that the Formula of Universal Law must play *no* role in Kantian moral theory—just that it cannot play this explanatory one. Nor have we concluded that Kantians must abandon moral constructivism altogether. For all I have said, it may be that Wronging First is consistent with forms of constructivist Kantianism that do not take the Formula of Universal Law as the construction procedure, such as Korsgaard's view in *The Sources of Normativity* (but see Tarasenko-Struc, "Kantian Constructivism and the Authority of Others," and cf. note 80 above).

it the case that Allen shows disrespect for the value of humanity? Presumably, that Allen shows disrespect for the value of *Barbara's* humanity. And this, in turn, is so because Allen uses Barbara's humanity merely as a means, which he does in virtue of the fact that he uses her as a means to an end in a way that relies on deceiving her about that end.[94]

We could keep going, but this will be enough for our purposes here. The proposal under consideration is that the fact that Allen wrongs Barbara by making the lying promise can be found among the facts we've just canvassed—that is, among the facts that makes it the case that Allen shows disrespect for humanity. At first glance, this proposal seems to fail. But there is a move available to the Kantian, which is to claim that one of the facts in this story *just is* the fact that Allen wrongs Barbara. For instance, the Kantian could endorse

> *Wronging is Disrespecting*: The fact that X wrongs Y by φ-ing just is the fact that in φ-ing, X shows disrespect for the value of Y's humanity.

Alternatively, the Kantian could endorse

> *Wronging is Using*: The fact that X wrongs Y by φ-ing just is the fact that in φ-ing, X uses Y's humanity merely as a means.

Either of these views would allow the Humanity-Grounded Kantian to affirm Wronging First, since they would each place the fact that Allen wrongs Barbara among the facts that make it the case that Allen shows disrespect for humanity, and therefore among the facts that make it the case that Allen acts wrongly. The Kantian could achieve the same result by going even further down the order of explanation, identifying the fact that Allen wrongs Barbara with the fact that Allen uses Barbara as a means in a way that relies on deceiving her (or with an even deeper fact that makes this the case), but the resulting account of wronging would be implausibly specific. Not

---

[94] See, e.g., Herman, *The Practice of Moral Judgment*, 228.

every wrong involves deception. Wronging is Disrespecting and Wronging is Using are thus the only plausible options for reconciling Humanity-Grounded Kantianism with Wronging First.

Wronging is Using is perhaps the more natural Kantian view, since wronging is associated with violations of perfect duty, which are supposed to involve using humanity merely as a means. And it's a plausible enough account of wronging in the abstract. Although much more would have to be said on this, the notion of using someone merely as a means arguably helps to make sense of some of the hallmarks of wronging: it makes sense that we should owe apology or restitution to those whom we have used merely as a means, and that they should be uniquely situated to forgive us for thus using them. On the other hand, we might worry whether all cases of wronging someone involve using them merely as a means. Consider a case of low-cost, high-stakes rescue. A toddler has fallen into the fountain and will drown, but I can easily pull her out. Plausibly, it isn't just wrong for me to fail to rescue the child—I would wrong her by ignoring her plight.[95] But if I fail to rescue her, have I used her (or her humanity) merely as a means? Not in any remotely ordinary sense, and not obviously in any specifically Kantian sense.[96] Indeed, it is supposed to be only the perfect duties whose violation involves using humanity merely as a means, yet even life-or-death cases of rescue like this one are supposed to fall under an imperfect duty of beneficence.[97] If the Kantian endorses Wronging is Using, then, it seems she must deny that I wrong the child by failing to save her.

---

[95] Cf. Wallace, *The Moral Nexus*, 209.
[96] For worries about a special Kantian sense of this phrase, see Parfit, *On What Matters*, vol. 1, 226–8. Though I will not press the point further here, it's not obvious that I have even, e.g., treated the child in a way to which she could not possibly consent, except on hopelessly broad interpretations of that phrase. See Korsgaard, *Creating the Kingdom of Ends*, 295–6, and cf. Parfit, *On What Matters*, vol. 1, 177ff.
[97] See Herman, *The Practice of Moral Judgment*, 65.

The Kantian can avoid this result by endorsing Wronging is Disrespecting instead, since I presumably show disrespect for the value of the child's humanity if I ignore her. And similar motivation can be given for Wronging is Disrespecting as we gave for Wronging is Using: it makes sense that we should owe apology or restitution to those whose humanity we have disrespected, that they should be uniquely situated to forgive us for thus disrespecting them, etc. On the other hand, we might worry that Wronging is Disrespecting overgenerates cases of wronging, in two ways. First, it seems to entail that we can wrong others just by adopting maxims. Suppose I adopt a maxim of targeted non-beneficence: I resolve never to assist *you in particular*. In adopting this maxim, I presumably show disrespect for your humanity—I fail to treat humanity in you as an end in itself—and so wrong you by the lights of Wronging is Disrespecting. But it's not obvious that this is the right result, especially since the acts of beneficence I am resolving to omit may well be ones that I would not wrong you by *actually omitting*. I don't generally wrong you if I (say) fail to offer you a ride to our shared destination; why should I wrong you by merely *intending* to forgo similar acts of beneficence? Second, Wronging is Disrespecting seems to entail that we can wrong ourselves. Suppose I adopt a maxim of neglecting my talents. In adopting this maxim, I show disrespect for my own humanity by failing to treat it as an end, and thus (according to Wronging is Disrespecting) wrong myself. But it's not clear that one can wrong oneself at all (let alone by adopting a maxim).[98]

The Kantian might respond that I have mischaracterized her account of imperfect duties. When I adopt a maxim of non-beneficence or self-neglect, she might say, I disrespect the value of humanity without disrespecting any particular person's humanity. Whatever plausibility this might

[98] For an excellent overview of the debate over duties to self, see Daniel Muñoz, "The Paradox of Duties to Oneself," *Australasian Journal of Philosophy* 98, no. 4 (2020): 2–6.

have as a claim about maxims of general non-beneficence, however, it doesn't seem plausible as a claim about the targeted maxims discussed above. If there is ever such a thing as failing to treat a particular person's humanity as an end in itself, surely it is my humanity that I fail to treat as an end in itself by adopting a maxim of never improving myself, and your humanity that I fail to treat as an end in itself by adopting a maxim of never helping you. And if we accept this, it seems the consistent thing to do is to apply the same reasoning back to the case of general non-beneficence: in adopting a maxim of never helping *anyone*, I disrespect the humanity in *everyone* rather than in no one. Far from being misplaced, then, our worry about targeted non-beneficence actually extends to maxims of general non-beneficence: if Wronging is Disrespecting were true, then we would wrong everyone if we adopted such a maxim. To generalize further, if it turns out that the only way of disrespecting humanity is to disrespect *someone's* humanity, Wronging is Disrespecting will entail that all wrong acts wrong someone.

So far we have discussed the implications of Wronging is Using and Wronging is Disrespecting separately, but the two views also have some noteworthy implications in common. Consider again the case of the lying promise. Allen uses Barbara merely as a means because he deceives her. But some Kantians would say that Allen also uses *every truthful promisor* merely as a means, since his lie only works by exploiting a practice of promising that is kept afloat by every truthful promisor.[99] This claim is plausible enough on its own, but combined with either Wronging is Using or Wronging is Disrespecting, it implies that when Allen makes his lying promise to Barbara, he wrongs not only Barbara, but also everyone who does her part to support the promising practice. Of course, the Kantian can point to an additional way in which the promisee is used as a mere means, and therefore wronged, that goes beyond the wrong to all the contributors to the

[99] See Korsgaard, *Creating the Kingdom of Ends*, 127.

practice.[100] But this will still strike some as an implausible proliferation of wronged parties, especially since the reasoning in this case presumably generalizes to other cases of free-riding.

We have just seen that Wronging is Using and Wronging is Disrespecting both have some surprising implications. But none of these implications is necessarily damning. Although each of these claims is controversial, the Kantian wouldn't be the first to hold that emergency rescue is obligatory but not owed to its beneficiary,[101] or that we can wrong ourselves,[102] or that all wrong acts wrong someone,[103] or that free riding wrongs all the participants in the practice.[104] Wronging by non-beneficent intention seems a harder pill to swallow, but perhaps not for the Kantian. If you really take the value of humanity seriously, she might say, you'll understand how a policy of never helping is disrespectful, and thus wrongs, while certain sorts of failure to actually help are not disrespectful, and thus do not wrong. Indeed, the Kantian might welcome a view on which we can wrong others solely in virtue of how we regard them. If we can disrespect someone's humanity by having certain beliefs about her, for example, then Wronging is Disrespecting can account for doxastic wronging.[105]

I mention these implications, then, not to counterexample Wronging is Using or Wronging is Disrespecting, but merely to show what is at stake in the choice between them. Kantians who are averse to any of these implications might take that as reason not to adopt the view that implies

---

[100] The Kantian would thus end up with a view on lying promises that is analogous to Kolodny and Wallace's view on promise-*breaking*, according to which promise-breaking wrongs all the contributors to the promising practice but also wrongs the promisee in an additional way. See Kolodny and Wallace, "Promises and Practices Revisited."
[101] See, e.g., Foot, "Euthanasia," 101–2 (and cf. p. 97), and Thomson, "A Defense of Abortion," 61. For the opposing view, see, e.g., Wallace, *The Moral Nexus*, 209.
[102] See note 98 above.
[103] See Wallace, *The Moral Nexus*; among the many opponents of this view is T. M. Scanlon, "Reply to Leif Wenar," *Journal of Moral Philosophy* 10, no. 4 (January 2013): 405.
[104] See note 100 above. For the contrary position (at least in the case of promise-breaking), see Scanlon, *What We Owe to Each Other*, 316.
[105] On doxastic wronging, see, e.g., Mark Schroeder, "When Beliefs Wrong," *Philosophical Topics* 46, no.1 (Spring 2018): 115–26; and Rima Basu, "What We Epistemically Owe to Each Other," *Philosophical Studies* 176, no. 4 (April 2019): 915–31.

them. But more open-minded Kantians might be pleased to find themselves forced into an account of wronging that settles so many controversial questions in a theoretically satisfying way.

It is worth pausing to remember how we got here. Our question was how Humanity-Grounded Kantianism might be made consistent with Wronging First, and the answer was that it must be paired with either Wronging is Using or Wronging is Disrespecting. We then examined these two views, and found that either of them would leave the Kantian with some difficult questions and surprising commitments. These questions and commitments are ones that come with Wronging is Using and Wronging is Disrespecting, but it was Wronging First that forced the Kantian to adopt one of these views in the first place.

It was also Wronging First that ruled out alternatives to these views. Faced with the choice between Wronging is Using and Wronging is Disrespecting, the Kantian might wonder whether a third option is available that splits the difference—perhaps, e.g., a disjunctive account that equates wronging someone with using her merely as a means *or* disrespecting her humanity through particularly egregious inaction. But Wronging First constrains the Kantian's options: it says that the Kantian must identify wronging with a property that she takes to be wrong-making. And the disjunctive property of *using someone merely as a means or disrespecting her humanity through egregious inaction* isn't a wrong-maker on the Kantian's account—it may be a property that many wrong acts have, but it plays no role in the Kantian's story about why wrong acts are wrong. I argued earlier that the only wrong-makers in Humanity-Grounded Kantianism that could plausibly be equated with wronging someone are (1) using her merely as a means and (2) disrespecting her humanity. Unless that argument was mistaken, Wronging is Using and Wronging is Disrespecting are the Humanity-Grounded Kantian's only options for accommodating Wronging First.

4.5 Conclusion

One question of this chapter was whether Kantians can accommodate Wronging First. We have now seen that the answer to this question is a qualified "yes." Kantians who ground right and wrong in respect for humanity, for instance, can accommodate Wronging First by equating wronging someone with disrespecting her humanity, or with using her merely as a means. There are likely other Kantian accounts of right and wrong, not explored here, that can also accommodate Wronging First. But there are also Kantian views that can't. There is, for instance, no room for wronging in a Kantian theory that grounds right and wrong in the universalizability of maxims.

But this chapter was never meant to be the final word on where wronging fits into Kantian ethics. Rather, it is meant to be the first word in a larger conversation about the place of wronging in moral theory more generally. Wronging First requires moral theories to make room for facts about wronging among their wrong-makers. As we've now seen, this requires a moral theory to offer its own account of what it is to wrong someone, one that identifies facts about wronging with facts that the theory takes to be wrong-making. This has not traditionally been viewed as a task for moral theories, most of which have had nothing to say about wronging. Wronging First thus demands a paradigm shift in moral theory, away from theories of monadic right and wrong that are silent on wronging and toward theories that illuminate wronging and its relation to wrongness. With Humanity-Grounded Kantianism, we saw how a theory of the old sort can be adapted into a theory of the new sort; with Universal Law Constructivism, we saw how some theories of the old sort will defy such adaptation. My hope is that these examples will prove instructive to moral theorists of all stripes as they try to adapt their own theories, or even develop new ones, to put wronging first.

Bibliography

Anscombe, G. E. M. “On the Source of the Authority of the State.” In *Ethics, Religion, and Politics*, 130–155. Oxford: Basil Blackwell, 1981.

Arpaly, Nomy. “Moral Worth.” *Journal of Philosophy* 99, no. 5 (May 2002), 223–245.

Ashford, Elizabeth. “The Demandingness of Scanlon’s Contractualism.” *Ethics* 113, no. 2 (January 2003), 273–302.

Basu, Rima. “What We Epistemically Owe to Each Other.” *Philosophical Studies* 176, no. 4 (April 2019), 915–931.

Berker, Selim. “The Unity of Grounding.” *Mind* 127, no. 507 (July 2018), 729–777.

Brandt, Richard. *A Theory of the Good and the Right*. Oxford: Clarendon Press, 1979.

Chappell, Richard Yetter. “The Right Wrong-Makers.” *Philosophy and Phenomenological Research* 103, no. 2 (September 2021), 426–440.

Darwall, Stephen. “Demystifying Promises.” In *Promises and Agreements: Philosophical Essays*, edited by Hanoch Sheinman, 255–276. Oxford: Oxford University Press, 2011.

Darwall, Stephen. *Morality, Authority, and Law*. Oxford: Oxford University Press, 2013.

Duarte d’Almeida, Luís. “Fundamental Legal Concepts: The Hohfeldian Framework.” *Philosophy Compass* 11, no. 10 (October 2016), 554–569.

Feinberg, Joel. “The Nature and Value of Rights.” *Journal of Value Inquiry* 4, no. 4 (December 1970), 243–257.

Feinberg, Joel. “Voluntary Euthanasia and the Inalienable Right to Life.” *Philosophy & Public Affairs* 7, no. 2 (Winter 1978), 93–123.

Fine, Kit. “Guide to Ground.” In *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press, 2012.

Frick, Johann. “Contractualism and Social Risk.” *Philosophy & Public Affairs* 43, no. 3 (Summer 2015), 175–223.

Fried, Barbara. “Can Contractualism Save Us from Aggregation?” *Journal of Ethics* 16, no. 1 (March 2012), 39–66.

Foot, Philippa. “Euthanasia.” *Philosophy & Public Affairs* 6, no. 2 (Winter 1977), 85–112.

Gilbert, Margaret. *Rights and Demands*. Oxford: Oxford University Press, 2018.

Hart, H. L. A. "Are There Any Natural Rights?" *Philosophical Review* 64, no. 2 (April 1955), 175–191.

Herman, Barbara. *The Practice of Moral Judgment*. Cambridge: Harvard University Press, 1993.

Hieronymi, Pamela. "Of Metaethics and Motivation: The Appeal of Contractualism." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 101–128. Oxford: Oxford University Press, 2011.

Hohfeld, Wesley Newcomb. *Fundamental Legal Conceptions*. New Haven: Yale University Press, 1923.

Hooker, Brad. *Ideal Code, Real World*. Oxford: Oxford University Press, 2000.

Kagan, Shelly. *Normative Ethics*. Boulder: Westview Press, 1998.

Kamm, F. M. *Intricate Ethics*. Oxford: Oxford University Press, 2007.

Kamm, F. M. "Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status." *Philosophy and Public Affairs* 21, no. 4 (Autumn 1992), 354–389.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor and Jens Timmerman. Cambridge: Cambridge University Press, 2012.

Kolodny, Niko and R. Jay Wallace. "Promises and Practices Revisited." *Philosophy & Public Affairs* 31, no. 2 (Spring 2003), 119–154.

Korsgaard, Christine. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

Kumar, Rahul. "Defending the Moral Moderate." *Philosophy & Public Affairs* 28, no. 4 (Autumn 1999), 275–309.

Kumar, Rahul. "Risking and Wronging." *Philosophy & Public Affairs* 43, no. 1 (Winter 2015), 27–51.

Markovits, Julia. "Acting for the Right Reasons." *Philosophical Review* 119, no. 2 (April 2010), 201–242.

May, Simon Căbulea. "Directed Duties." *Philosophy Compass* 10, no. 8 (August 2015), 523–532.

Muñoz, Daniel. "From Rights to Prerogatives." *Philosophy and Phenomenological Research* 102, no. 3 (May 2021), 608–623.

Muñoz, Daniel. “The Paradox of Duties to Oneself,” *Australasian Journal of Philosophy* 98, no. 4 (2020), 691–702.

Murphy, Liam. “Nonlegislative Justification.” In *Principles and Persons: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, 247–276. Oxford: Oxford University Press, 2021.

Parfit, Derek. *On What Matters, Volume One*. Oxford: Oxford University Press, 2011.

Parfit, Derek. *On What Matters, Volume Three*. Oxford: Oxford University Press, 2016.

Perl, Caleb. “Solving the Ideal Worlds Problem.” *Ethics* 132, no. 1 (October 2021), 89–126.

Poddar, Prashant and Vijaya Singh. “When Left Is ‘Right’! The Impact of Driving-Side Practice on Road Fatalities in Africa.” *Transport Policy* 114 (December 2021), 225–232.

Podgorski, Abelard. “Wouldn’t It Be Nice? Moral Rules and Distant Worlds.” *Noûs* 52, no. 2 (June 2018), 279–294.

Puri, Sunita. “My Patient Didn’t Die From Covid. He Died Because of It.” *The New York Times*, May 21, 2022, https://www.nytimes.com/2022/05/21/opinion/covid-deaths-million.html.

Railton, Peter. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13, no. 2 (Spring 1984), 134–171.

Rawls, John. *Lectures on the History of Moral Philosophy*. Edited by Barbara Herman. Cambridge: Harvard University Press, 2000.

Rawls, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1971.

Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.

Reath, Andrews. *Agency and Autonomy in Kant’s Moral Theory*. Oxford: Clarendon Press, 2006.

Reibetanz Moreau, Sophia. “Contractualism and Aggregation.” *Ethics* 108, no. 2 (January 1998), 296–311.

Ridge, Michael. “Introducing Variable-Rate Rule-Utilitarianism.” *Philosophical Quarterly* 56, no. 223 (April 2006), 242–253.

Rosen, Gideon. “Metaphysical Dependence: Grounding and Reduction.” In *Modality: Metaphysics, Logic, and Epistemology*, edited by Ben Hale and Aviv Hoffman, 109–136. Oxford: Oxford University Press, 2010.

Rosen, Gideon. “Might Kantian Contractualism Be the Supreme Principle of Morality?” *Ratio* 22, no. 1 (March 2009), 78–97.

Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press, 1930.

Scanlon, T. M. *Being Realistic About Reasons*. Oxford: Oxford University Press, 2014.

Scanlon, T. M. "Contractualism and Utilitarianism." In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams, 103–128. Cambridge: Cambridge University Press, 1982.

Scanlon, T. M. "Reply to Leif Wenar." *Journal of Moral Philosophy* 10, no. 4 (January 2013), 400–405.

Scanlon, T. M. *What We Owe to Each Other*. Cambridge: Harvard University Press, 1998.

Scanlon, T. M. "Wrongness and Reasons: A Re-Examination." In *Oxford Studies in Metaethics Volume Two*, edited by Russ Shafer-Landau, 5–20. Oxford: Oxford University Press, 2007.

Schaffer, Jonathan. "Grounding, Transitivity, and Contrastivity." In *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schnieder, 122–138. Cambridge: Cambridge University Press, 2012.

Schroeder, Mark. "When Beliefs Wrong." *Philosophical Topics* 46, no.1 (Spring 2018), 115–26.

Stratton-Lake, Philip. "Recalcitrant Pluralism." *Ratio* 24, no. 4 (December 2011), 364–383.

Suikkanen, Jussi. *Contractualism*. Cambridge: Cambridge University Press, 2020.

Tarasenko-Struc, Aleksy. "Kantian Constructivism and the Authority of Others." *European Journal of Philosophy* 28, no. 1 (March 2020), 77–92.

Thompson, Michael. "What is It to Wrong Someone? A Puzzle About Justice." In *Reasons and Values: Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 333–384. Oxford: Clarendon Press, 2004.

Thomson, Judith Jarvis. "A Defense of Abortion." *Philosophy & Public Affairs* 1, no. 1 (Autumn 1971), 47–66.

Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge: Harvard University Press, 1990.

Wallace, R. Jay. *The Moral Nexus*. Princeton: Princeton University Press, 2019.

Wood, Allen. "Humanity as End in Itself." In Derek Parfit, *On What Matters, Volume Two*, edited by Samuel Scheffler, 58–82. Oxford: Oxford University Press, 2011.

Wood, Allen. *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.

Zylberman, Ariel. "Relational Primitivism." *Philosophy and Phenomenological Research* 102, no. 2 (March 2021), 401–422.