# What We Owe to Ourselves:
# Essays on Rights and Supererogation

by

Daniel Muñoz

B.A., The University of Texas at Austin (2014)

Submitted to the Department of Linguistics and Philosophy
on May 13, 2019 in Partial Fulfillment of the Requirements
for the Degree of

Doctor in Philosophy in Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.



Signature of Author _____
Department of Linguistics and Philosophy
May 13, 2019

Certified by _____
Tamar Schapiro
Associate Professor of Philosophy
Dissertation Co-Chair

Certified by _____
Kieran Setiya
Professor of Philosophy
Dissertation Co-Chair

Accepted by _____
Bradford Skow
Laurance S. Rockefeller Professor of Philosophy
Chair of Committee on Graduate Studies

**WHAT WE OWE TO OURSELVES:**

**ESSAYS ON RIGHTS AND SUPEREROGATION**

by

Daniel Muñoz

Submitted to the Department of Linguistics and Philosophy on May 13, 2019 in Partial Fulfillment of the Requirements for the Degree of Doctor in Philosophy in Philosophy.

ABSTRACT

Some sacrifices—like giving a kidney or heroically dashing into a burning building—are supererogatory: they are good deeds beyond the call of duty. But if such deeds are really so good, philosophers ask, why shouldn't morality just require them? The standard answer is that morality recognizes a special role for the pursuit of self-interest, so that everyone may treat themselves as if they were uniquely important. This idea, however, cannot be reconciled with the compelling picture of morality as impartial—the view that we are each anyone's equal. I propose an alternative Self-Other Symmetric account of our moral freedom: the limits on what morality may demand of us are set by the duties we owe to ourselves. I begin with a defense of the Self-Other Symmetry: the idea that we owe the same duties to ourselves—and have the same rights against ourselves—as any relevantly similar other. Because we are consenting parties to our own actions, I argue, our rights against ourselves do not function like the rights of unwilling others. Instead, they play a permissive function, allowing us to block the demand to give up what is ours. I conclude by uniting, aggravating, and trying to solve some paradoxes of supererogatory permissions, guided by the idea that morality cannot be reduced to a ranking of options from best-to-worst. Rights against oneself are an irreducible second dimension, one that we need if we are to unify rights and supererogation into an impartial moral vision.

DISSERTATION COMMITTEE

Tamar Schapiro (Co-Chair), Associate Professor of Philosophy
Kieran Setiya (Co-Chair), Professor of Philosophy
Caspar Hare, Professor of Philosophy

By way of introduction it is to be noted that there is no question in moral philosophy which has received more defective treatment than that of the individual's duty towards himself.

—Immanuel Kant, *Lectures in Ethics* (117)

I am sure that when we look more closely at claims we will not find ourselves wanting to say that person has claims against himself or herself.

—Judith Jarvis Thomson, *The Realm of Rights* (p. 42)

Morality encourages the idea, *only an obligation can beat an obligation*…and in order to do what I wanted to do, I shall need one of those fraudulent items, a duty to myself.

—Bernard Williams, *Ethics and the Limits of Philosophy* (pp. 180–82)

I am to choose which of two possible outcomes is to be realized: in the one, *B* retains his arm intact and *C* dies; in the other, *B* loses his arm and *C* does not die. If the choice were *B*'s it would be permissible for him to choose the first outcome. But it is not permissible for me to make this same choice? Why exactly is this?

—John M. Taurek, "Should the Numbers Count?" (p. 302)

*For David, my brother*

# Table of Contents

**Acknowledgements**

If you are reading this, odds are decent that you are on my dissertation committee. So why don't I thank you first? In alphabetical order: thanks to Caspar Hare, for his willingness to follow and further the argument; to Tamar Schapiro, for seeing me through those long months when the ideas were on top of me—and helping turn the tables; and to Kieran Setiya, for his unfailing judgment and dedication. Having Kieran and Tamar as co-chairs has been the highlight of my (new) career.

Very special thanks to Nathaniel Baron-Schmitt, who helped me come up with the main idea of "Why Isn't Supererogation Wrong?" in January 2015. I will always be grateful for those late-night/early-morning conversations, especially in our first year of graduate school.

Next, more faculty. Those who most helped with this project were Jack Spencer, Brad Skow, Justin Khoo, Agustín Rayo, Selim Berker, Sally Haslanger, Dan Bonevac, Jonathan Dancy, Judy Thomson, and Stephen Yablo (whom I cannot resist singling out a final time—thank you, Steve).

To my fellow MIT graduate students: I cherish you all equally. Especially Thomas Byrne, Quinn White, the Davids (Balcarras, Builes, Grant), Lyndal Grant, Kevin Dorst, Kelly Gaus, Sophie Gibert, Sam Hesni, Anni Räty, Nicole Garcia, Ginger Schultheis, Eliot Watson, and Ryan Ravanpak. (Or rather, they are some of the ones who helped most recently.)

To my friends: Amy Veggeberg, Travis Smith, Mirjam Müller, Kari Rosenfeld, Nick Geiser, Shosh Williams, Sam Dishaw, Emily Tagtow, Ryan Westphal, Becca Rothfeld, Jon and Emily Drake, Leonard Katz, Blair Johnson, Thaisa Howorth, Sara Gottesman, Zoe Jenkin, Keshav Singh, Becca Besaw, Monica Thieu, Katy Meadows, Elliot Goodine, Carolyn Saund, Caitlin Fitchett, Zelda Mariet, Eric Stansifer, Kirun Sankaran, Joe Bowen, Cheryl Abbate, Isabel Ottlewski, Phil Bernard, Joe and Pat Crowley, Bert Xue, Nikki Pitt, Yael Loewenstein, Erik Shoener, and the staff at Flour and Vialé.

Finally, family: cousins, aunts, uncles; the Masutanis, Wadas, and Okadas; Conni and Wynn, my grandparents. Above all: my parents, Shanan and Homero, and brother, David, for their love.

## Preface

What do we owe to ourselves?

The standard answer, at least among modern moral philosophers, appears to be: not much. Morality is about what we owe to each other—how we hold one another accountable, litigate claims, cast aspersions, mete out punishment, promote the common good. Obligations to oneself are a relic at best (from a time when our bodies were someone else's "temples"), and at worst, pure paradox.

I argue that the standard line is missing something fundamental in the structure of our moral thinking, where duties to oneself serve as the irreplaceable ground of our moral freedom.

Consider an example. You have a spare kidney that properly belongs to you, and which you would quite like to continue using, but a stranger needs it more. Do you have to give the kidney? Intuitively, you do not, though it would be a fabulous thing if you did—the gift would be "supererogatory," lying splendidly beyond the call of duty. But many philosophers feel powerfully drawn towards the idea that, really, you *do* have the give the kidney. You aren't objectively more important than the stranger, after all, so why should you get to elevate your interests over theirs? Tough question. But change the case a bit. What if *I* were the one making the choice? Would I be permitted to harvest your kidney against your will? Easy: no. The kidney isn't *mine*, after all—it is yours, not just in the sense that it is within your skin, but in the sense that it is yours *by right*, and so you ought to have some say over what is done to it. But then shouldn't you also have a say when *you* are the one making the choice? How could it be wrong to keep the kidney given that it's yours?

This, in my view, is where obligations to oneself are essential. They are the flipside of interpersonal duties. Just as others' duties to us block them from overriding our wills, our duties to ourselves give us the power to *block moral demands*, to justify using our things and bodies how we like, within reasonable limits, even if we are not doing the greatest good. What we owe to others is limited because we owe the same things to ourselves.

That is the core idea of this dissertation: duties and rights are *Self-Other Symmetric*, and duties to oneself have the function of explaining why supererogatory sacrifices are optional instead of morally mandatory. Just as you have a right that others not take your kidney out, you have a right against *yourself* that *you* not take it out. The crucial twist: this right doesn't make it *wrong* or *bad* for you to give your kidney up. Instead, the right is purely permissive, allowing you to justify the choice to keep your kidney without at all detracting from the value of giving it to those in greater need.

The first papers develop the idea of duties to, and rights against, oneself. (I don't make a distinction between X's duty to Y and Y's right against X.) "The Paradox of Duties to Oneself" takes on some conceptual objections; "Rights Against Oneself" argues that the Self-Other Symmetry is a surprisingly snug fit for our intuitions about when we wrong ourselves, and when others may intervene in our self-regarding choices without wronging us. Here the central idea is that the rights we have against ourselves are nowhere near as restrictive as others' rights because we typically are willing parties to our own actions. Our rights against ourselves don't constrain us because, when we decide to do what they forbid, we waive them. Rights against oneself are thus like the rights of consenting others. (And we wrong ourselves when we do what would be a wrong even to a consenting other—grotesque debasement, egregious harm.)

But while rights against oneself aren't generally restrictive, they do have a valuable purpose: they can be used to justify doing less than best. That is the central argument in "From Rights to Prerogatives." I argue that *if* we have rights against oneself that are properly Self-Other Symmetric—forbidding the same acts and following the same principles as rights against others—then they serve as prerogatives instead of restrictions. A *prerogative* is something that tends to justify an action, making it permissible, without at all tending to make it required. Just because I have the prerogative to keep my kidney, for example, that doesn't make donating any worse, and it certainly doesn't make it morally wrong.

Why is a waivable right against oneself, like my right that I not take out my kidney, going to count as a prerogative? This question has two parts. First, why doesn't the right require compliance? Why isn't it just a moral reason? Here my answer is that we can waive our own rights at will, simply by choosing to do what they forbid, so they don't constrain us; a genuine reason doesn't evaporate when you act against it. If I choose to give my kidney, I violate no rights, because I have made myself into a consenting donor. But here's question two: why aren't I just *required* to make the gift? If the right against myself doesn't matter in the sense of being a reason, how could it block the demand to do what's best? Here my answer is that, so long as I don't actually waive the right, it is still available to cite in the course of moral justification, as when I insist, "it's *my* kidney, so I don't have to give it." So the right is a prerogative: it justifies without counting in favor or tending to require.

Now, admittedly, when I say "it's my kidney," it might not sound like I am leaning on a waivable right held against myself. But why shouldn't it? It is unfamiliar to think in terms of rights against ourselves. But if we think them through, drawing out their implications, we don't end up in unfamiliar moral territory. We end up with a natural account of supererogation: why it's valuable, permissible, and omissible.

The omissibility of supererogation—the permissibility of *not* doing it—is a simple fact, but strangely enough, the standard views struggle to account for it. For example, one classic story is that you may keep your kidney because your self-interest has extra weight *for you*. But then what's so good about sacrificing for the impartial good? Why isn't that flatly *wrong?*

I pursue this question in "Why Isn't Supererogation Wrong?" where I also develop counterexamples to any view that bases supererogatory permissions in self-interest. To explain the permissive character of prerogatives, and to explain why they protect what they protect, we must look beyond mere interest to the realm of rights, deriving prerogatives from waivable rights against oneself. (This is an argument "from prerogatives back to rights!") The result is a punchy account of

moral right and wrong: we must do whatever is best *unless* we have a right not to—i.e. a prerogative, consisting in a waivable duty to oneself.

This account of wrongness is easy to grasp, but in a way complex: it is irreducibly two-dimensional. On the one hand, there is the ranking of acts from bad to good—the ranking of which acts we have most (moral) reason to do—and on the other, there is the question of whether we have a countervailing prerogative to choose suboptimally, a right to go against the balance of reasons. The supererogatory act is better than other permissible options, but we have a prerogative not to do it.

In "Better to Do Wrong," I apply this two-dimensional view to a recent puzzle about supererogation and gratuitously bad acts—e.g. rushing into a building only to save one baby rather than two. (We have a prerogative not to go into the building; we have no prerogative to save one child while leaving the second.) In "Supererogation and Rational Choice: Intransitivity, Incommensurability, Independence," I draw out the links between this case and a familiar paradox involving intransitivity, before introducing a puzzle of my own, with the aim of pushing prerogatives to their limits (but no further!).

This thesis, then, is one line of thought, running from rights to prerogatives to the paradoxes of supererogation. I hope it is reasonably fun to read, with a mix of standalone puzzles (the alleged MIT staple) and full-on *Weltanshauung* (a scarce import?). I have tried my best to keep the main ideas totally exposed, so that you won't have to slash through a hedge maze just to get to them. No paper presupposes familiarity with any other.

That said, I recognize that I am asking a lot of you, the reader. Rights against oneself are widely seen as paradoxical—and here I am, throwing them at you in droves.

Let me close this preface with a high-level hunch about *why* people's kneejerk reactions are so hostile to rights against oneself. Skeptics about rights against oneself come in two kinds. (1) Some think these rights would be *reasons for action*—constraining us like the rights of an unwilling other,

with ridiculous results if rights are Self-Other Symmetric. Such rights would make it wrong, rather than supererogatory, to donate one's own kidney. (2) Others doubt that there could be rights against oneself precisely because they *could not* be reasons for action. As the holder of the right, the agent has the power to waive it. But a right that you can waive at will cannot give you any reason to comply, any more than you have reason to follow a law that you could repeal with the wave of a hand.

Both skeptics agree, however, that rights against oneself could only matter in one special way: as normative reasons. I reject this fundamental assumption. I think waivable rights against oneself are pure prerogatives, justifying without "counting in favor" like typical moral reasons.

Indeed, waivable rights are vital for ethics precisely because they aren't just reasons. Morality is like the law: a social institution to help us live in peace. Any analysis of the law that *only focused on agents' reasons for compliance* would be defective. Even granting that we have reasons to comply with legal duties, what would we say about laws that don't confer duties, like those that lay out the conditions for a valid wills and contracts? Legal powers, which we often have over our own rights (think: the power to trade rights over a bike for rights over some cash), don't bind us; they make us "competent to determine the course of the law within the sphere of" our private lives, to borrow a phrase from H.L.A. Hart (1961: 41).[1] Our powers over our own moral rights, I think, serve a similar function; they let us determine the course of *moral sanctions*, allowing us either to resist blame by leaning on our rights, or to waive them away—donating the kidney, making the sacrifice—and go beyond the call. Rights are constraints under control, and supererogation emerges from the possibility that we might control the constraints on our own choices.

In the end, it's still true that our duties to ourselves put some burdens on us, demanding self-respect and self-care—but they have a hidden aspect. The source of supererogation—the ground of our prerogatives—lies in what we owe to ourselves.

---

[1] *The Concept of Law*. Third Edition. Oxford: Clarendon Press.

**Abbreviations**

When citing other parts of this dissertation, I will give the chapter number along with one of the following abbreviations:

1. "The Paradox of Duties to Oneself" PDO
2. "Rights Against Oneself" RAO
3. "From Rights to Prerogatives" FRP
4. "Why Isn't Supererogation Wrong?" WSW
5. "Better to Do Wrong" BDW
6. "Supererogation and Rational Choice" SRC

# CHAPTER ONE

## The Paradox of Duties to Oneself

## The Paradox of Duties to Oneself

For centuries, philosophers have argued that there is something deeply paradoxical about the very concept of a duty to oneself—and especially the concept of a right against oneself. I review the literature, single out two possible paradoxes, and argue that neither is persuasive. There is nothing conceptually fishy about to duties to, or rights against, oneself.

### 1. Introduction

Morality says we owe each other plenty of things, like minimal care and basic respect. Since others tend to be quite a bit like *us*, you might think we also owe a few things to ourselves. There is indeed something gross about a servile suck-up, and something morally charged about a tragic suicide, even when the agents don't harm or demean anyone else. So you might think we have *duties to ourselves*.

Duties to self are entrenched in the history of ethics and still intuitive today. But they are haunted by a paradox, conjured by Marcus Singer (1959, 1963).[2] From three apparent truisms, we can show that duties to oneself are incoherent. Singer writes:

(1) If A has a duty to B, then B has a right against (or with respect to) A;

(2) if B has a right against A, he can give it up and release A from the obligation; and

(3) no one can release himself from an obligation. (Singer 1959: 203)

Taken one-by-one, these all sound plausible. Taken together, they rule out duties to oneself.

The Paradox of Duties to Oneself is the crucial test of any theory of self-obligation. It is also a sort of crossroads for the theorist. You have to choose: will you sever the link between duties and rights? Between rights and release? Or will you allow for the power to release oneself from duties? Your decision here determines what kind of theory you end up with.

As important as the Paradox is, it has not received much systematic treatment. Partly this is because, since Singer's time, "duties to oneself have largely disappeared from the radar of academic philosophers" (Cholbi 2015: 852). Another factor is that few of Singer's critics develop their

[2] Two historical precedents are Kant and Hobbes; see §3 below.

response in depth or survey other options; usually they are happy just to produce a few examples, content to poke a hole in a premise. (An exception is Paul Schofield, whose view I discuss in §2.) And even some deep contributions have gone unnoticed, being scattered across literatures.

This paper has two goals: first, to systematically collect and critique the responses to Singer's Paradox (§§2–3); second, to argue for the possibility of self-release—the denial of Singer's third premise (§§3–5). Although this premise is the most widely accepted, Singer says little to support it, other than his claim that it is "essential" to obligations that one can't wiggle out. I consider two ways to develop Singer's claim—one meta-normative, one using deontic logic—and present objections. My conclusion is that the Paradox should only make us skeptical of duties to oneself *if* we think a person has reason to comply with *any* duty, even one from which she might be released (§6).

Before we begin, let me clarify the topic. When we talk about duties *to* people, we are talking about what is *owed* to them, not just what duties are held "regarding" them (Singer 1959: 204).[3] If I promise you to water your snake plant, I have a duty to you, and merely regarding your plant. (A telltale sign: if I forsake my duty, I don't wrong the plant—I wrong you.) Now, just about everyone agrees that we can have duties regarding ourselves. For example, I might promise you that I won't smoke this week, even in private. My new duty is self-regarding—I'm supposed to be doing something with my own body—but not self-directed. I owe it to *you* not to smoke. Our topic is the possibility of owing bona fide duties *to oneself*.

## 2. Duties without release?

How can we get out of the Paradox? One option, of course, is to accept the conclusion—that there are no duties to oneself. But we could also contest a premise. That means three more options: we

[3] I will use "duty" only when talking about duties-to, not moral requirements in general, and I use "duty" and "obligation" interchangeably. For more on duties-to, see Thomson 1990: Chapter 1, Thompson 2004, Darwall 2006, Schofield 2015, and Gilbert 2018.

could (1) deny that duties entail rights, (2) deny that rights entail powers of release, or (3) accept that people can release themselves from duties, at least sometimes. Let's start with the most popular options, (1) and (2).

Singer's first two premises link duties to rights and rights to release. The entailment from duties to rights is, I think, fairly standard. If I have a duty to you not to harm you, you have a "correlative" right against me that I not harm you. If you owe me $5, I have a right to receiving $5 from you. Following Hohfeld 1919, many philosophers take this to be part of the definition of a "right in the strictest sense," also called a "claim right" (e.g. Thomson 1990: Chapter 1*, Johnson 2010). Other writers say that, by definition, the holder of a right can waive it, releasing the subject of the corresponding duty (e.g. Steiner 2013). But Singer's critics don't stick to any single use of "right," and so for us, the term might be more trouble than it's worth. It will be hard to separate those writers who reject (1) from those who reject (2). To avoid verbal disputes, I propose we drop rights-talk for now and focus on the substantive idea behind these two premises, namely this: if X owes Y a duty, it follows that Y can release X from that duty.

How could we break the link between duties and release? There are four main options to choose from.

First and foremost, there is the idea that duties to oneself aren't "juridical" or "legal," like duties to keep promises and refrain from embezzling. Juridical duties are *external* and *enforceable:* they concern our acts rather than our motives, and we can be coerced into complying. There is also a link between juridical duties and release. A wave of your hand can release me from a promise or forgive a debt. A signature on a form can permit the surgeon to cut me open or allow the studio to use my likeness in a movie. There is not supposed to be any such way to release someone from a non-juridical duty: they simply require us to act from good motives, and there is nothing anyone can do to dissolve that requirement. And so, the idea goes, Singer is right that we can't owe things juridically

to ourselves, but still it is possible to owe ourselves things non-juridically. We have duties to ourselves in the sense that we have reason to care about and respect ourselves simply for our own sakes.

This is the most popular response to Singer by a mile.[4] The problem, as I see it, is that it is hard to see why "non-juridical" duties should count as duties at all. The hallmark of duties and rights, as opposed to globs of value and normative reasons, is that they are fit to enforce by legitimate demands and external coercion (Hart 1955: 178). If you have a duty to pay me back, I may demand the cash. If I have a duty not to harm you, you may preempt my punches by hitting first (if that's what it takes). But we can't enforce good behavior in general—brushing one's teeth, donating a spare kidney, etc.—and non-juridical duties, by definition, just ask for good behavior from nice motives. Non-juridical duties aren't *supposed* to be enforceable.[5] They are more like sources of moral advice, telling us to what to value and why. We can call this sort of thing a "duty" if we want, but that is a big concession to Singer, who would just insist that real duties are juridical.

The second response is also concessive. Some writers relax the idea that duties must be owed *to oneself*. Knight (1961: 212), for example, claims that we have duties only "to" ideals. Some

---

[4] Warner Wick (1960: 161–62, 1961) was the first to make this response to Singer. Mary Mothersill (1961: 207) notes that, on Wick's view, "duties to oneself…must be of a completely different order from duties of the familiar sort." See also Knight 1961 on "non-legal" duties, and Fotion's (1970: 460) response to Eisenberg (1968) on "social contractual" duties. Margaret Paton (1990: 225–26) argues that duties to oneself are "non-contractual," so that they don't entail rights or powers of release; she also concedes to Singer that we can't making binding self-promises. Hills (2005: 138) agrees about self-promising and rejects juridical duties to oneself. These authors' main inspiration is Kant; see *MS* 6:383, *LE* 29:117, 632. (I quote from the Cambridge editions and use pagination from the Akademie Edition; *MS* is for *The Metaphysics of Morals* and *LE* is for *Lectures on Ethics*.) But note also that the concept of non-juridical duties was around before Kant. It has roots in Calvin's "precepts of love," which can be fulfilled only by acting from good motives, and in Hugo Grotius's "imperfect rights" and "laws of love," which aren't enforceable; for more on these writers, and for references, see Schneewind 1998: 79–80.

[5] Kant gives an argument here (*MS* 6:219). Non-juridical duties require us to act from the right motive. But we can't be coerced into having nobler motives—when coercion works, it *is* the motive!

neo-Kantians suggest that obligations are owed only among *parts* of people—no whole *person*, strictly speaking, owes any *person* anything, though we might loosely say that I have duties "to myself" when one part is obliged to another.[6] Then there are those like Meiland (1963) and Fotion (1965: 30) who retreat to the claim that we have self-regarding duties: duties that aren't necessarily owed to ourselves, but that do constrain how we treat ourselves; for example, the utilitarian duty to cheer oneself up is owed to no one; and I might owe it to a friend to keep my promise to stop smoking.[7] Sometimes this retreat is hailed as a triumph: Oakley (2017: 71) argues that we can release ourselves from merely self-regarding duties and dubs his paper "Good News for Duties to Oneself."[8]

These authors do manage to defend *something* like a duty to oneself. But again, if this is the best response to the Paradox, I think it's safe to say the Paradox wins.

The margin of victory only grows when we turn to the third response, which is that no one can release *anyone* from obligations (Wick 1960: 162; Denis 2001: 229 fn. 13, 230). This is blatant overkill. People release people from duties all the time. What else could I be up to when I let you out of a promise, or when sign my dentist's consent form?[9]

Finally, the most promising response. There might be *contingent reasons* why we are unable to release ourselves from certain duties, which opens up space for some (contingently unwaivable) duties to oneself.

The first to suggest this view was Daniel Kading, who gives the following case:

> Suppose A promises B to do *x* after B's death. We should certainly want to say that although B, after his death, cannot possibly release A, A does continue to have an obligation to B.

---

[6] "Thus the subject and the object of duties to one's self, though both part of the same human being, turn out not to be identical" (Timmermann 2006: 509). Timmermann (2006: 512) later seems to retract this idea. For more on whether Kantian duties to self require "two selves," see Denis 1997: 334–35, Potter 1998: 375, fn. 8, Reath 2006: 356, Kerstein 2008: 206, Schofield 2015: fn. 24, and Kant *MS* 6:417.

[7] For more on Meiland's view, see Mavrodes, Narveson, and Meiland 1963: 169–70.

[8] Oakley argues that we can act so that we aren't *required* to fulfill a certain duty, e.g. by giving ourselves even better things to do. But that's not what Singer, or anyone else, means by "release."

[9] Mothersill (1961: 206) raises the question about promises in her response to Wick.

(1960: 155)

Not exactly gripping drama, but the point is clear enough. Sometimes you can't be released because the person to whom you are obligated can't do what's necessary to release you. In Kading's case, the source of incapacity is death. But we could just as well imagine promisees who can't release anyone because they are napping, gagged, or off the grid.[10]

Singer (1963) has several responses here.[11] His main objection is that it doesn't matter if we are contingently unable to release ourselves. Unless we are *necessarily* unable, he thinks, there will still be a contradiction lurking:

> The proposition that, if B has a right against A, he can give it up and release A from his obligation, does not imply that B must always and in all circumstances retain the capacity or have the opportunity of releasing A. It only implies that it is not self-contradictory to speak of his doing so. And this is why the fact of B's death would be irrelevant to the situation. (1961: 134)

Here, Singer seems to be saying: anyone with a right *could* waive it in *some* coherent scenario—though perhaps not *any* scenario. This is a revision of (2). I don't quite know what to make of it.

But there is a simpler problem for Kading, which is that his cases don't have anything to do with the traditional lists of duties to oneself. Even if they show that we can't be released from *some* duties, there is no reason to expect that this will allow for the duties we intuitively owe to ourselves. It is not as if we can, post-mortem, break promises to our past dead selves.[12] Nor is this the sort of

---

[10] Kading (1960: 155) has two other cases where promisees lack powers of release—either because they have also promised us not to release us, or because they have transferred their rights of release to a third party. But as I see these cases, they aren't meant to show anything beyond what Kading's first case shows already. They are just more examples where the agent is obligated to someone who, for contingent reasons, doesn't have normal powers of release. (Kading's (1960: 356) fourth case is meant to show that I might have duties merely regarding someone who doesn't have the power to release me.)

[11] Most notably: (1) we would not say that the promisee has a *right* even after dying; (2) someone *else* might have the power to release the promisor after the promisee dies; and (3) the promisee has the power to release *before* his or her death (Singer 1961: 133–34). The first point doesn't seem persuasive (why couldn't Kading say that the dead have rights?), and the second two are irrelevant. For Singer's remarks on Kading's other cases, see Singer 1961: 133–35, Hill 1991.

[12] What about Kading's case where I promise someone who promises not to release me? (See

thing people have in mind when they say that we owe it to ourselves to take care of our bodies and carry ourselves with dignity.

There are some other authors who believe in unwaivable duties to oneself, and they tend to have more developed positive view than Kading (Hills 2003, Timmermann 2006: 516). But only one has a worked-out theory, and that is Paul Schofield (2015, 2018, forthcoming), who draws on ideas from Stephen Darwall (2006).

Schofield grants that waivable duties to oneself are impossible. But he thinks we might owe ourselves some unwaivable duties because we can't release ourselves when our acts affect our lives in the *future*.[13] To waive a duty, we must take up the right *perspective* (a "point of view" from which we act, desire, and perceive). But—and this is the key point—we can't take up our future perspective *now*, and so we can't release ourselves.

> A person might owe a duty to herself by virtue of occupying a perspective some time in the future to which interests and ends attach. And since she cannot, in the present, get a release from the future perspective that generated the duty—because the perspective is in the future, of course—it seems that the duty *could* actually bind in the way characteristic of genuine moral duties after all. (Schofield forthcoming, emphasis original)

The result is that we must treat our future selves, in a way, like unwilling others.[14] If we smoke for pleasure now, we impose on ourselves a hefty future cost. If we rack up debt, we hamstring our future choices. These acts are wrong, and they wrong *us* ourselves, and this is all possible because we can't take up our future perspectives to secure release.

Schofield's view of self-obligation is original and principled. There is a lot to like about it. But in its current form, the view suffers from counterexamples, in which an agent hurts her future

---

fn. 10, above.) That might allow for a very limited range of duties to self: viz. promises to oneself that include the promise *not* to release oneself. (Singer (1963: 134) gives an argument against the possibility of such promises; Hill (1991) responds.) But that would hardly be enough.

[13] Schofield forthcoming extends the account to synchronic duties.

[14] But note that Schofield is not committed to any view about future "selves" as distinct agents. He assumes the opposite: that people endure over time (2015: 13).

self for her own greater good. Suppose I allow the dentist to sedate me and take out my crooked wisdom teeth. For a while, it will be unpleasant: I'll be immobile in the chair, unable to do anything, groggy and uncomfortable. But the procedure is perfectly safe, and it will spare me even worse pains in the further future. Clearly, I'm not doing anything wrong here. But Schofield's view seems to entail that I am. After all, when I agree to the operation, I can't release myself from the duty *not to have my teeth damaged in the future* (even for my own greater good). Indeed, it is unclear why I should have even be able to release the *dentist* from her duty not to harm me later, since I couldn't then take up the requisite perspective. Schofield's view seems to entail that I can't make perfectly normal, morally healthy tradeoffs across my life. I may not cause myself a small future harm (or authorize others to do so) for my greater benefit, whether that benefit comes to me now or later.[15]

In sum: we have four options for allowing duties to oneself without release. They are:

(i) Construe duties to oneself as *non-juridical*.
(ii) Settle for duties that aren't strictly owed *to oneself*.
(iii) Give up on the idea that anyone can release *anyone* from duties.
(iv) Find *contingent reasons* why we might not be able to release ourselves.

But these all have their costs. (i) Non-juridical duties are more like nudges than obligations; (ii) it is a pyrrhic victory if we can only defend duties regarding oneself, or owed to one's parts; (iii) people seem to release each other from duties all the time; and (iv) two of the main views of unwaivable duties have trouble accounting for the cases: Kading can't vindicate our normal duties to oneself, and Schofield generates duties where we ought to be scot-free.

## 3. The possibility of self-release

If we want to defend duties to oneself, it might not be enough to block the entailment from duties to powers of release. So let's grant the entailment. We are now going to consider a more radical

[15] The same problem may arise for other views like Schofield's (e.g. Sartorius 1985: 245–46).

solution to the Paradox, which is to argue that one can *release oneself* from a genuine duty.

To start, let's consider some examples. Suppose I decide to drop out of a class mid-term. While enrolled, I have obligations to show up and participate—but once I sign the form to drop out, I release myself (Fruh 2014: 64). A second case, more controversial, is that I might promise myself to do something—like doing more to help my neighbors—and release myself when I learn that it will be much more onerous than expected.[16] There isn't anything obviously paradoxical about any of this. I start out obligated; I end up free.

And yet, Singer thinks that the very notion of self-release is "self-contradictory."[17] Why? Here is what he says:

> It is essential to the nature of an obligation that no one can release himself from an obligation by not wishing to perform it or by deciding not to perform it, or, indeed, *in any other way whatsoever*. In other words, no one can release himself from an obligation, just as no one can release himself from a promise. (Singer 1959: 202, emphasis added)

This isn't very persuasive. First, Singer draws an analogy with promises to oneself, which he assumes are paradoxical. But why should they be? The only problem he raises for self-promises is that they allow for self-release, which, for all we have seen so far, is perfectly possible.[18]

Singer's other point is that it is "essential to the nature of an obligation" that we can't release ourselves from one "in any…way whatsoever." There is something right about this. Obligations aren't optional, like hobbies or fancies. And so we might think: obligations must be *inescapable*.

---

[16] For more on release from self-promises, see Hill 1991, Habib 2009, Rosati 2011, and Fruh 2014. For more (extremely controversial) cases of self-release, see Chapter 2 (RAO).

[17] He writes: "a duty to oneself…would be a duty from which one could release oneself at will, and this is self-contradictory" (Singer 1959: 202–3). Kant, too, claims without much argument that a "contradiction" follows if we say that "the one imposing obligation (auctor obligationis) could always release the one put under obligation (subiectum obligationis) from the obligation (terminus obligationis), so that (if both are one and the same subject) he would not be bound at all to a duty he lays upon himself" (MS 6:417).

[18] Singer (1959: 203) has no other objection to self-promises, though he does suggest that our talk about self-promising could be interpreted as metaphorical; we use it to "express a settled determination" to act.

How can we respond? I think we should take a cue from G.A. Cohen's (1996) critique of Hobbes, who makes the analogous point about *legal* obligations. In the *Leviathan*, Hobbes argues that a Sovereign who can change the law isn't bound by the law:

> The Sovereaign of a Common-wealth, be it an assembly, or one man, is not subject to the civil laws. For having power to make, and repeal laws, he may when he pleaseth, free himself from that subjection, by repealing those laws that trouble him, and making of new; and consequently he was free before. *For he is free, that can be free when he will*: (1651 [1991]: Chapter 26, Section 2, 313, emphasis added; see also *De Cive*, XII.4, XI.14)

Cohen's response is that even the sovereign really is bound by "those laws that trouble him," *until* he elects to repeal. After all, the Sovereign really does have the authority to make laws, and the laws as they do in fact pronounce on how he is supposed to act. This is sufficient to make the Sovereign legally bound—even though, at will, he could undo it.[19] Cohen (1996: 170) concludes:

> The big mistake in [Hobbes'] argument is the supposition that if I *can* repeal the law, then it fails to bind me even when I have not *yet* repealed it. Hobbes is wrong that, if you can free yourself at will, then you are already free, that 'he is free, that can be free when he will'. But other important things do follow from my being able to free myself at will, for example, that I cannot complain about my unfreedom.

In a slogan: "can get free" doesn't mean "free already."

The same goes, I think, for duties to oneself. Even if we can free ourselves, we are bound until we do so. As it turns out, this point has been made already in more specific contexts. Rosati (2011: 134–35, emphasis original) makes the same basic point as Cohen, but about freeing oneself from a self-promise, not a self-made law:

> Suppose one insists…that in a two-party promise, the promisee can simply release the promisor at will. We would not be tempted to say that because the promisee can release the promisor at will, the promisor is not really obligated. So long as the promisee has not released the promisor, she is indeed bound. But then we should say the same thing about self-promises: although an agent, as promisee, can release herself, as promisor, at will, *so long as she does not release herself, she is indeed bound*."

[19] Suppose…that the law is indeed universal, or that it includes me within its scope by virtue of some other semantic or pragmatic feature of it. Then, if I had the authority to legislate it, it indeed binds me, as long as I do not repeal it" (Cohen 1996: 170).

And Kyle Fruh (2014: 64–5, emphasis original) has hit on the same basic idea, considering a case where a student releases herself from the honor code by dropping out:

> …simply because you are able to make it the case that *at some point* you have no honor code obligation to refrain from plagiarism does not mean that you have no such obligation *now*. Your subsequent decision to drop out only ceases the ongoing obligation—it does not undo what had up until that point been an obligation.

And this is exactly what I suggest might hold for waivable duties to oneself in general. To give just one example: I think we owe it to ourselves not to cause ourselves significant harm. But if we think about the matter seriously, and have relatively decent reasons, we can release ourselves from that obligation, making self-harm fine where before we were bound not to do it.

That is the core case for self-release. There are intuitive examples, and while Singer thinks they don't make sense, we seem to have a coherent way to say how they work: agents are bound *until* they free themselves. (It helps to remind ourselves that this kind of thing happens with all kinds of properties, not just obligations. I might be bound by ropes even though the slightest movement would make them fall off (cf. Hill 1991). A door might be locked even if it will automatically unlock should I ever try to go through (cf. Fruh 2014: 65–6). The ropes bind me, and the door locks me in, *until* the moment when they *normally* would kick in to constrain me.)

But could there be more to Singer's objection than I am giving him credit for? Isn't there *something* to the idea that we can't release ourselves from duties? I think we should try to develop Singer's worry here, first as a principle about the logic of duties, and second as a claim about the link between duties and normative reasons.[20] Let's start with logic.

---

[20] Here's a third way to hear Singer's objection: we can't release ourselves from an obligation because obligations are categorical, in the sense that they bind us no matter what we want (see e.g. Wick (1960: 161); Hills (2003: 133)). By contrast, a hypothetical norm, like "You have to take the A train, if you want to get to Harlem," binds us only in so far as we want or aim at a certain end. If you aren't off to Harlem, no need to take the train. But no one is suggesting that we release ourselves merely by not desiring to comply with our duty. Duties to oneself can be categorical, in the sense of not being based in desires, even though they are sensitive to our choice to release ourselves. (Analogously, in Hobbes' legal case, the laws are grounded in the act of legislation, not in the

## 4. Self-release: the logical objection

Singer's objection to self-release is that real obligations can't be shirked. As Hobbes puts it: "he is free, that can be free when he will."

You might wonder: is this a good principle in the *logic* of duties? Is there any logically natural—or even coherent—way to model counterexamples?

Hobbes' dictum, as it turns out, has a natural translation into deontic logic—the logic of obligation and permission. Here it is:

*S4 Axiom*
$\Diamond\Diamond p \rightarrow \Diamond p$

The diamond means "permissibly." So, this statement says, roughly, if a certain act $p$ is permissible to make permissible, then it's as good as permissible already.[21] "He is free, that can be free when he will." Whoever is free to become free is, automatically, just plain free.

But it is easy to construct countermodels to the S4 Axiom in standard deontic logic.[22] First, a case of legal obligation. Suppose I'm allowed to petition the state for an exception to its "No birthday cakes" law. (It's my birthday; I want cake.) In the actual world, I don't petition, and I don't bake anything. But I had the ability, and the right, to get an exception. If I had gotten one, I might still not have baked a cake (perfectly permissible). Or I might have gone ahead and baked one (permissible *only given* the exception). Here is a way to depict some of what's going on in these three possible worlds; if an arrow points from one world to another, then what goes on in the second is *permissible* by the laws as they are in the first:

---

lawmaker's desires, but the sovereign can choose to repeal them.)

[21] Standard deontic logic uses *sentential operators*, like 'permissibly', rather than *predicates of acts* (as in von Wright 1951). That is part of why my gloss is rough. For more, see McNamara 2018.

[22] It's called the S4 Axiom because it's characteristic of the modal logic S4—which is strictly stronger than the standard deontic logic D, on which the S4 Axiom is invalid.

| $w_1$ | → | $w_2$ | → | $w_3$ |
|---|---|---|---|---|
| I don't petition. | | I do petition. | | I do petition. |
| I don't bake the cake. | | I don't bake the cake. | | I do bake the cake. |

Here is the key point. *What goes on in* $w_3$ *is not permissible from the point of view of the law as it is in* $w_1$. No petition, no cakes allowed. And this is true even though $w_1$'s laws permit what happens in $w_2$, whose laws permit the cake-baking of $w_3$.[23] In $w_1$, it is legally permissible to make cake-baking legally permissible, but that isn't enough to make it actually permissible.

What about moral obligation? Suppose I am Carol's neighbor, and I'm allowed to ask permission to use her bike; if she complies, waiving her rights, I am morally permitted to ride it to work. We get the same structure as before:

| $w_4$ | → | $w_5$ | → | $w_6$ |
|---|---|---|---|---|
| I don't get permission. | | I do get permission. | | I do get permission. |
| I don't ride her bike. | | I don't ride her back. | | I do ride her bike. |

Riding the bike isn't permissible given my obligations as they are in $w_4$. But those obligations do permit me to ask for permission, and if do that, I am free to race off into the distance. My obligations change when I choose to ask permission.[24]

I conclude that there is a natural way to model self-release. (Leaving aside nice details like the *timing* of release.) We might be permitted to make an act permissible without its actually being permissible. Hobbes' dictum—"he is free, that can be free when he will"—isn't a logical truth.

---

[23] The arrow is the "accessibility relation." The hallmark of an S4 violation is that the relation is intransitive, as in our case. (To fully depict the case, we would add arrows from $w_3$ to $w_2$ and $w_1$; from $w_2$ to $w_1$; and from each world to itself.)

[24] We would have to add the same arrows as before for a full description; see fn. 23, just above.

## 5. Self-release: the meta-normative objection

"He is free, that can be free when he will" is not a logical truth. Nor is Singer's version of it: "No one can release himself from an obligation." But there is still *something* true about them.

Think back to Hobbes' sovereign. Because he can change the laws at will, there is a sense in which they never restrain him. He never has to give up what he wants in order to follow the law; he can change the law instead. If the sovereign wants free sandwiches from the bakery, he may legislate that the sovereign eats for free. If I want free sandwiches, I have to steal them.

But if I have the power of self-release, then at least when it comes to duties to myself, I'm a bit like the sovereign—unrestrained. I never have to give up something I want in order to comply with a duty to myself. I can waive it and do as I please. (Unless there are specific reasons why I *can't* waive; see §2 above.)

Is there anything fishy about norms that don't restrict us? Here is something you might worry about: such norms don't give us any *reason* to comply with them. A (normative) reason is a fact that counts in favor of some way of acting, making it more choiceworthy (Dancy 2004, Parfit 2011, Chang 2014). But the fact that I have a *waivable* duty doesn't seem to be a reason to comply. Even if the sovereign has reason not to break the law—and the law forbids littering—the law doesn't give him any reason not to litter. It only counts against littering *without changing the law*. Similarly, if I have a waivable duty to myself not to harm myself in certain ways, it's not a reason against harming; it only counts against *harming without waiving*.[25]

Is it possible to have a duty without having any reason to comply? Many say *no:* a duty to do X entails a reason to do X.[26] This suggests a meta-normative version of Singer's objection to self-

[25] Exception: the agent might have independent reason not to waive the duty. For example, suppose the sovereign has promised his friend not to repeal any laws. Then the fact that the law bans littering *does* entail that the sovereign has reason not to litter. He has reason to obey the laws if he doesn't change them, and he has reason not to change them.

[26] For example, Hills (2003: 131) presupposes that duties are reasons when she asks, "When

release. If we could release ourselves from duties, those duties wouldn't restrict us, and we wouldn't have reasons to comply. But it's essential to the nature of a duty that we have reasons to comply. So self-release is impossible.

This is the best objection we have seen so far. It is not plainly mistaken, like the logical objection; and it is more substantial than Singer's original. There is a genuine tension between self-release and the idea that we have reasons to comply with all duties.

That said, I don't see why we have to think duties *always* give reasons. What about duties that we don't know about? Suppose we make a deal: you let me use your skateboard today, and I give you $5 tomorrow. I really do owe you $5—i.e. I have a duty to you to give you $5. That remains true even if I wipe out on the skateboard, bonk my head, and forget my obligation. Since I have no idea that I owe you the money, the obligation doesn't give me any *reason*.[27] Why isn't that a case of duties without reasons? Now, someone might reply: there is still a conditional reason. *If* I knew about the obligation, *then* I would have a reason. But if all we need are conditional reasons, then self-release isn't problematic at all. Even I can say that waivable duties give us conditional reasons. We would have reason to comply with them *if they couldn't be waived*.

So I think waivable duties to ourselves are meta-normatively innocent. We don't have reasons to obey them, but the same is true of other bona fide duties, such as those we have forgotten. That said, my arguments here might not be the last word. There may be other grounds for thinking that duties must entail reasons. If so, then the meta-normative objection to self-release will still have some bite.

---

should reasons for action be classed as 'duties to the self'?" (See also Schofield 2015: 10.) Reath (2006: 241) defines duties-to as a kind of reason: one owes a duty to "the person who, under the circumstances, is the source of reasons for one to act." Fruh (2014: 65) claims that we "expect obligations to be weighty things that well-reasoning, conscientious, non-akratic agents will have to make room for in their deliberations"—which is what I have in mind when I say "reason."

[27] I assume that my reasons are sensitive to my ignorance but based in objective facts (Dancy 2000, Lord 2015). We could construct a similar case for duties that one becomes *unable* to fulfill.

## 6. Conclusion

Duties to oneself are "well embedded both in traditional moral philosophy and ordinary moral thinking," as even the skeptical Singer admits (1959: 202). I have tried to defend our ordinary thoughts from Singer's Paradox by arguing that we can coherently release ourselves from duties.

Let me close with a suggestion: we should flip the Paradox on its head, turning it into an argument for self-release. Singer himself grants that, intuitively, duties imply rights, rights imply powers of release, and we have duties to ourselves. So we can argue:

(1) If A has a duty to B, then B has a right against (or with respect to) A;

(2) if B has a right against A, he can give it up and release A from the obligation; and

(4) people have obligations to themselves.

Conclusion: people can release themselves from obligations (to themselves).

The inference is valid; the premises are solid (though we might want to allow some cases where duties don't imply powers of release); and I have argued that there is no decisive objection to the conclusion. Why not prefer this argument to Singer's?

This new argument—like most of this paper—leaves plenty to the imagination. We still have not seen any details about *what* we owe to ourselves, *when* a duty is waivable, and *how* we might release ourselves from duties. My goal here has not been to answer these questions, or to defend any theory of self-obligation. I have just been clearing a path. Duties to oneself are possible, even defensible, but there is still much to learn about how they work and why they might matter.

**REFERENCES**


Chang, Ruth (2014). "Practical Reasons: The Problem of Gridlock," in Barry Dainton and Howard Robinson (eds.), *The Bloomsbury Companion to Analytic Philosophy*. Continuum Publishing Corporation. 474–499.

Cholbi, Michael (2015). "On Marcus Singer's 'On Duties to Oneself'," in *Ethics* 125 (3): 851–853.

Cohen, G.A. (1996). "Reason, Humanity, and the Moral Law," in Korsgaard 1996: 167–188.

Dancy, Jonathan (2000). *Practical Reality*. Oxford: Oxford University Press.

-------(2004). *Ethics Without Principles*. Oxford: Oxford University Press.

Darwall, Stephen L. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge: Harvard University Press.

Denis, Lara (1997). "Kant's Ethics and Duties to Oneself," in *Pacific Philosophical Quarterly* 78 (4): 321–348.

-------(2001). *Moral Self-Regard: Duties to Oneself in Kant's Moral Theory*. Routledge.

Eisenberg, Paul D. (1968). "Duties to Oneself and the Concept of Morality," in *Inquiry: An Interdisciplinary Journal of Philosophy* 11 (1–4): 129–154.

Fotion, Nicholas (1965). "We Can Have Moral Obligations to Ourselves," in *Australasian Journal of Philosophy* 43 (1): 27–34.

-------(1970). "Eisenberg and Self-Obligations," in *Inquiry: An Interdisciplinary Journal of Philosophy* 13 (1–4): 458–461.

Fruh, Kyle (2014). The Power to Promise Oneself," in *The Southern Journal of Philosophy* 52 (1): 61–85.

Gilbert, Margaret (2018). *Rights and Demands: A Foundational Inquiry*. Oxford: Oxford University Press.

Hart, H.L.A. (1955). "Are There Any Natural Rights?" in *Philosophical Review* 64 (2): 175–191.

Hill, Thomas (1991). "Promises to Oneself," in *Autonomy and Self-Respect*. New York: Cambridge

University Press.

Hills, Alison (2003). "Duties and Duties to the Self," in *American Philosophical Quarterly* 40 (2): 131–142.

Hobbes, Thomas (1949). *De Cive or The Citizen*. Sterling P. Lamprecht (ed.). New York: Appleton-Century-Crafts.

-------(1651) [1991]. *Leviathan*. Richard Tuck (ed.). Cambridge: Cambridge University Press.

Hohfeld, Wesley Newcomb (1919). *Fundamental Legal Conceptions*. New Haven: Yale University Press.

Johnson, Robert N. (2010). "Duties to and Regarding Others," in Lara Denis (ed.), *Kant's Metaphysics of Morals: A Critical Guide*. Cambridge: Cambridge University Press.

Kading, Daniel (1960). "Are There Really 'No Duties to Oneself'?" in *Ethics* 70 (2): 155–157.

Kant, Immanuel (1996). *Practical Philosophy*. Mary Gregor (ed.). Cambridge: Cambridge University Press.

-------(1997). *Lectures on Ethics*. J.B. Schneewind (ed.), translated by Peter Heath. Cambridge: Cambridge University Press.

Kerstein, Samuel J. (2008). "Treating Oneself Merely as a Means," in Monika Betzler (ed.), *Kant's Ethics of Virtue*. Berlin: New York: de Gruyter: 201–218.

Knight, Frank H. (1960). "I, Me, My Self, and My Duties," in *Ethics* 71 (3): 209–212.

Korsgaard, Christine M. (1996). *The Sources of Normativity*. Onora O'Neill (ed.). Cambridge: Cambridge University Press.

Lord, Errol (2015). "Acting for the Right Reasons, Abilities, and Obligation," in *Oxford Studies in Metaethics, Volume 10*. Oxford: Oxford University Press.

Mavrodes, George I.; Narveson, Jan & Meiland, J. W. (1964). "Duties to Oneself," in *Analysis* 24 (5): 165–171.

McNamara, Paul (2018). "Deontic Logic," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of*

*Philosophy* (Fall 2018 Edition). URL = <https://plato.stanford.edu/archives/fall2018/entries/logic-deontic/>.

Meiland, Jack W. (1963). "Duty and Interest," in *Analysis* 23 (5): 106–110.

Mothersill, Mary (1961). "Professor Wick on Duties to Oneself," in *Ethics* 71 (3): 205–208.

Oakley, Tim (2016). "How to Release Oneself from an Obligation: Good News for Duties to Oneself," in *Australasian Journal of Philosophy* 95 (1): 70–80.

Parfit, Derek (2011). *On What Matters, Volume 1*. Oxford: Oxford University Press.

Paton, Margaret (1990). "A Reconsideration of Kant's Treatment of Duties to Oneself," in *Philosophical Quarterly* 40 (159): 222–233.

Potter, Nelson (1998). "Duties to Oneself, Motivational Internalism, and Self-Deception in Kant's Ethics," in Mark Timmons (ed.) *Kant's Metaphysics of Morals: Interpretive Essays*. Clarendon Press: 371–390.

Reath, Andrews (2006). *Agency and Autonomy in Kant's Moral Thoery: Selected Essays*. Oxford: Oxford University Press.

Rosati, Connie 2011. "The Importance of Self-Promises," in Hanoch Sheinman (ed.), *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press. 124–155.

Sartorius, Rolf (1985). "Utilitarianism, Rights, and Duties to Self," in *American Philosophical Quarterly* 22 (3): 241–249.

Schneewind, Jerome B. (1997). *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.

Schofield, Paul (2015). "On the Existence of Duties to the Self," in *Philosophy and Phenomenological Research* 90 (3): 505–528.

-------(2018). "Paternalism and Right," in *The Journal of Political Philosophy* 26 (1): 65–83.

-------(forthcoming). "Practical Identity and Duties to the Self," in *American Philosophical Quarterly*.

Singer, Marcus G. (1959). "On Duties to Oneself," in *Ethics* 69 (3): 202–205.

-------(1963). "Duties and Duties to Oneself," in *Ethics* 73 (2): 133–142.

Steiner, Hillel (2013). "Directed Duties and Inalienable Rights," in *Ethics* 123 (2): 230–244

Thompson, Michael (2004). "What is it to Wrong Someone? A Puzzle About Justice," in R. Jay Wallace, Philip Pettit, and Michael Smith (eds.), *Reason and Value: Themes from the Philosophy of Joseph Raz*. Clarendon Press. 333–384.

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

Timmermann, Jens (2006). "Kantian Duties to the Self, Explained and Defended," in *Philosophy* 81 (3): 505–530.

von Wright, G.H. (1951). "Deontic Logic," in *Mind* 60: 1–15.

Wick, Warner (1960). "More About Duties to Oneself," in *Ethics* 70 (2): 158–163.

-------(1961). "Still More About Duties to Oneself," in *Ethics* 71 (3): 213–217.

# CHAPTER TWO

## Rights Against Oneself

## Rights Against Oneself

I argue that we have the same basic rights against ourselves as against others: waivable moral rights forbidding harm and debasement, killing and kidney-removal. This is the Self-Other Symmetric view of rights. Straightaway the view seems to face counterexamples—pairs of cases where a certain act (like poking an eye) violates a right when done to others and not when done to oneself. But what explains the difference is not that we lack rights against ourselves: it is that these rights are typically harder to violate, because when we decide to act, we give ourselves consent, and consent waives rights. After considering some hard cases—accidental harms, alcoholic sex, killings with consent—I conclude that rights against oneself are not just defensible but fruitful, yielding a fresh, simple, and impartial theory of self-obligation. We wrong ourselves just when actions done to ourselves would wrong an equally willing other in a position like ours.

### 1. Introduction

We owe a great deal to others, morally speaking. We may not harm them, take their stuff, or trespass where they live; we may not slip antidepressants into their coffee to brighten their day or abscond with their kidneys to buy someone else another decade. These acts aren't just morally wrong: normally, they wrong other people—putting them in a position to resent or forgive us, to demand apology or restitution—by virtue of a failure to give what is owed. But what, if anything, do we owe to *ourselves*? Can we wrong ourselves in the same sense in which we wrong others? Or is it just plain wrong or pure idiocy, wronging no one, when we expose ourselves to pointless risk, let others walk all over us, neglect our health and happiness?

It's natural to think that we owe ourselves at least some basic respect. Abject servility and reckless self-destruction do not feel morally neutral; they can seem like direct offenses to one's own moral status—the kind of thing that might be legally forbidden, and for which one might sensibly resent oneself. (Think of how you might feel after an all-too avoidable injury.)

Even more obvious, it seems, is that most of what we owe to others we *don't* owe to ourselves. I am free to spike my own coffee, donate my kidney, and toss away my loose change; I wrong no one when I "invade" my own body (e.g. by brushing my teeth), "trespass" unannounced

in my own kitchen, or cause myself slight harms—plucking hairs, poking my own eye. So long as I don't do anything too gruesome or debasing, I cannot deliberately wrong myself. Wronging others is relatively easy.

And so ethicists conclude that moral obligations must be *Self-Other Asymmetric*. I owe different things to myself than to other people, and the reason why is that they are others, and I am me. The self/other difference makes a moral difference. This much is standard in the literature on self-obligation, where the really live question is whether we can rescue *any* duties to oneself from incoherence. The main threat, best pressed by Marcus Singer (1959), is that since obligations entail corresponding rights, obligations to oneself must entail those "paradoxical" things, the most maligned monstrosities of 20th-century normative ethics: rights against oneself. Singer surmises that duties to oneself must be impossible. (He thinks duties to others, and corresponding rights, are fine.) His critics insist that duties to oneself are coherent, but concede that they must be uniquely limited: they don't proscribe the same things as duties to others, and they imply no waivable rights.

Against this consensus, I argue that our basic rights and duties are *Self-Other Symmetric:* we owe the same obligations to ourselves as to others, and these obligations correspond to bona fide moral rights against ourselves. I begin by showing that the Symmetric view does not really count minor harms to and "trespasses" against oneself as wrongings (§3). The key fact here is that we consent to our own decisions, and so our rights against ourselves are like the rights of a cooperating other—sure to be waived before they can be violated. In such cases, I say that rights are "finkish" and do not provide normative reasons for compliance (§4). *What we owe to ourselves* is not always part of *what we should do with ourselves*.

That said, sometimes our rights do seem to constrain us; it is possible to wrong ourselves. But how? And when? The Symmetric answer is that we wrong ourselves by violating rights, just as

we wrong others, and this happens when we fail to waive rights via consent.[28] Failures could occur for a number of reasons (§§5–6). In some cases, we lack the authority to waive certain awesome rights, like the right to life. Another possibility is that our self-consent is flawed, e.g. by booze or lies. Regardless, the limits of what I may do to myself turn out to be nothing strange or *sui generis;* they are simply the limits of what I may do to any equally willing person relevantly like me. The cases that first strike us as proof of a Self-Other Asymmetry turn out to be tectonically Symmetric.

After dealing with these problems for Symmetric duties, we move to Singer's paradox (§7), which threatens any duties to oneself. My response is that Singer has no good argument for his key premise—the impossibility of self-release—and that finkish rights are counterexamples. I conclude (§8) that Self-Other Symmetric rights have been misunderstood. They are not paradoxical; they do not imply that we must treat ourselves as gingerly as we treat unwilling others. At heart, our commonsense views are Self-Other Symmetric, and the special character of duties to self is due to the fact that we typically, though not always, act with the blessing of our own consent.

## 2. What is a right?

Before we get to rights against oneself, we face the more basic question of what rights are in general. What is it to have a right against *anybody,* whoever it may be? As it turns out, there is no easy answer. Every theory of rights is controversial, and "right" in normal English is paradoxically polysemous, denoting either freedoms (the right to free speech) or constraints (rights against harm).[29]

Fortunately, we don't need to prove that any theory or definition is uniquely best. For our

[28] Whenever I capitalize "(a)symmetry," I am talking about the Self-Other Symmetry.

[29] The main theories of rights are "interest" theories (e.g. Raz 1986), "will" theories (e.g. Hart 1955, Steiner 1994), and more recently, Margaret Gilbert's (2018) theory of demand-rights as arising from joint commitments. Sumner (1987: 211) suggests that interest theories, though not will theories, allow for rights against oneself. Gilbert (2018) uses her view to argue for the coherence of rights against oneself.

purposes, we just need a working concept: something thin enough to be uncontroversial, but robust enough to capture our target notion of rights against oneself. We can express our concept in three truisms.

First, rights are a matter of what we owe: if I have a right against you that you not poke my eye, then you owe a duty (or "obligation") to me not to poke it—and vice versa.[30] Now, the words "duty" and "obligation" might be misleading; they tend to connote laws and institutional roles. One hears them and thinks of "jury duty," "fiduciary obligation," and "duties of the office." But we are talking about moral obligations, which are often informal: we owe each other basic respect and debts of gratitude. These debts aren't meant to be enforced by brick-and-mortar institutions; there is no "morality police." Still they are part of what we owe to each other, and so they are a matter of what is ours by right.

Since there is no morality police, there is also no "morality jail." No one in a badge is going to lock you up for being immoral. So why do rights matter? What is the significance of infringing a moral right, going derelict on a moral obligation? This leads to our second truism: rights can be informally *enforced* (Hart 1955: 178). If you are about to infringe a right, others may stop you using unusual means, and once you have done it, you deserve a sanction at the right holder's discretion. For example, if you are about to poke my eye, I may defend myself by slapping you. If you do end up harming me, then (other things equal) you wrong me, and I gain the standing to resent you (Strawson 1962), seek restitution, or perhaps demand compensation. Set out to violate a right, and

[30] This is the classic Hohfeldian equivalence of duties with "rights in the strictest sense," i.e. "claims" or "claim rights." See Hohfeld 1919, Thomson 1990: Chapter 1, and Gilbert 2018; see also Johnson 2010 for an argument that Kantians, too, should accept the correlation between rights and duties. Throughout this paper, I use "right" only to mean "claim." In other literature, "right" may also refer to a liberty (the absence of a right against the agent), a power (the ability or standing to change rights), or an immunity (the absence of a power over the agent's rights). One last verbal point: following Thomson (1990) and others, I use "violate" to signal the *wrongful* infringement of a right; "infringements" aren't necessarily wrong.

you may end up losing some moral privileges. (By contrast, if you choose not to brush your teeth tonight, that is your business; the norms of adult dental hygiene aren't enforceable.)

Finally, unlike divine laws and intrinsic values, our rights are under our direct control: rights can typically be *waived* by consent. If I give you permission to poke my eye, you may go ahead; I can no longer sensibly resent you for doing it, just as I can't demand an apology from the dentist I allow to drill my teeth. Rights thus offer a kind of flexible protection; one may lean on them or give them up as one desires, all thanks to the mechanism of consent. (For now we can leave open the questions of when consent is "valid" enough to waive rights, and of whether any rights are unwaivable.)

That is our core concept. At a minimum, when we talk about rights, we are talking about moral obligations that are (other things equal) both enforceable and sensitive to consent.[31] Now we turn to the question of whether our most familiar moral rights are Self-Other Symmetric.[32]

### 3. Simple harms and self-consent

On a Self-Other Symmetric view, we have the same rights against ourselves as against others; the self/other difference doesn't affect who has which rights against whom, if other factors are held equal.

At a first, most naïve glance, Symmetry might seem like a nice default. It is simple and principled: it says that my rights apply the same to similar people, no matter who, which resonates

---

[31] For clarity, I will use "obligation" only to talk about duties owed *to* people, not moral requirements in general. Also, to keep things simple, I will ignore the big question of how legal and moral obligations are related; on this issue, see e.g. Thomson 1990: Chapter 1, Gilbert 2018: Chapter 4.

[32] I will not argue that *all* possible systems of rights are Self-Other Symmetric. If we wanted to, we could all agree to dole out rights in an Asymmetric way. But I do think that there is a principled reason why even conventional systems of moral rights tend to be Symmetric: see Chapter 3 (FRP) for an argument that Self-Other Symmetric rights are needed to avoid the systematic undercutting of rights and prerogatives.

with the deeper idea that the moral point of view is impartial. The Self-Other Symmetry is a specific case of the lovely idea that no one should be morally special just because of who they are, that each of us is "one among many, equally real" (Nagel 1970). This is the moral vision that makes fairness fundamental and gives the Golden Rule its luster. If we like the simple and impartial, we will see Symmetry as the view to beat.

But look again, and rights can seem *obviously* Self-Other Asymmetric: your rights appear to bind everyone but you, as mine bind everyone but me. According to the standard line, we can observe this Asymmetry at work in various cases, minimal pairs where everything is held fixed except the self/other difference. The classic example involves a simple, pointless harm. For instance: if I poke your eyeball out of the blue, I wrong you by violating your right not to have your eyeballs poked; poking my own eye, however, is no violation. Maybe it's a senseless thing to do. Probably I shouldn't do it. But it doesn't wrong anyone, and it certainly doesn't violate a right. The classic conclusion is that we have this particular right against others but not against ourselves.

(For other rights, we construct other pairs. Pulling someone's hair, barging into a bedroom, publicizing passwords, spiking a drink, riding a bicycle to work, taking money from a wallet, deleting someone's emails—these are transgressions if I'm messing with *your* inbox, bike, and hairdo; not so, apparently, if my acts touch only what is *mine*.)[33]

But does merely stating this "minimal pair" really establish a Self-Other Asymmetry? Hardly. We still have to check whether the cases are equalized, in the sense that the only relevant difference

[33] Here are the main examples from the literature. Stocker writes: "If Nick knowingly puts his hand in a fire, causing himself considerable pain and harm, he does not act wrongly, he does not fail an obligation or a duty. But clearly, he would act wrongly, would fail an obligation or a duty, were he to put someone else's hand in a fire, causing that person considerable pain and harm" (1976: 210). Later, we have Slote: "Even if one may not cut up another person to furnish healthy organs that will save the lives of five injured or sick individuals, there is no immediate moral bar to cutting oneself up in order to save five other people" (1984: 183). Finally, Pettit: "One may commit suicide for a higher cause, but not murder" (1986: 409). See also Sider 1993. My cases are most like Stocker's, where the harmful acts have no good effects.

is that I harm *myself* in the one case and someone *else* in the other. The cases won't prove anything unless what I do solo perfectly parallels what I do to the other.

The problem is that imperfect parallels—non-minimal pairs—will generate all kinds of bogus "Self-Other Asymmetries." Example: you're deathly allergic to peanuts and I'm not. I seem to have a very special reason not to put peanuts in *your* breakfast, and no such reason to avoid putting peanuts in *mine*. That doesn't prove that legume-based reasons are Self-Other Asymmetric! The only "asymmetry" here is between those who are allergic and those who are not. Your allergy, not your otherness, is responsible for the reason to be careful. This becomes clear once we equalize the cases. If we are *both* allergic, I also have a special reason not to give *myself* peanuts; and I have no reason in either case if we are both allergy-free. When allergies are equal, the self/other difference makes no difference. That is the hallmark of a bogus "Self-Other Asymmetry;" the test is to check whether it survives after equalizing.

What, then, should we make of the pair of eye-pokings? Do they prove a Self-Other Asymmetry, or are they just poorly equalized? My view is that they clearly aren't equal; there is a crucial asymmetry in *consent*. When I choose to poke my own eye, the person being poked is an active participant, not an unwilling bystander; there is consent "implicit in actions we do to ourselves" (Slote 1984: 190). But when I poke you out of the blue, your consent is not implied.

Moreover, we have a positive reason to think that self-consent is the difference-maker here. Remember our test: what happens when we equalize the cases? Well, if *you* consent to having your eye poked, I violate no rights by poking away. Maybe it's a senseless thing to do. Probably I shouldn't do it. But it doesn't wrong anyone, and it certainly doesn't violate a right. That's a sign that this "Asymmetry" is bogus. It would seem, in Slote's words, that the Self-Other Asymmetry of rights turns out not to be a "deep feature of morality but rather derivative from and justifiable in terms of

the moral importance of consent" (1984: 191).[34]

This should be great news. The appeal to self-consent gives us a simple, Symmetric way to unify two realms of morality—self-regarding acts and the treatment of consenting others. The view is not ad hoc. It does not introduce any fundamental factors or invoke any highfalutin metaphysics of self and other. The basic idea, which has been understood for decades,[35] simply flows from the idea that we consent to our own actions plus the truism that consent waives rights.

So why isn't anyone convinced? Why are Symmetry and self-consent dismissed in footnotes while the Self-Other Asymmetry is treated as a *fait accompli*?[36] One stock objection is that self-consent

---

[34] What about the case where I poke my own eye without consent? It's complicated. There is no way to intentionally poke your own eye against your own will, so there is no one-person analogue to the case where I poke your eye after you refuse consent (Kagan 1989: 214). As for accidental self-injury, see §5, below.

[35] Stocker writes: "It could be said that it is as wrong to cause oneself pain as it is to cause someone else pain, but that in agent-regarding cases, patients waive their right not to be caused pain" (1976: 213, see also 211). Slote writes: "If I harm myself or avoid a benefit, I presumably do this willingly, whereas the agent whom I refuse to benefit does not consent to this neglect (and when she does there is nothing wrong with what I do)" (1984: 190–91). Pettit suggests that the moral difference between self- and other-killing might be due to "the fact that one can consent to one's own deprivation, while the other is not allowed the chance to consent to this" (1986: 410). See also Kagan 1989: Chapter 6, Section 2. Sider (1993: 125) discusses preferences, but not consent.

[36] "The presence or lack of consent seems to have little to do with the self-other asymmetry" (Portmore 2003: 311, fn. 16). Meanwhile, the Self-Other Asymmetry of rights and duties is widely accepted with hardly any argument. Judith Jarvis Thomson (1990: 42) writes: "I am sure that when we look more closely at claims we will not find ourselves wanting to say that person has claims against himself or herself" (she doesn't revisit the issue). Hill (1991a: 4) [1973: 87] says that it "does seem absurd to say that a person could literally violate his own rights or owe himself a debt of gratitude." Onora O'Neill (2001: 428), in an encyclopedia entry on duties, reports that "the idea of a right against oneself is generally thought paradoxical." Peter Jones (1999: 94) concurs: "It is generally accepted that one cannot hold rights against oneself." Michael Cholbi (2013: 111) calls it an "unlikely assumption" that we have at least the same duties to ourselves as to others. Frances Kamm asserts that "one cannot have rights against oneself" (2007: 235), elsewhere dismissing them as "implausible" and not "possible" (2007: 241). Thomas Pogge (2002: 62) at least gives a reason to support his Asymmetry: "Speaking of rights against oneself…is problematic because of the connection between having rights and being entitled to claim and defend one's rights as well as to protest and sometimes punish the infringement of those rights. We do not engage in such claiming, defending, and punishing activities against ourselves…." A similar reason is given by Hillel Steiner (2013: 240, fn. 21): "I cannot have rights against myself; I cannot be both plaintiff and defendant in a legal suit nor, presumably, in its moral counterpart." See also Haase 2014a: 7, 2014b: 131. The Asymmetry of rights is also built into the popular view that morality as a whole is other-regarding

can't explain the full range of cases; in further, trickier minimal pairs, our intuitions are allegedly more Asymmetric. But before taking on these hard cases, we should consider two conceptual objections that apply even in our easy case of simple harms. We begin with the worry that self-consent is impossible.

## 4. Conceptual problems: self-consent and finkish rights

Consent, you might think, is not the sort of thing one can give to oneself. There has to be one party who offers—signing the waiver, flashing the "OK," nodding—and another party who accepts. It's a moral tango; it takes two. If this is right, then the Self-Other Symmetry is doomed, since there will be no way to waive a right against oneself, and simple self-harms will be judged wrong.[37]

Now, maybe there are special cases where consenting to oneself is like consenting to others. If I am in charge of both the inventory and event-planning at my company, I might need to fill out a form consenting to my own use of the party supplies, signing my name twice. Even if I presently wish to smoke, my "future self" might not consent to the health risks (see Kagan 1989: 214). But the possibility of fragmented agents won't help with this objection. Unless we can consent to our own choices *even in normal cases,* like intentional eye-pokes, the Symmetric view collapses. Without self-consent, nothing will prevent simple self-harms from ludicrously violating rights.

But there are good reasons why so many different philosophers, including Kant (*LE* 27:340, *MS* 6:314, 422), Stocker, and Kagan, are happy to grant that self-consent is coherent.[38] For one thing,

---

(see e.g. Baier 1958, Finlay 2007).

[37] Although I often hear this objection in conversation, I have not found it in print (though Haase 2014a gives an illuminating argument for the idea that *claiming a right* requires two parties; see also Moran 2018).

[38] For contemporary references, see fn. 35, above. When quoting Kant, I will use the Cambridge Editions; pagination and volumes will be from the Akademie edition; and I use these abbreviations: *GW* (*Groundwork for the Metaphysics of Morals*), *MS* (*Metaphysics of Morals*), *LE* (*Lectures on Ethics*), *CP* (*Critique of Practical Reason*).

it is just a verbal choice if someone wishes to define "consent" as interpersonal. The real issue is whether making decisions is *like* giving consent in ways that allow for the waiving of rights. And the two are clearly similar. A consenting party, like someone who has made a decision, has established their willingness; they might even express their consent in decisive terms ("Let's do this"). If anything, deciding involves something more than consenting. Consent is an input to a joint decision; actually *making* a decision solo is more analogous to a complete joint decision in which consent is taken up and acted on. In light of this, we seem to have a good rationale for saying that we can "consent" to our own actions—even in cases where we are mentally unified, like simple, synchronic self-harms.

But self-consent raises a deeper problem even in these cases. On a Symmetric view, I have rights against everyone, myself included, that they not poke my eyeball. These rights tend to give others a strong reason against poking; there is a threat of violating my right, since I might withhold consent. But there is no way that I can withhold consent from myself. (Unless I am acting unintentionally—one of the hard cases in §5, below.) No matter how I try to violate my own right, I will fail, because the very decision that would lead to a violation also issues a waiver. The upshot is that rights against oneself are *impossible to violate*. This can seem doubly strange. For one thing, some think it is essential to moral constraints that, unlike laws of nature, they can be flouted.[39] More important: I have no good *reason* to comply with a right that I can't violate, just as I have no reason to avoid a smart-mine that I know for certain I can't trigger. By this, I mean that the right doesn't affect what is choiceworthy or sensible to do, doesn't ground "ought" claims, doesn't count in favor of compliance. In this sense of "reason," rights against oneself aren't reasons. (Which fits the cases. If someone offers me $500 to poke my own eye, my rights shouldn't count against taking the deal.)

[39] One example is Helen Beebee. She writes: "it seems to me to be essential to the idea of a moral law that it is breakable. It's hard to see how something could count as a moral law—or a rule of cricket, or whatever—if nobody was capable of breaking it" (2000: 581).

And yet most ethicists think that rights must be, perhaps by definition, a kind of reason for action.[40]

Now, in my view, *some* rights against oneself do give us reason to comply, because they can be straightforwardly violated. I have in mind unwaivable rights, like my right not to be pointlessly maimed. But set these aside for now. The Symmetric view is also committed to waivable rights, like my right not to have my eye poked, so the view has to respond to the strangeness head on. If waivable rights against oneself can't be violated—and therefore can't be reasons—does that make them intolerably weird?

The first thing to say is that everyone can agree on the easy point: rights are *generally* going to give agents a reason to comply. Other things equal, we should cough up what we owe. The real question is whether there is room for principled exceptions. I think there is, and once we understand how the exceptional rights work, we will also see why they have eluded notice.

Consider an analogy. Fragile things are *generally* going to shatter if whacked; they are disposed to break upon impact. But dispositions can be hidden. Imagine that a fragile vase is hooked up to a device—called a "fink"—that senses incoming projectiles, and that makes the vase invincible when struck (but not before or after). This vase is still fragile; it's still made of the same breakable material, still disposed to break. But the conditions that would normally cause the disposition to manifest are *also* conditions in which the manifestation is prevented. In this sense, the vase's fragility is *finkish*

---

[40] In *The Realm of Rights*, Judith Jarvis Thomson assumes that rights amount to a "behavioral constraint," cashing out in facts about what the person bound ought to do, other things equal; and though Thomson (1990: 34) denies having any argument for this claim, she calls on it repeatedly (1990: 2–3, 34, 64–7, 69, 77–8, 85–6, 98, 120, 123, 149, 165, 174–75, 197–98, 200–02, 208, 214–15, 221–22, 224, 227, 232, 252, 269–71, 292, 340). Hills (2003: 131) open her discussion by asking, "When should reasons for action be classed as 'duties to the self'?"—presupposing that duties are reasons. Schofield (2012: 10) makes the same presupposition in his argument that duties to self are reasons of a special kind (viz. "second-personal" reasons to protect one's own "interests and autonomy"). Reath (2006: 241) builds the same assumption into a definition of duties-to: one owes a duty to "the person who, under the circumstances, is the source of reasons for one to act." Fruh (2014: 65) is less strident, but insists that we "expect obligations to be weighty things that well-reasoning, conscientious, non-akratic agents will have to make room for in their deliberations"—i.e. reasons.

(Martin 1994, Lewis 1997). The disposition's triggers double as silencers.

A similar thing is going on with my right that I not poke my eye. This right is disposed to make self-harm wrong, but the disposition is hidden, because when I decide on the harm I consent to it and waive the right. My waivable rights against myself are in this sense *finkish rights:* the decision to do what would normally violate them also waives them, so that they are out of the way in time. That is why they are exceptions to the rule of thumb that rights are reasons.[41] Finkish rights don't "manifest" to make our actions wrong. And this, in turn, makes finkish rights rather elusive, which would help explain why no one ever noted them as exceptions.

My view is that rights against oneself are normally finkish, and that finkish rights make good sense. But let me be clear: there are critical exceptions where we might infringe our rights. We do not always consent to our own actions, as in the case of accidental harms. And even when we consent, that might not be enough for a waiver. Sometimes the right is too important to waive, at least not for trivial purposes. Other times, the right is in principle waivable, but our consent is undermined, because it is not informed and competent.[42] These are the cases in which our rights against ourselves aren't finkish; they are also the main hard cases wielded against the Self-Other Symmetric view.

---

[41] This principle—that reasons cannot evaporate when one acts against them—in effect requires that reasons have a stable normative upshot under different outcomes of deliberation. For some exploration of this requirement, giving a more detailed argument that finkish rights aren't reasons, see Chapter 3 (FRP).

[42] Consent is valid when it is informed, competent, and voluntary. (I won't discuss the voluntariness condition.) This view has been the norm in bioethics since the Nuremberg Code; it is the heart of Faden and Beauchamp's (1986: 155) conception of valid consent as "Autonomous Authorization"—which is itself the "prevailing view" among bioethicists, according to Miller and Wertheimer (2009: 80–1). A similar view is true of interpersonal consent in legal contexts. Faden and Beauchamp (1986: 38) report that, according to "well-established case law in such diverse areas as tort, contract, and fourth-amendment search-and-seizure law, a person can effectively waive a legal right only if the waiver is informed, reasoned, and voluntary."

## 5. Hard cases: irrationality, ignorance, intoxication

Now that we have dealt with conceptual objections to the Self-Other Symmetry, we turn to some hard cases, which are supposed to be more stubbornly Asymmetric than simple harms. In these cases, the conditions for valid consent are not met, and yet an action done to oneself may appear morally fine—or at any rate, less wrong than doing the same to others.

Slote gives the example of consent marred by *irrationality:*

> If someone irrationally asks me to harm or kill him, it will presumably be irrational and wrong of me to kill him, more wrong at any rate than if I irrationally choose to kill myself; yet the consent seems equal in the two cases. (1984: 191)

But *why* should irrational self-harm be any less wrong? It can't be that the self-harmer alone is blameworthy. Both agents who do harm are so irrational that they can't give valid consent; that's enough discombobulation to count as an excuse for doing wrong, to make blame inapt.[43]

Moreover, there is a good reason for thinking that rights are infringed in the case of self-harm, even if the agent is blameless. Recall that rights are enforceable in two ways. Besides sanctions like blame, onlookers might have the *prerogative to prevent violations* by means that would otherwise wrong the aggressor. If I stumble on two irrational hotheads who have agreed to poison each other, causing many hours of intense pain, I may interrupt their plans even if that means holding them down or knocking them out. I have a prerogative to stop them from infringing their rights against harm—even if they are blameless! This gives us a new test for the presence of rights. Do we have *permissions* to prevent an irrational self-injury, as we have permissions to prevent the irrational harmings of irrationally consenting others? Apparently, we do. Imagine that the two have decided together that each will poison *himself*. Intervention still seems permissible; I may still hold them down or knock them out without wrongly violating their rights. The finer details of whose poison

[43] I assume that, by "irrational," Slote means "incompetent to consent," not just "foolish." If he does mean "foolish," then the present case is a minor variant on the case of suicide I discuss in §6, below.

goes in whose body don't make a fundamental moral difference; what matters is the nature of the harm and the quality of consent. Hold those equal, and enforcement prerogatives are equal, too.[44]

Enforcement prerogatives may also be present when consent is absent due to *ignorance*. Here is how Michael Stocker puts the case:

> …even if because of ignorance nothing happens which can be taken as waiving a right, causing oneself pain is not wrong. (1976: 213)[45]

Indeed, accidentally causing oneself pain doesn't feel wrong. But what about accidentally causing someone *else* pain? Why should that be *more* wrong? Imagine that you have just developed a fairly nasty peanut allergy, and nobody knows it. As we sit down for breakfast, I put some peanuts on our bowls of oatmeal; you take a bite; and the result is that you endure serious pain for several hours before finally stabilizing. Was it *wrong* for me to put peanuts on your oatmeal, knowing that you would take a bite, but unaware of the risk to your health? The case has the same profile as irrational harm with irrational consent. I'm not *blameworthy* for the harm; at the time, the peanuts seemed like a nice touch. And yet my action tests positive for a rights infringement, because people have the *prerogative* to stop me (by force—e.g. grabbing my hand or throwing something at me—if there isn't a gentler way to get my attention). Again, one of the hallmarks of a wrongful violation is present; another is absent.

The same is true, I think, in the parallel case of unwitting harm to self. When *you* are the one who sprinkles the peanuts, and who is about to cause your own reaction, you don't deserve any

---

[44] Someone might object to this test: "Enforceability doesn't always signal a rights-infringement, since we may also stop 'impersonal' wrongdoers, who do wrong without infringing rights (e.g. by spoiling natural wonders)." A fair point. But in our cases, the wrongness presumably won't be impersonal, since the only decisive wrong-makers around are rights.

[45] A second case: "And even if a person asks us to cause him or her pain, it may be wrong for us to do so. And here, at least as much has been done to waive the right against us as we do *vis-à-vis* ourselves when we deliberately cause ourselves pain" (Stocker 1976: 213). But if a pain is so awful that it's wrong to inflict even on request, why should self-infliction be permissible? Stocker's dismissal is too quick to be persuasive.

blame, but you could permissibly have been stopped by force.[46] Whether the harm falls on self or other, the case features enforcement prerogatives without blameworthiness.[47]

But *pace* Stocker, we do sometimes deserve blame for accidentally harming ourselves; ignorance is no excuse when one is *negligent*. Consider a twist on our last example: it's common knowledge between us that you have a serious peanut allergy, and yet, carelessly (though not maliciously), I put peanuts on your oatmeal. As before, my action reflects no ill will and may nonetheless be prevented by force. But this time what I do is blameworthy; you even have grounds to resent my carelessness, taking it personally. The same goes, I think, for negligent harm to self. If you negligently add the legumes, it makes perfect sense that you might resent yourself. Others could also blame you, thinking that you were careless in way that is unacceptable.[48] I myself have felt this way about friends drifting into self-destructive lifestyles.

Speaking of which, we should also consider agents who are stupendously *intoxicated*. Like irrational agents, drunk people aren't always in command of their faculties, and in extreme cases, they are not competent to give valid consent. But wrongs done to drunk people, you might think, are the most Self-Other Asymmetric cases so far. The most striking cases involve sex. There is an

---

[46] Of course, you can't intentionally stop *yourself* from *accidentally* infringing your own right. (Ignoring fragmented agents.) But I assume that others can enforce your rights just as vigorously as you could. (The exception: when you object to being defended by others—again, something that only a psychically split agent could do *vis-à-vis* herself). For a defense of this view, with a focus on rights against harm, see Parry 2017.

[47] Interestingly, we find the same intuition in the work of John Stuart Mill, who is sometimes cited as a skeptic about duties to oneself (e.g. Denis 2001: 4–5). As Saunders puts it, "Mill says that someone can be prevented from crossing an unsafe bridge, though they are the only one at risk of harm. True, this should only be long enough to warn him of the danger; once he is aware of it, Mill insists that he should 'not forcibly [be] prevented from exposing himself to it" (Saunders 2016: 1012, Mill 1859: 294). This is how I imagine the case of the peanut allergy. We may not stop people from moderately hurting themselves *if* they consent.

[48] To be sure, your friends might have other reasons not to wag their fingers; perhaps your pain is enough to teach you a lesson, or maybe the root cause of your negligence was a feeling of worthlessness, which blame would only aggravate. Similar reasons could also count against blaming those who negligently harm others; there isn't any Self-Other Asymmetry lurking in here.

ongoing campaign across American universities to persuade students that it's wrong to have sex with very drunk people even with consent. There is not, to my knowledge, any comparable effort against drunken masturbation. If bibulous consent is invalid, Symmetry predicts that it's invalid whether given to oneself or another. And yet sexual touching seems like a rights violation only when done to an intoxicated other, not oneself.[49] Is this a counterexample?

Again, I don't think so, because the self/other distinction stops mattering when we equalize other factors. Here are a few of the main ones. First, there is a serious risk of *misunderstanding* someone else when they (and we) are drunk; we don't have the same access to their feelings and thoughts as we do to our own, so we might not realize that we are pressuring them, and might not know that they are reluctant or confused. The risk is greatest if the other person is a stranger. But the risk is minimal, and consent seems more valid, when the drunk parties are trusting and intimate, like an older married couple in touch with one another's feelings. (Notice that there isn't a campaign to get happy couples to abstain from drunken sex!) Second, having sex, even with verbal consent, may psychologically *harm* or distress someone else, especially if there is miscommunication or regret, whereas we tend to assume that this can't happen with self-gratification. Maybe it can: imagine a monk who deeply values chastity, and who would feel horribly violated if he were not to remain abstinent. Probably he shouldn't do anything sexual while seriously drunk, and should instead make that kind of decision when he has his wits about him. Third, it's useful to have a social *rule* of treating drunk interpersonal consent as categorically invalid, even if it sometimes isn't, because alcohol is the most commonly used date rape drug in the United States.[50] We have no such use for

---

[49] Some religious traditions frown upon drinking and masturbation, alone or in any combination. That might be too priggish. But the alleged problem for Symmetry is more disturbing. The worry is that these two are severally permissible but together tantamount to rape. (I am discussing this objection because—if you can believe it—it is among the most common responses I receive when presenting my view.)

[50] See this fact sheet from the Office of Women's Health, URL = <https://www.womenshealth.gov/a-z-topics/date-rape-drugs>.

any policy on drunk masturbation.

But above all, the key factor, I think, is that since we are so familiar with ourselves, we don't need fresh consent every time we use our bodies. Whenever a person is intimate and trusting with someone, as in the intimate couple, some rights are *waived by default*: The same goes, I think, for the ultimate intimates—you and yourself. Just as intimate couples have tacit agreements, you have made plenty of standing decisions. This makes your relation to your own body quite unlike your relation to a wary stranger's.

So I think even intoxicated consent turns out to be Self-Other Symmetric; we just have to make sure that the two-person case involves equal chances of harm and miscommunication, and that we aren't confusing nice rules and default statuses with independent moral fact. If the act is known to be safe, then even quite a bit of intoxication doesn't undermine consent (think of the drunk married couple). If it is risky—which it typically isn't in the solo case—then I suggest that we will be less inclined to say that the choice is morally neutral (think of the traumatized monk).[51]

I conclude that cases of absent and flawed consent are far from "minimal pairs," cleanly proving a Self-Other Asymmetry. The more carefully we look, the more Symmetrically our intuitions line up.

And we are not just looking in random directions. My responses have been following a simple, two-step method, which we can reuse on further might-be minimal pairs. Step one: *refine* the cases. Fill in missing details, if the examples are too skeletal. (When I choose death, am I rational or incompetent? When I trigger your peanut allergy, am I negligent or just unlucky? Is the drunk sex between teenager strangers after prom or a happy couple on their 35th anniversary?) Sometimes a detail in one case is omitted in the other—like the patient's consent. Here we need to equalize; either

[51] The monk's decision may be like the irrational choice to self-harm: blameless, but stoppable by force.

make both patients consenters or neither. Step two: *revise* initial hunches. With the refined cases in view, consider whether the action done to oneself still strikes you as morally special. The new details might incline you to think that the solo case is more problematic than expected (as in negligent harms), or that the two-person case is more innocent (as in unlucky accidents).

Of course, there is no guarantee that the "refine and revise" method will always work; maybe the real counterexamples are still out there. But when it does work, the point isn't to brutally bend our intuitions to the examples. It's supposed to be a process of discovery. Through tweaking and equalizing and reconsidering, we unmask the Self-Other Asymmetry in each pair as some asymmetry among the factors we already understand—the patient's consent, the agent's incompetence, the risk of harm. On a Symmetric view, these are the things that really matter in the realm of rights; the focus on "selves" is a distraction.[52]

## 6. Hard cases: unwaivable rights

For our final hard cases, we turn to rights that can't be erased even by the soberest and savviest agents: *unwaivable* rights. These are as far away from finkish as it gets.

A right is unwaivable, in a context, if it can't be suspended by consent, no matter how

---

[52] Slote has another minimal pair:

> … if I can avoid either an enduring pain to myself or a short-lived pain to you, you and I might both agree that it would be foolish of me to prevent the shorter pain to you; judging the matter objectively, you might not consent to my taking the longer pain upon myself in order to save you from the shorter pain. Yet there would be nothing morally wrong…in such a sacrifice. But when positions are reversed and I can avoid a short-lived pain to myself or a longer one to you and it is morally right that I should do the latter, you will presumably not consent to my doing the former and it will be wrong if I do so. (1984: 191)

The main problem with these cases is that they are not equalized. When I take on the big pain to spare you the little pain, *the person with the bigger pain* consents. Not so when I allow the big pain to befall you in order to spare myself the small pain. (Another problem is that Slote's cases are so abstractly described that it's unclear why anyone should have a right not to endure the big pain.)

competent, informed, and voluntary. Even with ideal consent, I violate a right if I kill a healthy, innocent person for no good reason. (Perhaps it would be fine in a context where there is a good reason to kill, e.g. because studying their viscera would help cure a terrible disease.) Another example: according to liberals (like Mill 1859: 300), consenting to slavery doesn't waive your rights against your "master." (Libertarians think we have broader powers to waive.)

Unwaivable rights—granting that they might exist—are the basis for another kind of Self-Other Asymmetric minimal pair. Consider the right not to be killed:

> We call a soldier a hero who throws himself on a grenade to save his battalion, but there is something ghoulish, for instance, about a man who throws a quadriplegic radio operator on a grenade, even if the radio operator begs him to do it. (Elliott 1992: 4)

Self-killing is "heroic" and other-killing is "ghoulish," even holding fixed the martyr's consent. This case isn't quite what we are looking for, however, since it's not clear how ghoulishness hooks up to rights and wrongness. Elliott's other case suits us better. Suppose a patient offers to give up his heart for a transplant, and a doctor is required for the procedure:

> …the puzzle is why I, for instance, might have the dual intuitions that in this case it is morally praiseworthy for a person to offer to donate his heart, and that it would be morally wrong for me to assist him. (Elliott 1992: 1–2)

That's the key intuition. If the patient could somehow transplant his own heart, killing himself in the process, that would be fine, but he cannot validly authorize a doctor to perform the transplant, even with competent consent. The right is present in the two-person case, but appears absent solo. Is this the killer pair? Or can we refine the cases and revise our intuition?

Fortunately, the cases leave plenty of room for refinement, with at least three crucial factors to equalize. (1) We need to make sure that the agents in both cases have equally excellent access to the donor's state of mind—whether the donor is sincere, competent, and so on. It is hard to imagine that a doctor could have such access; perhaps that is why we have the intuition that the doctor ought to refrain from operating. (2) We need to factor out any special obligations that the doctor has that

the patient doesn't. Perhaps the doctor owes it to every patient, donor *and* recipient, to value their needs equally, and this tells against the heart transplant (Elliott 1992: 9). Again, while this might be a great reason in practice against performing the surgery, it's no proof of a Self-Other Asymmetry. Finally, (3) as with the case of intoxicated sex, we have to distinguish the intuition that something would be a *good social rule* from the intuition of intrinsic moral *oomph*. Perhaps we don't want doctors to be cavalier about consent; we think that they should err on the side of caution when evaluating their patients, and that this would make a good norm. If so, this might be coloring our intuitions. We should stipulate that no such norm is in place in our cases.[53]

My revised intuition is that the patient can, in principle, authorize the doctor to perform the transplant, waiving his rights against her. The right isn't really unwaivable in either case. For completeness, we should also consider what happens when we adjust the harm/benefit ratio. What if the surgery is lethal torture and the benefits to the recipient are barely perceptible? My sense is that the doctor who performs the procedure with consent is a murderer, and that the patient who does this to himself is doing something morally wrong. Other things equal, both agents are equally blameworthy if they lack an excuse, and both are equally liable be stopped by force. If the patient's right against the doctor is unwaviable—and only if it is unwaivable—so is the corresponding right against the patient himself.

There is also an a priori reason for expecting unwaivable rights to be Self-Other Symmetric.

---

[53] We also need to rule out the idea that the doctor's motives are impure, which might cloud our intuitions. To this end, Elliott imagines that the operation is also costly to the doctor.

> Our moral evaluation of the physician who performed a heart transplant would also change if the physician underwent a sacrifice - say, if she risked catching a fatal infectious disease by operating. (1992: 7)

But obviously (as Elliott would agree) the addition of a harm to the doctor won't change whether she violates the patient's rights. The point of the case, as I see it, is to remove the "ghoulish" vibe of an agent who seems blasé about her duties to others, and who would help others with a sacrifice she wouldn't make herself.

If I may kill myself though you may not kill me, even with consent, then I am authorized to do things to myself that I cannot possibly authorize others to do on my behalf. Why should that be? How could my own authority, vis-à-vis myself, outstrip my power to authorize cooperating others? Authority isn't a liquid that gets diluted the further it spreads. The Symmetric view avoids this puzzle; it posits just one mechanism, consent, by which I can equally authorize myself and others.

So I suggest that the right not to be killed is Self-Other Symmetric. In carefully built cases, I think, we have permissions to kill consenting others, just as we have permissions to take our own lives—but this is just my intuition. To be clear: even if you disagree, you can still accept Symmetry. Maybe you think that killing others is always problematic even with consent. Maybe you think the choice to commit suicide is always permissible, even when there is much to live for. Symmetry just asks that you be consistent, at least in cases that are truly equalized, whether the person killed is the agent herself or another. Once you refine, which way to revise is up to you.

And with that, we have finished treating the hardest cases for the Self-Other Symmetry. No doubt there are scores left to puzzle through, and of course we should be open-minded in extending our defense to other kinds of rights—privacy, autonomy, property, etc.[54] Still, I hope to have shown that the *modus operandi* of Symmetry's critics—blurt out a case in one sentence, announce an intuition, vanish—isn't going to cut it anymore. Symmetry is shockingly stable even when pelted

[54] Rights to autonomy and privacy raise a special worry: they seem *essentially interpersonal*, in the sense of forbidding acts that can only be done to others—acts like snooping and eavesdropping, interference and domination. Doesn't that rule out privacy and autonomy rights against oneself? My response: privacy rights can be seen as forbidding a cluster of ordinary acts—like installing spyware on a laptop, looking down a person's shirt, or combing through sensitive documents—which can be done to one's own body and property. The same point holds for autonomy rights, which forbid chaining people up and locking them in boxes, no matter whether the victim is oneself or another. This response, by the way, doesn't require us to *reduce* autonomy and privacy rights, in the way that Thomson (1986) [1975] reduces privacy rights (to Scanlon's (1975) chagrin). Maybe autonomy has a *sui generis* value; maybe privacy rights can't be reduced to prior rights against ogling and listening. No problem. All Symmetry requires, to avoid this problem, is that the *acts* forbidden by privacy and autonomy rights be potentially reflexive. On duties of privacy to oneself, see Allen 2013, 2016.

with examples, and it elegantly delivers a lot of right answers to the questions of whom we may blame and whom we may protect by force.

## 7. The paradox of duties to oneself

By deciding on an act, I consent to it. That is how Self-Other Symmetric views explain why we so rarely infringe our many (finkish) rights against and duties to ourselves; we consent, so that we are released just in time. But even if self-consent makes sense, the idea that we *release ourselves from moral duties* might seem fishy. If I can shirk a duty at will, in what sense could it bind me? Aren't obligations supposed to be the kind of thing one can't just wiggle out of?

These questions are at the heart of Marcus Singer's paradox for duties to oneself, which has received more attention than any other modern work on the topic—and rightly so. It is the clearest, gravest threat to any theory of duties to oneself, including the Self-Other Symmetric view. We are going to need a solution.

Singer's argument has three premises. In his words:

(4) If A has a duty to B, then B has a right against (or with respect to) A;

(5) if B has a right against A, he can give it up and release A from the obligation; and

(6) no one can release himself from an obligation. (Singer 1963: 133)

Which together entail that there are no duties to (or rights against) oneself.

To get out of the puzzle, we have three basic options. First, follow Singer and give up duties to oneself—not what we're after.[55] Second, give up on the duties-rights link, or the rights-release link, and rest content with a defense of *sticky* duties to self, i.e. the ones from which the agent can't

[55] Even Singer grants that duties to oneself seem "well embedded both in traditional moral philosophy and ordinary moral thinking" (1959: 202). His view is costly, but he thinks we have to pay the price. For more skepticism about duties to oneself, see Baier (1958: 215; 231), as well as Frankena (1969: 692) (who is skeptical that they are "moral" duties). For some valuable criticism of Baier's arguments, see Neblett 1969. See the appendix for more references.

be released. This is the most popular solution by far.[56] But it's not an option if we want Self-Other Symmetry, since most duties to others *aren't* sticky: we can be released from promises, as well as duties not to harm ("Go ahead, Doc"), trespass ("Come on in!"), and use others' property ("Just give it back later"). Again: Symmetry says that if we have these duties to others (and they correlate with rights), we also have them to ourselves (along with corresponding rights); and if others can release us, then so can we release ourselves.

What we need—our third and final option—is to exonerate the power of self-release. Well, what are the charges? What's so bad about releasing oneself from a duty? Here is Singer:

> It is essential to the nature of an obligation that no one can release himself from an obligation by not wishing to perform it or by deciding not to perform it, or, indeed, *in any other way whatsoever*. In other words, no one can release himself from an obligation, just as no one can release himself from a promise. (Singer 1959: 202, emphasis added)

These are his only two objections; they're a bit thin.[57] First is an analogy with promises to oneself, which he takes to be incoherent. But this isn't persuasive. The supposed problem with self-promises is *that they would give us powers of self-release*, and we still haven't been told why these powers are impossible.[58]

The real sticking point, for Singer, is that it is "essential to the nature of an obligation" that we can't release ourselves from one—no matter what. Presumably, the worry is that obligations

---

[56] Kading 1960: 156, Wick 1960: 161–2, 1961: 216, Meiland 1963: 107, Jones 1983: 174, Denis 1997: 335, 2001, Hills 2003: 131, Reath 2006: 236, Timmermann 2006: 516, Schofield 2015. For an illuminating critique of Singer's second premise (the rights-release link), see Rosati 2011: 128–132; see also Cholbi 2015. ("Sticky" is my word.)

[57] Elsewhere Singer repeats his point without addition: "a duty to oneself…would be a duty from which one could release oneself at will, and this is self-contradictory" (Singer 1959: 202–3). Kant, too, claims without much argument that a "contradiction" follows if we say that "the one imposing obligation (auctor obligationis) could always release the one put under obligation (subiectum obligationis) from the obligation (terminus obligationis), so that (if both are one and the same subject) he would not be bound at all to a duty he lays upon himself" (MS 6:417).

[58] Singer (1959: 203) has no other objection to self-promises, though he does suggest that our talk about self-promising could be interpreted as metaphorical; we use it to "express a settled determination" to act.

would be pointless if they could be escaped at will; besides inescapable binding, they have no other upshot. But we have been studying just such upshots this whole time. Why do we have the prerogative to stop people from harming themselves when they are ignorant, incompetent, or intoxicated? Answer: their rights haven't been fully waived, because their self-consent isn't valid. Why does it make sense to resent oneself after negligent injury? Answer: again an unwaived right—this time, with culpability.

The whole theory of wronging oneself, in my view, comes from self-release—in particular, from the idea that we don't always do it right. We shouldn't just accept that certain acts express a lack of self-respect, end of story. We can *explain* why various awful things wrong oneself by showing that they would wrong a cooperating other, whose consent would be missing or inadequate; we can also explain why, and when, we have the paternalistic prerogative to stop those who are about to harm themselves.

Singer is wrong that self-release would ruin self-obligation: even finkish rights against oneself have an explanatory purpose. At most, we can admit that self-release does make these rights a good bit less restrictive. But that doesn't make them incoherent: it makes them exactly right—not too binding, not too toothless—for explaining the elusive nature of duties to oneself.

## 8. Conclusion

The Self-Other Symmetric view of rights is almost universally dismissed. Singer says it is incoherent; others say it is coherently ridiculous. I have argued that, however strange, it's true. We have a great many rights against ourselves that we are constantly waiving. Moving limbs, putting food in our mouths, taking coins from our pockets, broadcasting secrets, nail-clipping, teeth-brushing, undressing, perusing emails, getting arm tattoos and caffeine buzzes, putting socks on and contacts in. We have rights against everyone, including ourselves, that th ey not do these things to us. But just

as we can authorize others via consent, we authorize ourselves by making decisions. Our rights against ourselves are therefore finkish—waived by the same decisions that would otherwise lead to violations—and that is why they do not restrict us like the rights of unwilling others.

But our rights aren't *perfectly* finkish; in special cases, where self-consent is absent or invalid, it is possible to infringe rights against oneself. But when is self-consent invalid? What happens when it is absent? Here is where Symmetry starts to shine. The key insight is that we can reuse what we already know about the limits of *interpersonal* consent, since Symmetry predicts that limits of what we may do to ourselves just are the limits of what we may do to cooperating others relevantly like us. What we know about rights in general applies to the case of rights against oneself, which in turn—if we are careful in constructing cases—allow us to explain and explore our duties to ourselves, without giving up the compelling picture of morality as impartial.

This gives the Symmetric view an edge over its two rivals. Singer's "nihilist" view, which rejects all duties to self, can't explain our prerogative to stop people from innocently harming themselves (e.g. Mill's case of unwitting self-harm); and, of course, it cannot even allow for, much less explain, cases of self-wronging (e.g. negligent self-harm). There is also the "exceptionalist" view that we have some duties to ourselves that are somehow fundamentally unlike duties to others. If the view rejects rights against oneself, it will also have trouble explaining enforcement prerogatives. It can allow for self-wronging, which is an improvement over Singer, but it still can't explain *why* we wrong ourselves when we do. There is nothing like the Symmetric explanation that self-wronging happens when self-consent is compromised. Symmetric rights give deeper explanations.

But what's most exciting about rights against oneself, I think, is that they suggest a new project for moral theory. The old utilitarian project is to "expand the circle." The presumed starting point is that people care only about themselves, then ethicists implore them to extend that care to family, then friends, then the nation, then all of humanity, then all sentient life. First they see

themselves as the center of the universe—then another cosmic dot. Self-Other Symmetric rights take us on a different journey, starting not with care for self, but *respect for others*. In better moments, we naturally see neighbors and strangers as independent beings whose decisions, even when imperfect, carry real authority; we often take pains to respect others' wills, and we insist that those around us respect the wills of one another. What happens when we turn that respect inward, toward ourselves?

## Appendix

### A Legacy of Skeptics

What do philosophers think about duties to oneself? Unfortunately, there are few surveys available. Here I try to fill this gap in the literature, focusing on contemporary work, which I argue has been driven by a bipartisan desire to rule out rights against oneself.

On the issue of self-obligation, moral philosophers have been of two minds. There is a tradition of writers who defend some basic duties of self-respect: duties forbidding such things as servility, slavery, self-destruction, self-deception, and suicide. (Primmer moralists add other examples: gluttony and intoxication, prostitution and masturbation, lying to others, selling one's hair.)[59] Big figures here include Plato (see e.g. Socrates' denunciation of self-deception in the *Cratylus*, 428d); British Moralists like Richard Price (1974) and Bishop Butler (1900); contemporary Kantians like Thomas Hill, Jr. (1991a); and above all Kant himself, who proclaims that duties to self are "the most important of all," because whoever "violates" them "throws away his humanity, and is no longer in a position to perform duties to others" (*LE* 341).

But there is also a skeptical tradition that rejects all duties to self. Bernard Williams (1985: 182) calls them "fraudulent items," cooked-up counterweights to the needs of the collective. Aristotle (1138a 4–1138b 11) argues that there cannot be injustice against oneself, and later claims that suicide wrongs not oneself but the state; his reasoning is that the suicidal person "suffers

[59] In condemning "partial murder," Kant writes that "cutting one's hair in order to sell it is not altogether free from blame" (*MS* 6:423). There are some other questionable items on Kant's catalogue of duties to oneself, which mainly consists of the following: cultivating kindness towards animals and a love of nature (*MS* 6:563–64), cultivating one's talents (*GW* 423), "having religion" (*MS* 6:436), being honest with oneself about one's own worth or lack thereof (*MS* 6:562–63), being honest with others (*MS* 6:552–54), not committing suicide (*MS* 6:546–8, *GW* 422) even to save another's life (*CP* 266), not committing "partial murder," e.g. castration to improve one's singing or the removal of a tooth with an eye towards selling it (*MS* 6:547–8), chastity (*MS* 6:548–50), abstinence from gluttony and intoxication (*MS* 6:550–52), avoiding avarice (*MS* 6:555–57), and avoiding "servility," which for Kant involves, e.g., letting others violate your rights, taking favors, begging, flattering, whining, genuflection, and becoming poor (*MS* 6:557–59).

voluntarily," and "nobody suffers injustice voluntarily."[60] Echoing Hobbes (and Kant), Marcus Singer (1959, 1963) says that all duties to oneself are paradoxical; his argument is that they would entail corresponding rights against oneself, which are "surely nonsense," since they would give us the paradoxical power (as right-holders) to release ourselves from obligations. I am only *obligated* to pay debts and keep promises, Singer thinks, because I cannot get out of these predicaments at will; the other party, who holds a right against me, has to release me. But if I am the only "other," I am off the hook whenever I wish. So I cannot be obligated at all.

Singer's paradox, in particular, hit like a bomb—sensational, destructive, intimidating. The skeptics have dominated the debate ever since. Now most positive work on duties to self begins from the assumption that they are the suspects of paradox, and the modest goal is to eke out some nook where they, or an attenuated version of them, can be contained without causing too much embarrassment.[61] I don't think Singer's argument ultimately works (see §4, above). But it has been overwhelmingly successful in extracting concessions from opponents.[62]

The main response to Singer is that duties to oneself aren't "juridical" or "legal," like the duty to keep promises and stick to contracts, which means we never owe it to ourselves simply to *do* anything, but only to act from self-respecting motives. (Promise to pay me, and normally you owe

---

[60] Some scholars have a dim opinion of Aristotle's treatment of suicide, along with the rest of Section 11 of Book 5 in the *Nicomachean Ethics*; see Winthrop 1978: 1211, fn. 22.

[61] This is true of one big strand of the literature, where the focus is on whether we have any duties to oneself. In another strand, Kantian ethicists use duties to oneself to shed light on Kant's moral theory (Hill 1991a: 4 [1973: 87], Denis 1997, 2001, Reath 2006) and apply Kantianism to concrete moral problems (Jeske 1996 argues for Kantian duties to promote one's own happiness, with Bustos 2008 demurring; in bioethics, Chadwick 1989 considers duties not to sell body parts, Velleman 1999 critiques the idea of a right to die, Cholbi 2013 takes up the issue of paternalism, and Bauer 2018 examines chemical cognitive self-enhancement). A third theme is social critique: several writers argue that women and racial minorities owe it to themselves to resist being subordinated (e.g. Hill 1973 [1991a], Straumanis 1984, Hampton 1993), and Rawls (1999: 155–59) [1971: 178–82] argues that the value of self-respect supports his principles of justice.

[62] The next two paragraphs draw on material from Chapter One (PDO).

hard cash, not good intentions.)[63] The point of this move is to avoid linking duties to rights, in order to steer clear of rights against oneself; unfortunately, the move also makes us say that duties to self aren't fit to be *enforced* by external coercion, since we can't be coerced into having self-respecting motives. (When coercion works, the threat *is* the motive!) But if these "non-juridical" duties to self can't be enforced, in what sense are they even duties? Enforceability is arguably an essential property of any obligation (Hart 1955: 178). Without it, duties to oneself are so radically unlike pacts and debts that we might not want to count them as duties at all. They seem more like conscientious guides, nudging us to cherish and nourish ourselves out of genuine self-concern.[64]

The next response to Singer is that we do have duties to ourselves, but no one can release us from them (Kading 1960, Hills 2003, Timmermann 2006: 516, Schofield 2015).[65] Some writers arrive

---

[63] The first to make this response to Singer was Warner Wick (1960: 161, 1961); soon after, Mary Mothersill (1961: 207) noted that, on Wick's view, "duties to oneself…must be of a completely different order from duties of the familiar sort." See also Knight 1961 on "non-legal" duties, and Fotion's (1970: 460) response to Eisenberg (1968) on "social contractual" duties. Margaret Paton (1990: 225–26) argues that duties to oneself are "non-contractual," so that they don't entail corresponding rights or powers of release; she also concedes to Singer that one can't make a binding promise to oneself. Hills (2005: 138) takes a similar line on self-promising and rejects any "juridical" duties to oneself; she emphasizes that only juridical duties are enforceable, in the sense that others can hold us accountable for failures of performance. The main inspiration for these authors is Kant; see *MS* 6:383, where he claims that duties of virtue (which include all duties owed to anyone) can't be rightfully enforced, whereas "[all] legal duties can be coerced" (as he says in *LE* 29:632); see also the following passage from his *LE* 117: "My duty towards myself cannot be treated juridically; the law touches only our relations with other men; I have no legal obligations towards myself…." But I should note that the concept of non-juridical duties isn't original to Kant. It has roots in Calvin's "precepts of love," which can be fulfilled only by acting from good motives (see his Institutes of the Christian Religion: III.X.5), and in Hugo Grotius's (1925: 34–9, 75) "imperfect rights" and "laws of love," which aren't enforceable; for more on these writers, see Schneewind 1997: 79–80. I should also note that Kant's word for duty, pflicht, does not have the same legalistic connotations in German that "duty" has in English.

[64] The move toward "non-juridical" duties to oneself has been met with limited resistance. Hill (1991b), Habib (2009), Rosati (2011), and Fruh (2014) argue that we can be bound by promises to ourselves—though even then, Hill concedes that the duty is only "non-moral."

[65] Schofield (2015), drawing on Darwall 2006, argues that we owe it to ourselves not to cause self-harm just if the harm falls in the future, since we can't now release ourselves from our future perspective. His view is original and important, but as I argue in Chapter One (PDO), Schofield's view seems to entail that I must treat myself in the future like an unwilling other—so, e.g., I can't take a pill to cure my big headache now if it gives me a medium one later. I can't even cause myself

at this claim by denying that anyone can release *anyone* from obligations (Wick 1960: 162; Denis 2001: 229 fn. 13, 230). This is an extreme concession. (What else could I be doing when I let you out of a promise, or when I give my surgeon the go-ahead?)[66] But Knight (1961: 212) goes even further, denying that anyone *has* any obligations to anyone (only "to" ideals); and some Kantians suggest that obligations are owed only among parts of people—no *person*, strictly speaking, owes any *person* anything, though I can loosely be said to have duties "to myself" when one of my parts is obliged to another.[67] Then there are those like Meiland (1963) and Fotion (1965: 30) who retreat to the claim that we have self-*regarding* duties: duties that aren't necessarily owed to ourselves, but that do constrain how we treat ourselves; for example, the utilitarian duty to cheer oneself up is owed to no one; and I might owe it to a friend to keep my promise to stop smoking.[68] Sometimes the retreat is hailed as a triumph—Oakley (2017: 71) defends merely self-regarding duties and dubs his paper "Good News for Duties to Oneself."

So despite all the controversy over duties to self, the skeptics and the defenders share a deep presupposition: that we cannot have waivable rights against ourselves. Rights might be unwaivable,

---

future harms for my own (greater) future good! (The same problem afflicts Sartorius 1985: 245–6, notable for defending unwaivable rights against oneself; see also Jones 1983.)

[66] Mothersill (1961: 206) raises the question about promises in her response to Wick.

[67] "Thus the subject and the object of duties to one's self, though both part of the same human being, turn out not to be identical" (Timmermann 2006: 509). Timmermann (2006: 512) later seems to retract this idea. Kant himself warns that a paradox follows if "the I that imposes obligation is taken in the same sense as the I that is put under obligation" (*MS* 6:417). Schofield (2015: fn. 24) says that Kant's own solution (in *MS* 6:418) is to distinguish noumenal from phenomenal selves (or aspects), and hold that the noumenal self binds the phenomenal one. There is good evidence for this reading, but Reath (2006: 356) and Kerstein (2008: 205–6) argue that Kant's view requires that the noumenal self, or "homo noumenon," binds itself (see *MS* 6:239, 295 and LE 27:627–28). Kerstein (2008: 206), echoing Denis 1997: 334–35, suggests an interpretation on which the homo noumenon has real duties to itself because it is bound under one aspect (as "faculty of choice," or Willkür) and binding under another (as "legislative reason," or Wille). But as Kerstein notes, Kant himself doesn't mention these two aspects in his treatment of duties to oneself. Clearly, however we interpret him, Kant has left many readers wary of duties owed by and to whole people—but not all readers. Nelson Potter (1998: 375, fn. 8), for example, thinks Kant's split between selves is "irrelevant" to the paradox of self-release and is "not a helpful reply."

[68] For more on Meiland's view, see Mavrodes, Narveson, and Meiland 1963: 169–70.

unreal, decoupled from duties, directed solely at others. But whatever we owe to ourselves, we are not bound by waivable, Self-Other Symmetric rights. This is the consensus I hope to challenge.[69]

[69] Some interesting exceptions: Kagan (1989: Chapter 6, Section 2) considers the project of deriving prerogatives from waivable rights against oneself, and Hurley (1995) tries deriving them from unwaivable rights against oneself. For discussion, see Chapter Three (FRP).

**REFERENCES**


Allen, Anita L. (2013). An ethical duty to protect one's own information privacy? *Alabama Law Review* 64 (4): 845–866.

-------(2016). The duty to protect your own privacy. In Adam Moore (ed.) *Privacy, Security, and Accountability: Ethics, Law, and Policy*. Rowan & Littlefield.

Aristotle (1999). *Nichomachean Ethics, Second Edition*. Terence Irwin (ed., trans.). Cambridge: Hackett Publishing Company.

Baier Kurt, (1958). *The Moral Point of View*. Ithaca: Cornell University Press.

Bauer, Katharina (2018). Cognitive self-enhancement as a duty to oneself. *Southern Journal of Philosophy* 56 (1): 36–58.

Beebee, Helen (2000). The non-governing conception of laws of nature. *Philosophy and Phenomenological Research* 61 (3): 571–594.

Bustos, Keith (2008). Defending a Kantian conception of duties to self and others. *The Journal of Value Inquiry* 42 (2): 241–254.

Butler, Joseph (1900). *Fifteen Sermons Preached at the Rolls Chapel*, in *The Works of Bishop Butler, Vol. 1*, ed. J. H. Bernard. London: Macmillon.

Chadwick, Ruth F. (1989). The market for bodily parts: Kant and duties to oneself. *Journal of Applied Philosophy* 6 (2): 129–140.

Cholbi, Michael (2013). Kantian paternalism and suicide intervention. In Christian Coons and Michael Weber (eds.), *Paternalism: Theory and Practice*. Cambridge: Cambridge University Press. 115–133.

-------(2015). On Marcus Singer's "On Duties to Oneself". *Ethics* 125 (3): 851–853.

Darwall, Stephen L. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge: Harvard University Press.

Denis, Lara (1997). Kant's ethics and duties to oneself. *Pacific Philosophical Quarterly* 78 (4): 321–348.

-------(2001). *Moral Self-Regard: Duties to Oneself in Kant's Moral Theory*. Routledge.

Eisenberg, Paul D. (1968). Duties to oneself and the concept of morality. *Inquiry: An Interdisciplinary Journal of Philosophy* 11 (1–4): 129–154.

Engstrom, Stephen (1998). Deriving duties to oneself: comments on Andrews Reath's "Self-Legislation and Duties to Oneself". *Southern Journal of Philosophy* 36 (S1): 125–130.

Faden, Ruth R. and Beauchamp, Tom L. (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press.

Finlay, Steven (2007). Too much morality. In Paul Bloomfield (ed.), *Morality and Self-Interest*. Oxford: Oxford University Press. 136–154.

Fotion, Nicholas (1965). We can have moral obligations to ourselves. *Australasian Journal of Philosophy* 43 (1): 27–34.

-------(1970). II. Eisenberg and self-obligations. *Inquiry: An Interdisciplinary Journal of Philosophy* 13 (1–4): 458–461.

Frankena, William K. (1966). The concept of morality. *Journal of Philosophy* 63 (21): 688–696.

Fruh, Kyle (2014). The power to promise oneself. *The Southern Journal of Philosophy* 52 (1): 61–85.

Gilbert, Margaret (2018). *Rights and Demands: A Foundational Inquiry*. Oxford: Oxford University Press.

Haase, Matthias (2014a). Am I you? *Philosophical Explorations* 17 (3): 358–371.

-------(2014b). For oneself and toward another: the puzzle about recognition. *Philosophical Topics* 42 (1): 113–152.

Hampton, Jean (1993). Selflessness and the loss of self. *Social Philosophy and Policy* 10: 135–165.

Hart, H.L.A. (1955). Are there any natural rights? *Philosophical Review* 64 (2): 175–191.

Hill, Thomas (1973). Servility and self-respect. *The Monist* 57 (1): 87–104. Reprinted in Hill 1991a: 4–

18.

-------(1991a). *Autonomy and Self-Respect*. New York: Cambridge University Press.

-------(1991b). Promises to oneself. In Hill 1991a: 138–154.

Hills, Alison (2003). Duties and duties to the self. *American Philosophical Quarterly* 40 (2): 131–142.

Hohfeld, Wesley Newcomb (1919). *Fundamental Legal Conceptions*. New Haven: Yale University Press.

Hurley, Paul 1995. Getting our options clear: a closer look at agent-centered options. *Philosophical Studies* 78: 163–188.

Jeske, Diane (1996). Perfection, happiness, and duties to self. *American Philosophical Quarterly* 33 (3): 263–276.

Johnson, Robert N. (2010). Duties to and regarding others. In Lara Denis (ed.), *Kant's Metaphysics of Morals: A Critical Guide*. Cambridge: Cambridge University Press.

Jones, H. J. (1983). Treating oneself wrongly. *Journal of Value Inquiry* 17: 169–177.

Jones, Peter. (1999). *Human Rights Quarterly* 21 (1): 80–107

Kading, Daniel (1960). Are there really "no duties to oneself"? *Ethics* 70 (2): 155–157.

Kagan, Shelly 1989. *The Limits of Morality*. Oxford: Oxford University Press.

Kamm, Frances (1996). *Morality, Mortality Volume II: Rights, Duties, and Status*. Oxford: Oxford University Press.

Kant, Immanuel (1996). *Practical Philosophy*. Mary Gregor (ed.). Cambridge: Cambridge University Press.

-------(1997). *Lectures on Ethics*. J.B. Schneewind (ed.), translated by Peter Heath. Cambridge: Cambridge University Press.

Kerstein, Samuel J. (2008). Treating oneself merely as a means. In Monika Betzler (ed.) *Kant's Ethics of Virtue*. Berlin: New York: de Gruyter: 201–218.

Knight, Frank H. (1960). I, me, my self, and my duties. *Ethics* 71 (3): 209–212.

Lewis, David (1997). Finkish dispositions. *The Philosophical Quarterly* 47 (187): 143–158.

Martin, C. (1994). Dispositions and conditionals. *The Philosophical Quarterly* 44 (174): 1–8.

Mavrodes, George I.; Narveson, Jan & Meiland, J. W. (1964). Duties to oneself. *Analysis* 24 (5): 165–171.

Meiland, Jack W. (1963). Duty and interest. *Analysis* 23 (5): 106–110.

Mill, John Stuart (1859). *On Liberty*, in Robson 1977: 213–310.

Miller, Franklin G. and Wertheimer, Alan (2009). Preface to a theory of consent transactions: beyond valid consent. In Franklin Miller and Alan Wertheimer (eds.) *The Ethics of Consent: Theory and Practice*. Oxford: Oxford University Press: 79–106.

Moran, Richard (2018). *The Exchange of Words*. Oxford: Oxford University Press.

Mothersill, Mary (1961). Professor Wick on duties to oneself. *Ethics* 71 (3): 205–208.

Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

Neblett, William (1969). Morality, prudence, and obligations to oneself. *Ethics* 80 (1): 70–73.

O'Neill, Onora (2001). Duty and obligation. In Lawrence C. Becker and Charlotte B. Becker (eds.), *Encyclopedia of Ethics, Second Edition*. Routledge Press: 423–428.

Oakley, Tim (2016). How to release oneself from an obligation: good news for duties to oneself. *Australasian Journal of Philosophy* 95 (1): 70–80.

Parry, Jonathan (2017). Defensive harm, consent, and intervention. *Philosophy and Public Affairs* 45 (4): 356–396.

Paton, Margaret (1990). A reconsideration of Kant's treatment of duties to oneself. *Philosophical Quarterly* 40 (159): 222–233.

Pogge, Thomas (2002). *World Poverty and Human Rights*. Cambridge: Polity Press.

Potter, Nelson (1998). Duties to oneself, motivational internalism, and self-deception in Kant's ethics. In Mark Timmons (ed.) *Kant's Metaphysics of Morals: Interpretive Essays*. Clarendon Press:

371–390.

Price, Richard (1974). *A Review of the Principal Questions in Morals*. D.D. Raphael (ed.). Oxford: Oxford University Press.

Portmore, Douglas (2003). Position-relative consequentialism, agent-centered options, and supererogation. *Ethics* 113 (2): 303–332.

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

Reath, Andrews (1997). Self-Legislation and Duties to Oneself. *Southern Journal of Philosophy* 36 (S1): 103–124. Reprinted in Reath 2006: 231–249.

-------(2006). *Agency and Autonomy in Kant's Moral Thoery: Selected Essays*. Oxford: Oxford University Press.

Robson, John M. (1977). *Collected Works of John Stuart Mill. Volume XVIII: Essays on Politics and Society.* Toronto: University of Toronto Press and London: Routledge and Kegan Paul.

Rosati, Connie 2011. The importance of self-promises. In Hanoch Sheinman (ed.), *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press. 124–155.

Sartorius, Rolf (1985). Utilitarianism, rights, and duties to self. *American Philosophical Quarterly* 22 (3): 241–249.

Scanlon, Thomas (1975). Thomson on privacy. *Philosophy and Public Affairs* 4 (4): 315–322.

Schneewind, Jerome B. (1997). *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.

Schofield, Paul (2015). On the Existence of Duties to the Self. *Philosophy and Phenomenological Research* 90 (3):505–528.

Singer, Marcus G. (1959). On duties to oneself. *Ethics* 69 (3): 202–205.

-------(1961). *Generalizations in Ethics*. New York: Russell and Russell.

-------(1963). Duties and duties to oneself. *Ethics* 73 (2): 133–142.

Slote, Michael (1984). Morality and self-other asymmetry. *The Journal of Philosophy* 81 (4): 179–192.

Steiner, Hillel (2013). Directed duties and inalienable rights. *Ethics* 123 (2): 230–244

Stocker, Michael (1976). Agent and other: against ethical universalism. *Australasian Journal of Philosophy* 54 (3): 206–220.

Straumanis, Joan (1984). Duties to oneself: a basis for self-liberation? *Journal of Social Philosophy* 15 (2): 1–13.

Strawson, Peter F. (1962). Freedom and resentment. *Proceedings of the British Academy* 48: 1–25.

Thomson, Judith Jarvis (1975). The right to privacy. *Philosophy and Public Affairs* 4 (4): 295–314. Reprinted in Thomson 1986.

-------(1986). *Rights, Restitution, and Risk: Essays, in Moral Theory*. William Parent (ed.). Cambridge: Harvard University Press.

-------(1990). *The Realm of Rights*. Cambridge: Harvard University Press.

Timmermann, Jens (2006). Kantian duties to the self, explained and defended. *Philosophy* 81 (3): 505–530.

Wick, Warner (1960). More about duties to oneself. *Ethics* 70 (2): 158–163.

-------(1961). Still more about duties to oneself. *Ethics* 71 (3): 213–217.

Williams, Bernard Arthur Owen (1985). *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press.

Winthrop, Debra (1978). Aristotle and theories of justice. *The American Political Science Review* 72 (4): 1201–1216.

Wood, Allen W. (2009). Duties to oneself, duties of respect to others. In Thomas E. Hill (ed.), *The Blackwell Guide to Kant's Ethics*. Wiley-Blackwell.

Velleman, J. David (1999). A right of self-termination? *Ethics* 109 (3): 606–628.

# CHAPTER THREE

## From Rights to Prerogatives

## From Rights to Prerogatives

Act consequentialists think we always have to make things go best. There are two main ways in which deontologists disagree: they opt for rights, which may forbid doing what has the best outcome, and they posit prerogatives, which may permit suboptimal acts. There are powerful signs of a link between rights and prerogatives. But what exactly is the link? I argue that prerogatives just are a kind of right—not rights against others, but waivable rights against oneself. After responding to objections, I conclude that Self-Other Symmetry—the view that we have the same rights against ourselves as against others—is not just paradox-free, but explanatorily powerful. With Symmetry, we can reduce prerogatives to rights, giving deontology an elegant, impartial, and unified foundation.

### 1. Introduction

Traditional act consequentialists think we always have to make things go best. Ordinary morality disagrees in two ways: it says that (1) other people have *rights*, which forbid us from mistreating them even for the greater good; and that (2) we have *prerogatives*, which permit us to forsake the greater good even when others' rights aren't in the way.[70] Thanks to your rights, I can't take your life to save five others. Thanks to my prerogatives, I don't have to save five lives if it means giving up my own.

Rights and prerogatives seem deeply linked: interpersonal rights represent the limits of what we may take from unwilling others, and prerogatives set the limits of what we ourselves must give. There is also a known clue that the two are entwined: the failure of "hybrid" non-consequentialist views featuring one without the other. Consider the peculiar moralizing of a view with rights and no prerogatives. On W.D. Ross's (1930) view, we all have to promote the greatest good whenever it doesn't violate rights—when we can do best without breaking deals, legs, and laws. So even though

[70] Instead of rights, some talk more generally about *restrictions*, i.e. features that can make it wrong to promote the greatest good (e.g. Kamm 1996). There may be some peripheral restrictions that aren't rights—perhaps it's wrong to smash one natural wonder to save five—but we can ignore them, since they have no link to prerogatives. Also, I won't use the familiar term "agent-centered prerogative" (Scheffler 1992), which suggests that prerogatives are based in agent-relative value. I prefer the more neutral "prerogative," which just refers to *whatever* makes it permissible to do less than best.

you have a right over your cash and kidneys, you must give them up whenever others need them more, if only barely. The result is that Ross's rights are undercut: morality forbids others from seizing your stuff for the greater good, and then demands that you hand it over, anyway.

Or consider Scheffler's (1994) lax hybrid, which features prerogatives to promote self-interest but no rights. Say your prerogatives allow you to keep $10,000 for yourself rather than give it to charity and save a life. Then your prerogatives should also let you *take* such a life to *acquire* $10,000. After all, your interests matter more, and you don't have to worry about your victim's rights—they have none.[71] Worse, since *you* also lack rights, nothing protects others from denying *you* the chance to act on your prerogatives; others may, and sometimes must, kill you for their own good, steal from you to give to the needy, etc.—and if they can't themselves take your cash or life, they may have to coerce you into self-sacrifice. The result, again, is systematic undercutting. What's so great about prerogatives if others won't let you exercise them?

So we need rights to limit and protect our prerogatives, and we prerogatives in order to permissibly lean on our rights. Somehow, rights and prerogatives have got to be connected. But how? Do they share a source? Does one somehow explain the other? I argue that prerogatives can be *derived* from rights, but not in the ways one might expect. Prerogatives aren't rights against others; they aren't the absence of others' rights against us (§2). They are a much stranger item—waivable rights we have against ourselves (§3). Prerogatives can be explained given the simple, extraordinary assumption that rights are *Self-Other Symmetric:* that we have the same basic moral rights against ourselves as we do against others, rights that proscribe things like harm and bodily intrusion, and that can be waived by valid consent.

---

[71] This kind of objection to Scheffler, first made by Kagan 1984: 251, is now a classic; see Kamm 1985, 1996: 213, Kagan 1989: 19–24, Schmidtz 1990: 625–26, Scheffler 1992: 379–387, Das 2000, and Mack 2005: 379. Kamm (1992: 357–58, 2007: 16) gives a related objection: that Scheffler's hybrid allows us to run others' lives without their consent, so long as we care about their welfare.

---

How could rights against oneself ground a prerogative? One possibility, which I think is almost right, is embodied by Kagan's (1989) fairly outré *Self-Constraint Argument*, which assumes the Self-Other Symmetry of rights (§§4–5). The basic idea is that a waivable right against oneself is an optional restriction. By default, the right makes self-sacrifice wrong (and therefore not required), just as rights generally make it wrong to sacrifice the interests of unwilling others. But when a right is held by the agent *herself*, she has the option to waive it by consenting to the sacrifice, which she does, in effect, by deciding to make it. Self-sacrifice is optional because it is up to the agent whether her rights make it wrong.

A concrete example might help. Suppose that Amanda has a right against herself that she not crush her own arms, but crushing them is the only way to hold open an escape route for Bert, a stranger trapped in a collapsing building. If Amanda declines to save Bert, she isn't *doing* wrong; she is *avoiding* a wrong. Her rights function like the rights of an unwilling third party. It is as if she were choosing not to crush the uninvolved Clayton's arms against his will—an obligatory, and so permissible, omission. But what if Amanda chooses to save Bert? Then she effectively consents to losing her arms, and therefore waives her right. Now it is as if she had crushed Clayton's arms with his consent, and her action is perfectly permissible. Conclusion: whether she crushes her arms for the greater good or keeps them, Amanda doesn't act wrongly—so her sacrifice is optional. From a right, we derived a prerogative. Indeed, since it is the right that permits Amanda to keep her arms, there's a case for saying that the right *is* her prerogative.

The Self-Constraint Argument—pulled from thin air by Shelly Kagan (1989: Chapter 6, Section 2), who would immediately clobber it with objections—has received almost no attention from philosophers. Probably, that's because it relies on rights against oneself, which are routinely dismissed offhand (Thomson 1990: 42, Hill 1991: 4 [1973: 87], Kamm 2007: 235, 241) or declared

paradoxical (Singer 1959, O'Neill 2001: 48). But I will argue that rights against oneself make for a better premise than one might expect; the Argument is surprisingly promising (§3).

Let me also mention that the Self-Constraint Argument has greater *ambition* than Kagan ever imagined: it could give the theory of prerogatives its strongest basis. There are troubling cracks in the old foundations—the doctrine that prerogatives are derived from self-interest, magnified by an agent's "personal point of view" (Scheffler 1994). So construed, a prerogative is just a license to treat one's own welfare and projects as *eo ipso* more important. This kind of view, I think, puts the wrong limits on our moral freedom, since it implies that we must let others use our body parts and rightful holdings for the greater good so long as the use won't *harm* our interests, won't leave us or our projects worse off. Intuitively, we have some prerogative not to sacrifice our bodies and things even when doing so is costless or slightly rewarding. Even if I wouldn't miss my kidney, it's still *mine*, and sometimes that's enough to justify keeping it for myself. The best way to capture this intuition is by basing prerogatives not in interests, but in rights. If so, rights-based prerogatives are essential for understanding the source and scope of our moral freedoms.[72]

So I think the Self-Constraint Argument is exciting—but it's not quite sound. As we will see, it only works if our rights somehow make self-sacrifice *wrong*, but clearly they can't; self-sacrifice is optional (§6). To solve this "Wrongness Problem," we will have to show how rights could make a difference without being wrong-making reasons (§7). But we begin with the Self-Constraint Argument and its advantages over the more familiar paths from rights to prerogatives.

[72] Here are two more clues that prerogatives are a kind of right. Woollard (2015: 130) argues that prerogatives, like rights, weigh more when a sacrifice requires direct harm as a causal means than when the harm to one's interests is merely foreseen. For example, if you are threatened by a runaway train, I do not have to save you by putting my most prized possessions in its way. But if the train is coming for both you and my treasure, I have to save you, if I must choose one. Second, Saba Bazargan-Forward (2018) argues that prerogatives can be vested in proxies, just like rights.

## 2. Do we need waivable rights against oneself?

The Self-Constraint Argument tries to establish prerogatives—permissions to do less than best—by starting from waivable rights against oneself. But why is *that* the starting point? Can't we use some safer, less exotic kind of right? Surprisingly, no. Let's take a hard look at the "safe" options, starting with interpersonal rights.

A tempting thought is that prerogatives are just rights against others that they leave us alone, that they not force our hand or frustrate our actions (Benn 2017: 278). Such rights are called "protective claims" (Gilbert 2018) or rights "against interference" (Bolinger 2017: 43); I call them *interference rights* for short. Interference rights are nice; they forbid others from taking our choice into their hands, which grants us a kind of protection. But this protection doesn't work like a prerogative, since it doesn't make our choices any more *permissible*. Indeed, interference rights often protect actions that are wrong.[73] In the simplest case, someone starts out without a prerogative and gains the interference rights by agreement. Suppose we all promise not to stop members of a certain sect from feeding their children a harmful herb. Now the group has interference rights, but their actions are still wrong; they don't have a prerogative to harm their kids. (Another example: voting for a nightmare candidate in a legitimate democracy.)[74]

The other temptation is to think of prerogatives as the absence of rights against the agent—what is known as a privilege (Hohfeld 1919) or a *liberty* (Gilbert 2018). I have a liberty to eat the salad just if people lack a right against me that I not eat it. (I have the liberty "as regards" whoever

---

[73] This possibility is what Enoch (2002), Herstein (2012), and (arguably) Waldron (1981) have in mind when they talk about the "right to do wrong." See Bolinger 2017 for illuminating discussion.

[74] People can also have prerogatives without interference rights. In the simplest case, the agent starts with both and waives the interference rights, as when I say: "You're welcome to take my treasure—if you can find it!" I retain the prerogative to keep my loot, but I waive the right that you let me. (In §1, I said that prerogatives and rights are problematically "undercut" unless they go together. Let me clarify: they should *typically* go together. Exceptions are to be expected, since we know that rights can be changed by agreements.)

lacks the right.) Liberties are also nice, since they ensure that one's actions won't violate any rights. But liberties don't entail prerogatives. It's possible to do wrong despite having all the liberty in the world: some actions are *impersonally* wrong, wronging no one, because they are opposed by moral reasons with another source besides rights. Suppose I dump sludge onto an unowned natural wonder, like the moon. That's wrong; I have massive moral reasons not to ruin lovely things (Wedgwood 2013: 44, fn. 9).[75] But since no one owns the moon, I am at liberty (as regards everyone) to spoil it. Liberties prevent wrong*ings*, but not wrong*ness*.[76]

(The same goes for "protected liberties" (Gilbert 2018: 19), which combine a liberty with interference rights. If everyone promises me not to interfere with my lunar pollution schemes, my liberty to pollute is protected, but spoiling the moon is still wrong.)

Interpersonal rights, or lacks thereof, can't be the source of prerogatives. But before we turn to waivable rights against oneself, we have one last "safe" option: derive prerogatives from *unwaivable* rights against oneself. The basic idea, due to Paul Hurley, is that everyone is morally protected: an agent has "patient-protecting" reasons not to interfere with anyone, even for the greater good. When the patient is another person, that means the agent has decisive reasons not to kill one to save five, crush two arms to save a life, etc. But when the patient is the agent? Then the "patient-protecting" reasons demand the provision of a "protected sphere" of acceptable acts. "The agent must leave himself a range of morally permissible options" (Hurley 1995: 176).

---

[75] Owens (2012: 45) gives another example: concreting over the Grand Canyon. Impersonal wrongness may also show up in what Parfit (1984: 356) calls "different people choices," where our decision affects who exists. Suppose you are presented with a button. Push it, and you create an unreachable galaxy where trillions live fabulous lives. Don't push, and nothing happens. It's wrong not to push, but you have the liberty not to; those whom you don't create will never have rights. (Note that it's only wrong not to push if we have moral reasons to create happy people—a controversial claim. See e.g. McMahan 1981 on the Procreation Asymmetry.)

[76] There is also a simpler reason why prerogatives can't be liberties. Prerogatives need to have *weights*, since they can be outweighed by reasons. (If my kidney can save 10,000 lives, I must donate.) But liberties have no weights. They are just the absence of rights—and how could an absence be weighty?

Hurley deserves credit here for pioneering a Self-Other Symmetric view, but I don't think the machinery quite works. The problem is with "patient-protecting reasons" (his name for the reasons given by rights). What do these reasons favor? What are they reasons *for?* The answer is easy enough in two-person cases. The reasons forbid harming and harassing the patient (without consent); these are reasons against plain old acts like arm-crushing. But Hurley can't say that Amanda has strong reasons against crushing her *own* arms, otherwise it would be wrong, not optional! Hurley seems to think that her reasons favor a special act—permitting a sphere for herself—but it's unclear what this could mean. Amanda doesn't *decide* that she is permitted to keep her arms instead of saving Bert; she just *is* permitted. There appears to be no room for patient-protecting reasons vis-à-vis oneself[77,78]

What's going wrong with these proposals? Why can't we squeeze out a prerogative from privileges, liberties, and reasons to permit? The problem, I think, is that prerogatives have got to weigh against the agent's (otherwise compelling) reasons to make the sacrifice. Interference rights just ward off meddling; liberties just prevent wrongings; and "reasons to permit" only weigh against reasons not to permit. None of these can counterbalance the reasons to sacrifice for the greater good; so, none can stop those reasons from grounding a requirement.

What would do the job? What kind of right could weigh against the reasons for self-sacrifice? We need something that bears on the choice to sacrifice: a right *against the agent* that she not make it—e.g. a right against the organ donor that he keep his kidney, a right against the

---

[77] At times, Hurley (1995: 176) seems to say that rights against oneself aren't reasons *for agents to act*, but only "reasons" in a broader sense; they weigh against moral demands (see also Kagan 1989: 208). But then they aren't symmetric with rights against others, which *are* reasons for and against plain old actions.

[78] Frances Kamm (1992: 358–59, 1996, 2007: 16, 17) also believes that prerogatives and rights share a source; for her, that source is the conception of persons as "ends-in-themselves." But unlike Hurley and Kagan, Kamm doesn't try to derive prerogatives from rights, and she doesn't take seriously the possibility of rights against oneself (see e.g. Kamm 2007: 235, 241).

philanthropist that she not send off her belongings. Clearly, such a right isn't going to be held by someone other than the agent; so, the claimant must be the agent herself. Moreover, the right has to be waivable, since if it were unwaivable (Hurley-style), it would make the agent's sacrifice wrong rather than optional. And with that, we have come back around to (Kagan-style) waivable rights against oneself, and so we have arrived at the first step of the Self-Constraint Argument. Waivable rights against oneself are, to be sure, a strange first premise, but if we are to derive prerogatives from rights, there is no other place to start.

### 3. Can we defend rights against oneself?

The Self-Constraint Argument assumes that we have the same rights against ourselves as against others: rights against oneself have the same content, confer the same kind of moral status, and are waived in the same ways. This is the conception of rights as *Self-Other Symmetric*.

The standard line on Symmetry is that it is obviously false. Counterexamples have been kicked around for decades, and rights against oneself are rarely mentioned except as the *absurdum* of a *reductio*. Many philosophers, I expect, will see the Self-Constraint Argument as doomed from the start; my view is that it is doomed only about halfway through. Let me explain, then, how we might defend the Self-Other Symmetry from the three main objections.

The classic objection is that rights against oneself are paradoxical, because they would imply an impossible power:

> *Paradox*
> If you have a right against yourself, two things follow. You are under an obligation (since the right is *against* you), and you have a power to waive away the obligation (since you *hold* the right)—and yet it should be impossible to release oneself from an obligation. (Singer 1959)

The problem, we are told, is that people can't release themselves from obligations ("in any…way whatsoever," adds Singer 1959: 202). But why not? Self-release seems to happen all the time. A student can release himself from the obligation to show up for exams by signing a form to drop a

class (Fruh 2014). A sovereign, Cohen (1996) argues, can release herself from pesky legal obligations by changing the law. More controversially, some would say that I can release myself from a duty to go to the gym by undoing the promise I made to myself on New Year's (Habib 2009, Rosati 2011: 135, Fruh 2014). It isn't clear why we have to call these cases "paradoxical." In each, the agent seems to be bound by an obligation *until* the moment of self-release. The sovereign is bound until she changes the law; the student is bound until he signs the form; I am bound until I nullify my promise. The same is true, on a Symmetric view, in the case of self-consent. My rights bind me *until* I waive them by making decisions and becoming a willing party. Just as the surgeon has a duty not to harm me until I give consent, I have the same duty to myself until—however silently and implicitly—I grant myself release.

Self-release is also the key to the second objection, which is that Symmetry is supposed to make catastrophic predictions about certain cases. For instance:

> *Minimal Pair*
> If I poke your eyeball, a right is violated. If I poke my own, no violation. Doesn't that show that people have this right only against others? (See Stocker 1976: 211, Slote 1984: 180–81, Sider 1993: 122)

But these cases don't prove a Self-Other Asymmetry; instead they reflect the familiar "asymmetry" between patients who do and don't consent. If poking your eye violates a right, that is because you aren't a consenting party to my action. When I poke my own eye (on purpose), I am acting with my own consent, and so I waive my rights. There is no violation, because I release myself from the obligation. The appeal to minimal pairs, it seems, is only effective if we hear the Symmetric view as saying that we have to treat ourselves like unwilling others, tip-toeing around our own rights. The real view, as I see it, is that our rights bind us just like the rights of a *relevantly similar other*, someone equally willing and intrinsically like us in the ways that matter. We have the same rights against ourselves as others only if other things are held equal, and the key thing to equalize is consent.

Finally, we might worry that waivable rights against oneself, objections aside, are just a

gimmick. No one ever argues for them, and so it is natural to think that they must lack a rationale:

*No Motivation*
There is no positive reason to believe that we have waivable rights against ourselves.

But besides their use in grounding prerogatives, these rights have two motivations. First, if obligations entail corresponding rights, rights against oneself are essential for the possibility of obligations to oneself, which seem "well embedded both in traditional moral philosophy and in ordinary moral thinking," as even a skeptical Singer admits (1959: 202).[79]

There is also a deeper motivation: impartiality. Many philosophers are moved by a vision of morality as impartial, a vision on which no one counts for more just because of who they are; we all have the same moral status, since each is "one among many, equally real" (Nagel 1970: Chapters 9-12). The impartial picture is powerful, but it is often rejected because of a felt tension with our ordinary belief in prerogatives (e.g. Slote 1984, Scheffler 1994). If everyone is equal, the idea goes, nothing is permissible except what is best for all; either morality is partial, or we must maximize the good. Symmetry provides a way out of this dilemma. If our prerogatives come from rights against ourselves, rather than brute permissions to self-favor, we can reconcile impartial morality with our intuitive belief in moral freedom. The promise of Self-Other Symmetric rights is to unite a mishmash of deontological hunches into something more compelling.

That, in a nutshell, is the case for waivable rights against oneself.[80] I hope it convinces you that they are worth taking seriously, even if they remain open to serious doubts. (Indubitable premises aren't the only ones worth exploring.)

---

[79] The view that obligations entail rights is hardly unique to Singer. It is standard in the tradition of Hohfeld 1919 (e.g. Thomson 1990, Gilbert 2018); Johnson 2010 argues that Kantians should accept it, too.

[80] My full argument for Self-Other Symmetric rights is the first chapters of this dissertation. In Chapter One (PDO), I try to dispel the air of paradox around rights against oneself. In Chapter Two (RAO), I take up the issues raised by unwaivable rights (e.g. not to be killed for trivial reasons), non-ideal consent, and accidental harms.

That said, let me be clear about what I need to assume about rights. They are, after all, controversial. Philosophers disagree about their ultimate source. (Are they natural or social? Based in interests or wills? Can they be grounded in outcomes, or are they situational constraints?) Tricky cases are disputed. Thankfully, we won't need any recherché conception of rights or bold case verdicts. All we need to assume, besides Symmetry, is this: rights constrain the choices of the person they are held against, except insofar as they are waived by the claimant's consent.[81]

## 4. The Self-Constraint Argument I: permitted to sacrifice

The Self-Constraint Argument starts from Self-Other Symmetric rights and concludes that self-sacrifice is optional. Now, being optional is a conjunctive matter. The act has to be permissible and omissible; we may do it, and we may refrain. So our Argument will need two parts. Let's start with the part that tries to show permissibility.

> Why is self-sacrifice allowed? Here is Kagan (1989: 209):
>
> > …if the agent is willing to make the sacrifice, then he need only grant himself permission to do so—and it will no longer violate the constraint. That is, if the agent wants to make the sacrifice, he is permitted to do so.

This should be clear enough. But let's put it in terms of our example from before.

Amanda has a right that she not crush her own arms, even if it would save Bert from being squashed. But Amanda still *may* make the sacrifice, because although she has the right—and the right in some sense binds her—it doesn't bind her inescapably. By deciding to sacrifice, Amanda waives her rights, making it permissible instead of wrong. This is symmetric to the case where Amanda saves Bert by crushing a third party Clayton's arms after his consent makes her action permissible.

---

[81] This assumption is shared by certain non-traditional consequentialists, like Portmore (2011) and Setiya (2018), who argue that the force of rights reduces to the values of (or reasons to prefer) certain outcomes. My view is that even these "consequentialized" rights can be prerogatives, since they can be finkish (see below).

Here is how this part of the Argument looks laid out:

1. Amanda has a right against herself that she not crush her arms.
2. If Amanda crushes her arms, she thereby waives that right.
3. If crushing waives the right, she is permitted to crush her arms.

*So:* Amanda is permitted to crush her arms.

This argument is valid. And the premises seem true, given our assumption of the Self-Other Symmetry: premise 1 is just Symmetry plus the setup of the case; premise 2 holds so long as deciding entails self-consent (and so long as the only way to crush the arms is via a decision—which we can stipulate is true in Amanda's case); and premise 3 is obvious, given that the sacrifice would be optimal and would infringe no one's rights. So far, so good.

## 5. The Self-Constraint Argument II: permitted not to sacrifice

Amanda may sacrifice her arms for the greater good, but that doesn't yet show that she has a *prerogative*. For all we've said, she might still be required to sacrifice, as one is sometimes required to impose sacrifices on a consenting other.[82]

So why doesn't Amanda have to make the sacrifice? Here is how Kagan (1989: 209, emphasis original) puts it:

> …suppose it violates a constraint for an agent to treat an individual in a certain way unless the agent has the permission of that individual. Then this is true regardless of who the agent is; so it is true even if the agent is the given individual herself. If forcing a certain sacrifice on a person without his permission is forbidden, then it is forbidden for that person to force the sacrifice upon himself against his will. Thus, if an agent does not desire to make that sacrifice, he need only withhold his permission qua patient from himself qua agent. Making the sacrifice in such a situation is forbidden—for it violates the constraint. A fortiori, it is not *required*. Therefore, if the agent does not want to make the sacrifice, it is not required.

[82] Imagine that Clayton says to Amanda: "Please, for the love of God, crush my arms and save Bert!" If the stakes are fairly high, she might have to oblige him. (Perhaps she still has *some* prerogative not to get involved—but it's weaker than the prerogative to keep one's limbs.)

Again, fairly clear, but we had better break it down.

Start with the three-person case. Clayton has a right over his arms. So, by default, it's wrong for Amanda to crush them for Bert's greater good, though there are exceptions for special circumstances (e.g. when Clayton consents, or when he forfeits the right by attacking someone, or when his arm is needed to save the world). Saving Bert, alas, doesn't merit an exception. His needs are outweighed by Clayton's rights, so if Clayton refuses consent, Amanda would be wrong to save Bert; *a fortiori*, she is permitted not to save him. By Symmetry, the same is true in her choice of whether to sacrifice her *own* arm for Bert's sake. By default, the sacrifice is wrong, because of Amanda's rights against herself; since she doesn't consent, the default remains in play; and so the sacrifice ends up wrong and therefore omissible. Amanda doesn't have to crush her arms.

Laid out a bit more carefully:

1. Amanda has a right against herself that she not crush her arm.
2. If she has that right, then it's wrong for her to crush the arm, unless there are special circumstances.
3. There are no such circumstances in Amanda's case (the right isn't waived, outweighed, etc.).
4. If an act is wrong, not doing that act is permissible.

*So:* Amanda is permitted not to crush her own arm.

Looks valid to me. And premises 1 and 3 are just stipulations about the case: that Amanda has a right against herself, which she hasn't elected to waive.

What about the other premises? Premise 2 is a truism about rights: it expresses the idea that violating rights is wrong unless there is a special reason why not. Special reasons come in three types: factors that take the right out of play (like waivers), huge benefits to infringing on the right, and clashes with bigger, bossier rights. And we are supposing that the right isn't waived, that the benefits aren't too huge, and that there is no huge right enjoining Amanda to crush her arm. Premise 4, meanwhile, is just good deontic logic. The only cases in which it might be false are dilemmas,

where all of one's options are forbidden (Marcus 1980), but that is not at all like Amanda's choice. In general, crushing one person's arms to save another person's life is either simply wrong or simply permissible, depending on the presence of consent.

And so we seem to have a sound argument for the permission not to sacrifice. Combine it with its other half, and we have an argument from rights against oneself to the optionality of self-sacrifice—we have shown how rights might entail prerogatives to do less than best. This completes the Self-Constraint Argument.

## 6. The Wrongness Problem

Just one problem: even granting rights against oneself, the Self-Constraint Argument has a false premise. The culprit is premise 2 from the Argument's second half:

1. Amanda has a right against herself that she not crush her arm.
2. **If she has that right, then it's wrong for her to crush the arm, unless there are special circumstances.**
3. There are no such circumstances in Amanda's case (the right isn't waived, outweighed, etc.).
4. If an act is wrong, not doing that act is permissible.

*So:* Amanda is permitted not to crush her own arm.

From premises 1 and 2, we can infer that it's wrong for Amanda to crush her arm, and from there, we infer that she doesn't have to do it. But there is no sense in which the act really is *wrong*. No matter what she chooses, Amanda isn't forbidden from sacrificing her interests for the greater good; by hypothesis, the sacrifice is optional.

This is the *Wrongness Problem:* the Self-Constraint Argument needs self-sacrifice to be wrong for it to be omissible—but it's not wrong, so we can't prove omissibility. Premise 2 has got to go.[83]

[83] Can we repair the Argument? A quick fix, suggested by Kagan's (1989: 209) phrasing, is to make the permissions and wrongness conditional on the agent's choice. Amanda is permitted to

## 7. The nature of prerogatives: deliberation and defense

How could rights against oneself make a sacrifice omissible without making it wrong? How do we solve the Wrongness Problem? We need a different approach, beyond the Self-Constraint Argument. We need a way for rights to matter without being wrong-makers.

Well, what is a wrong-maker? I think wrong-makers are opposing *reasons*.[84] Wherever there is a wrong action, there is a preponderance of reasons against it. Rights against oneself, therefore, can't be reasons, and yet they still have to allow us to self-preserve, if they are to be prerogatives. For prerogatives aren't reasons—counting against or in favor of actions, constraining deliberation—but still they help to justify actions, to ensure that they aren't wrong. That's the whole job description: prerogatives are whatever can justify doing less than best, whatever can put a gap between "most reason" and "must."

Now we have a clear, two-part strategy for replacing the Self-Constraint Argument. We need to show (1) that rights against oneself aren't reasons, but also (2) that they are able to weigh against reasons and prevent them from grounding requirements.

Why aren't rights against oneself reasons? The key fact, I think, is that there is no way to violate them intentionally.[85] For example: I have a right against myself that I not poke my own eye. But if I decide to poke it, I thereby consent to being poked, so I waive the right just in time. Same goes for Amanda and her right not to have her arms crushed; if she decides to crush them and save

---

crush *if* she chooses to, and permitted not to *if* she doesn't (because it's wrong to sacrifice, *if* she doesn't consent). But then we haven't shown that the sacrifice is optional *unconditionally*.

[84] There are other views. Maybe the wrongness of an act isn't grounded in the balance of an agent's reasons and prerogatives, but is instead grounded in facts about which *rules* are useful to adopt (Hooker 2000), or which *social contracts* are (reasonable to put) in force (Scanlon 1998, Parfit 2011). I'm not saying that this is how rule-based and contractualist views must conceive of wrong-making, but it's a possibility.

[85] Although there are *some* cases where we can violate our own rights—e.g. if they're inalienable—we can set these cases aside; for these are precisely the ones where rights against oneself really are reasons, not prerogatives. The line between alienable and inalienable rights is also the boundary between supererogation and wronging oneself.

Bert, she waives the right, and no one is wronged. There is a word for this kind of feature, the kind that evaporates in cases where it would normally have an influence: such factors are *finkish*.[86] My claim is that finkish rights, which are waived by the choice to do what they forbid, aren't reasons. I think this is a fairly intuitive claim, but it might sound unfamiliar, so let me say a bit in its favor.

First, a series of examples suggests that, in general, finkish factors aren't reasons. Imagine that there is a nifty prize behind a door. Other things equal, you should go through and collect it. Next, suppose the door is rigged with a bomb, which will grievously injure you upon approach; the bomb is a strong reason against entering the door, and so you should forsake the prize. But what if there is a way to disable the bomb, e.g. by clapping your hands? Then the bomb is no longer a reason against entering; it's only a reason against *entering without clapping;* the bomb constrains your choice of means. Finally, suppose you know for certain that the bomb can't harm you, because it's programmed to deactivate whenever you are nearby. Now you are free to enter the room however you like. Since the bomb is guaranteed not to blow up on you, it is not something that you have to take account of in deliberation. It is no longer a reason at all. From a deliberative point of view, the bomb is as good as gone. The bomb, of course, is finkish much like a finkish right, which is sure never to morally "blow up." You're guaranteed not to violate the right, so it won't ever make your actions wrong. It therefore isn't a reason at all.

Second, I have an argument based on the idea that reasons must be *stable* over different outcomes of your deliberation—they can't depend for their reason-giving powers on your actually listening to them:

---

[86] See Martin 1994, Lewis 1997. Martin imagines a machine, called a (reverse) "fink," hooked up to a live wire. Normally, live wires shock you if you touch them. But the fink detects incoming touchers and, during contact, makes the wire "dead." Intuitively, even though it's false that the finked wire will shock if touched, the wire is still disposed to shock. The disposition is just finkish: the conditions in which it normally manifests are also conditions where the manifestation is thwarted.

*Stability*
If R is a reason for A to $\varphi$, then R cannot have force only conditional on A's $\varphi$-ing.

The intuitive idea is that reasons have to be able to "talk you into" doing something. That can't happen if one is already convinced.[87] Consider an example from Nagel (1970: 53), who might be the first to endorse something like Stability:

> …to have a reason to promote an end, one must expect that there will be a reason for it whether one undertakes to promote the end or not. In this way we eliminate cases in which a single act both creates a problem for the future and then contributes to its solution. For example, if I remove the door of my office from its hinges I shall be in possession of a door to install in the now doorless entrance to my office. But the reason for possessing such an unattached door will not exist independently of my adopting this measure, so removing the door does not promote an end for which there is going to be a reason independently.[88]

One needs a door only if one removes the door—so the need isn't a reason for the removal. The same holds for finkish rights. A finkish right—e.g. against self-harm—would not be a stable reason, since the choice to harm would waive the right, sapping it of its force. The right is only in play if one in fact isn't going to harm.

Let me also mention that there is an argument for the stability principle, based in a substantive view about what reasons are. The view: reasons are premises in good practical reasoning (see e.g. Setiya 2014). A reason is something it makes sense to use in reasoning about what to do. But good reasoning can't be circular. Deliberation about what to shouldn't assume that one will in fact decide one way rather than the other. (The same is true of theoretical reasoning; it's fallacious to assume a conclusion in order to "derive" it.) Now consider this unstable would-be reason: Amanda has a right that she not crush her arms. Since the right is finkish, it only counts *against* the crushing if she *doesn't* decide to crush them; the choice to crush would waive the right. This disqualifies the right from being a reason against crushing. The right only counts given that Amanda won't sacrifice her

---

[87] A stronger principle would say that R must have force whether or not A $\varphi$s.

[88] Nagel's principle is put in terms of what an agent expects. (As is the stability principle criticized in Hare and Hedden 2016.) My principle doesn't mention belief or credence, since none of my cases turn on uncertainty.

arms, but she is in the middle of deciding whether to do so; as a good open-minded reasoner, she can't take her decision's outcome for granted. She can't treat her unstable right as a reason.

That's our first conclusion: rights against oneself aren't reasons because they are finkish.[89] But why doesn't that just erase them from the moral landscape? How could such flighty rights have any permissive power? Here, we need to remember that reasons—construed narrowly as constraints on good practical deliberation—aren't the be-all and end-all of ethics. We don't just reason our way to choiceworthy decisions; we must also *defend* our actions when the moral community comes along demanding better. Imagine that I go up to Amanda and say, "How dare you not crush your arms! That's the only way to save Bert's life, and you have most reason to save him!" She might defend herself by leaning on her rights. "But they're *my* arms," she could insist, "and I'm not willing to give them up." Here Amanda is not making excuses (as if to say, "Give me a break—I'm biased"). Nor is the point that she has special reasons to keep her arms intact (as if she were a master surgeon). The point is that she doesn't *owe* me a reason. Her rights allow her to act against the balance of reasons; they make it defensible to do less than best.

When we take the defensive stance, our aim isn't to argue that our decision was optimally guided by reasons, or that it was excusably flawed. To defend one's decision is to show that it was justifiable, that it was good enough by the standards of the moral community. Now, since defending oneself isn't the same as deliberating about what to do, we no longer have to worry about the sin of circular reasoning. In defense, we may take for granted that we aren't doing the very best thing. That's why Amanda is allowed to cite her unwaived finkish right in defense of her choice not to sacrifice, even though she couldn't cite that right in deliberation. Her right evaporates when acted against, but not when leaned on as a justification. "I have a right that I'm unwilling to waive" is a poor reason but, in this case, a fine defense.

---

[89] On this point, see also Waldron (1981: 28) on why "rights to do" are not reasons.

That is my replacement for the Self-Constraint Argument. Finkish rights aren't reasons, because they vanish when you act against them; but since you can still bring them up to defend yourself, these rights have power as prerogatives.

This derivation, besides laying the ground for a theory of prerogatives, also suggests a lesson about the nature of rights. They are not always reasons for action, meant to guide deliberation.[90] The more essential link is between rights and the standing to defend oneself, to negotiate demands. (It is no wonder that rights are said to confer authority and "status," as in Thomson 1990: 2 and Kamm 1992.) To have a waivable right against someone is, in part, to have a say over what they must do; one may claim the right—pressing the demand—or waive it. Normally we think of waivers as a sweet release. But they also make the freshly unbound agent vulnerable to other demands that had been outweighed by the right. If I release you from our lunch plans, you can no longer say "No" when other people come along asking for your time. If I insist on the plans, you don't have to say "Yes" to the others. The effect of my right is that *I have direct control over what you are obligated to do*. This fits exactly with the idea that rights against *oneself* are prerogatives. To have a prerogative is just to have some control over what one has to do. One may decide to sacrifice, waiving a right and venturing beyond the call of duty, or one may leave the right in play in order to use it in defense. Rights against oneself aren't an ad hoc concoction; they are the natural home of prerogatives, and that is why moral freedom flows so nicely from this part of the realm of rights.

## 8. Kagan's objections

Now that we have a debugged story of how rights lead to prerogatives, let's test it against Kagan's four objections to the Self-Constraint Argument. The first three, like the Wrongness Problem, will

[90] Many writers assume that rights and duties must be a kind of reason, or at least a ground of "oughts" (see e.g. Thomson 1990: 2–3, 34, 64–7, 69, 77–8, Hills 2003: 131, Reath 2006: 241, Fruh 2014: 65, Schofield 2015: 10).

call for the distinction between deliberation and defense.

**Impossibility**

Kagan objects that rights against oneself can't be prerogatives because they aren't possible to violate:

> [I]f the constraints upon which the self-constraint argument is supposed to operate are waivable—as the argument requires—then this is presumably because for such constraints the morally offensive feature that underlies violations of the constraint cannot be displayed in cases where the patient is a cooperating partner. In such cases, then, the constraints in question will not protect me from myself. They cannot, therefore, be used in any general way to yield options. The self-constraint argument fails. (Kagan 1989: 213)

The basic idea is that rights can only "protect" us from someone by making their actions wrong; a right's strength consists in the deterrent possibility of violations. My rights aren't prerogatives because they can't make my actions wrong, and so they "will not protect me from myself."

But my prerogatives don't have to protect me from *my actions*—only from morality's *demands*. Rights against oneself can override the demand to be optimal, since they carry weight in moral defense, even though they don't constrain deliberation as wrong-making reasons. Unless I am of two minds, I can't sacrifice my life unwillingly. And yet, "I'm unwilling to give it up" is enough to justify, not just excuse, the choice to preserve a life that's mine to lead.

The key to this objection, as with the Wrongness Problem, is that prerogatives don't need to make anything wrong. Our finkish rights aren't wrong-making reasons, since we can't violate them intentionally, but we can still lean on these rights to defend our choices as permissible.

**The Consent Dilemma**

Now we turn to a dilemma about deciding and consent. The Symmetric view is that we can waive our rights by making decisions and giving ourselves consent. But is it *possible* to violate a right against oneself by somehow withholding consent and acting anyway?

There are two options, Kagan says, both disastrous:

> *Can Act Without Own Consent*
> If I can act without my own consent, then I can violate my own rights by failing to waive them. But this is "bizarre;" in particular, it is bizarre to suggest that "I can violate a *waivable* constraint by imposing sacrifices on myself unwillingly—that to do so is forbidden." (Kagan 1989: 213, emphasis original)

> *Can't Act Without Own Consent*
> If I can't act without my own consent, then I can never violate my own rights. "But the reason that the self-constraint argument gave for thinking that it was permissible to *refrain* from taking on the sacrifice if I did not want to make it, was that *making* the sacrifice while I withheld permission would violate the constraint. If, however, I *cannot* violate the constraint protecting myself then the self-constraint argument will have provided no reason for thinking that taking on such a sacrifice cannot be required. And so no option will be established." (Kagan 1989: 212, emphasis original)

That is Kagan's *Consent Dilemma*. In short: if you can act without your own consent, then you can violate your own waivable rights, which seems ridiculous; if you can't act without your own consent, then you can't ever violate those rights, which means that they must be powerless, and therefore they can't ground any real prerogatives. What to do?

Kagan is right, I think, that it would be bizarre if I could intentionally harm myself sans consent. This shouldn't be possible for any decently mentally unified agent: when I act, I'm willing; without consent, I'm not.

So the real issue is what happens if I *can't* act without my own consent. Here Kagan has an argument. He says:

1. The Self-Constraint Argument works only if self-sacrifice *would* violate my rights.

Because otherwise the rights are "powerless." But since deciding waives the rights before I can ever violate them, it's also true that:

2. If deciding goes with consent, self-sacrifice *would not* violate my rights.

And this does sound right: if I crush my arm, and that entails consenting to the crushing, then I *will not end up violating* a right, because my rights will have been waived.

But we can see that Kagan is equivocating. To say that something "would violate" my rights

might, first, just be a way of saying what the content of the right is, what it proscribes. Unless a right of mine proscribes self-sacrifice, it can't ground a prerogative to self-preserve. In this *content-giving* sense of "would violate," (1) is true, but (2) is false. But there's also a *counterfactual* sense in which only (2) is true. Making the sacrifice "would not violate" my rights, in the sense that *if I were to do it, no right would end up violated*—it would have been waived just in time. In this sense, the Self-Constraint Argument does not rely on the idea that self-sacrifice would violate rights.

The crucial point here—as with the Impossibility objection and Wrongness Problem—is that a right doesn't need to be violable in order to be prerogatives. It can ground a prerogative even if it is finkish and has no hope of making actions wrong in any counterfactual scenario, so long as it counts in the context of defense.

**Fragmentation**

Kagan's third objection is that the Self-Constraint Argument requires agents to be unrealistically fragmented, as if torn between two institutional commitments.[91]

If a person occupies two social roles, Kagan suggests, it might be true that she is forbidden from doing an act that *would* be permissible if she gave herself permission (1989: 213–14). Example: for the Treasurer to buy the trinkets, she needs permission from the Secretary (same person!), who is forbidden to authorize such wasteful purchases. We can see why, for this conflicted agent, the trinkets are off limits. But it is harder to see why *normal* agents (like Amanda) should feel that the acts they could make permissible (like self-sacrifice) are in any sense off limits. Kagan's cases are meant to be exceptions that prove the rule: people aren't bound by rights against themselves.

Two points. First, Kagan is right that there is a disanalogy here. "Unlike the institutional case, there is nothing to correspond to the withholding of permission once the *person* has chosen to

[91] Kagan (1989: 214–15) makes a similar point using the case of one agent over time.

perform the act" (Kagan 1989: 213, emphasis original). I agree: deciding entails self-consent, so people can't (normally) violate their own rights. But, second, their rights could still be prerogatives. The crucial response, one last time, is that the force of a prerogative doesn't rest on the possibility of violations, but only on its potential to be used in defense. Your finkish rights can be prerogatives even if you aren't fragmented enough to violate them.

**Required to Waive**

Kagan's last objection is his simplest. We might be *required to waive* our rights against ourselves:

> …even if we grant that failure to make the sacrifice would be permissible were I to withhold permission, the self-constraint argument may *still* fail to establish the existence of options. For the moderate has given no reason to think that it is permissible for me to withhold my permission when it would be optimal for me to give it. (This is so, regardless of whether the relevant permission is at the time of acting, or at some later time.) If it is not permissible, then I am required to waive the protection which the constraint gives me from myself—and so that protection cannot be used to prove the existence of an option. (1989: 216)[92]

This objection, unlike the others, can't be solved with the idea of moral defense. The worry is that leaning on a prerogative in defense might itself be indefensible.

Kagan is right, I think, that it is at least conceivable that we might have to waive our rights. Suppose that I have promised Frieda to donate blood for her upcoming surgery, for which she might need a transfusion. I show up on the big day; the nurse asks if I am ready to give. Don't I have to do it? Other things equal, yes. Even though I retain my body rights (the nurse may not draw my blood *sans* consent), it would be wrong to insist on these rights after promising not to. My rights are there granting protection, but invoking that protection is forbidden.

This is the position we are left in, Kagan is saying, with respect to all of our rights, all of the time. Kagan has no *argument* for this unhappy conclusion, though he might observe that *we* (so far) lack an argument *for* the view that waiving is optional. Fair enough. But do we need an argument?

---

[92] Kagan's "moderate" believes in rights, prerogatives, and moral reasons to promote the good.

It's a bizarre thought that we are never meaningfully allowed to invoke our rights. Granted, Frieda's case shows that we do *sometimes* have to waive them for the greater good. But the idea that we must *always* do so, even for the puniest marginal benefit, is just a dystopian skeptical hypothesis.[93]

Still, Kagan raises a good question. If we aren't required to waive our rights for the greater good, why not? Presumably we must have a prerogative to lean on our prerogatives—a "meta-prerogative," if you like. Of course, even these won't do any good if we are required to waive them, so we will also need meta-meta-prerogatives, along with meta-meta-meta-prerogatives—etc.—which might sound a bit extravagant. But the view is really not so odd. It's just a bit recursive: we have prerogatives to lean on our rights (other things equal). That certainly sounds better than the alternative: that all rights *must* be surrendered whenever they impede the greater good, and that it is always wrong to say "It's *mine*" rather than consenting to donation.

Without meta-prerogatives, our rights are systematically undercut; we have to waive whenever it's optimal. That is one point in favor of meta-prerogatives, but not the only one. We can also give an independent argument for meta-prerogatives—conceived as rights over one's rights—based on Warren Quinn's (1989: 308–10) classic argument for rights over one's *body*.

Quinn's first premise is that your body is *yours*, not just in the physical sense that you can will its movement, but in the normative sense that you have some authority over it. Premise two is that this authority requires a restriction on how your body is to be used. If others were always permitted to redistribute your organs for the best, then your kidneys could not really belong to you—only to the collective. Your moral status, Quinn thinks, would be reduced to that of a cell in the whole, a dependent being with no bodily integrity. Real ownership requires rights, and since you really own

[93] A dialectical note: unlike Kagan's opponent, I am not trying to vindicate the intuition that we have rights and prerogatives. I am *presupposing* that we have them, and my goal is to understand their internal structure, since they are evidently linked in ways we don't yet understand.

your body, you must have decisive rights over its use.[94]

The same kind of argument can be run to establish *meta-rights*. As body rights forbid the removal of organs, meta-rights forbid the removal of rights. A meta-prerogative, in particular, is a right against oneself that one not waive one's rights via consent. We have the analogous premises:

1. Your rights are yours, normatively speaking.
2. If something is yours, normatively speaking, then you must have rights against people that they not put it out of your possession.

Conclusion: you must have rights against people that they not put your *rights* out of your possession.

And given Symmetry, you must have a meta-right against *yourself* that you not waive your rights. This meta-right is finkish, like all normal rights against oneself, so it isn't a reason against waiving, but it still counts in defense. If you choose to lean on your prerogative, keeping your spare kidney, I cannot say that you have done something impermissible. The prerogative is *your* right, so it is to some real extent up to *you* whether you are going to waive it for the greater good, just as it is up to you whether you are going to give your things and organs.[95]

---

[94] As Fiona Woollard puts it, "if others are *permitted* to take something from me *whenever* this is for the best, if there is no normative restriction that counts my needs and desires for more than theirs, then the object does not genuinely belong to me in the first place" (2015: 107, emphasis original). But note that "restrictions" on others aren't quite enough for ownership. As Kamm (2007: 82) notes, my body isn't *mine* if I lack all prerogatives over it; so, since Quinn argues only for rights against others, his argument is "incomplete." Note that Symmetry would let us complete the argument: rights against others entail Symmetric rights against oneself, i.e. prerogatives.

[95] In a footnote, Kagan (1989: 211, fn. 3) gives a fifth objection: rights can't ground the prerogative not to give aid in a case where I may either save myself or someone else. (Never mind that, in such a case, I shouldn't *need* a prerogative to permissibly save someone else.) Kagan's idea seems to be that helping the other can't have anything to do with my rights. I think that's false. If I give up my last dose of medicine, I waive a property right against myself (viz. the right that no one else possess it). If I swim out into the ocean, I waive my body rights. Kagan doesn't see any such rights in his cases because, as he describes them, they are perfectly abstract—save oneself, or save the other. But the presence of rights—and prerogatives—crucially depends on concrete details.

## 9. Conclusion

Waivable rights against oneself are prerogatives: one may cite them to defend one's conduct as permissible, but they are not reasons that inescapably constrain deliberation. Why aren't they reasons? Because they are finkish—they are waived by the same choices that would normally lead to violations—and finkish factors can't constrain us. And yet even finkish rights offer a kind of protection, since they may still be cited to defend one's choices and ward off moral blame. "The kidney is mine" doesn't count against donation, but it does block the moral demand that one surrender one's belongings for the collective weal.

This line of thought replaces Kagan's Self-Constraint Argument, which fails because of what I have called the Wrongness Problem. The Argument has to say that self-sacrifice is wrong—even though it's optional, by hypothesis—in order to show that it's omissible. The bad assumption here, which also underlies Kagan's main objections to the Argument, is that rights against oneself can only make a difference if they are wrong-making reasons. These rights should instead be seen as purely permissive: they count in defense but don't constrain deliberation. We can derive prerogatives from rights, but not if we stick to a theory with reasons alone. It is only by examining the distinctive practice of moral defense—in which we justify ourselves to the community, rather than aiming at the choiceworthy—that we will uncover the contribution of our finkish rights.

At last, we can answer our big question. Why do rights and prerogatives go together? Because prerogatives *just are* finkish rights against oneself. Nonconsequentialist ethics doesn't have to be a hybrid hodgepodge; though it may never be as simple as consequentialism, it can be elegantly unified if we take rights to be Self-Other Symmetric—if we believe in equal rights against oneself.[96]

[96] For helpful written comments, I would like to thank Alex Byrne, Brendan de Kenessey, Caspar Hare, Kieran Setiya, Brad Skow, and Peter Vallentyne. For helpful discussion, I owe thanks to David Balcarras, Nathaniel Baron-Schmitt, Renee Jorgensen Bolinger, Seth Lazar, Suzy Killmister, Agustín Rayo, Jack Spencer, Judy Thomson, Quinn White, and no doubt many others. Insightful comments from Aleksy Tarasenko-Struc led to a complete rewrite. Encouragement and criticism from Tamar Schapiro

**REFERENCES**

Bazargan-Forward, Saba (2018). "Vesting Agent-Relative Permissions in a Proxy," in *Law and Philosophy*.

Benn, Claire (2017). "Supererogatory Spandrels," in *Ethics & Politics* 19 (1): 269–290.

Bolinger, Renee Jorgensen (2017). "Revisiting the Right to Do Wrong," in *The Australasian Journal of Philosophy* 95 (1): 43–57.

Cohen, G.A. (1996). "Reason, Humanity, and the Moral Law," in Korsgaard 1996: 167–188.

Das, Ramon (2000). "Prerogatives Without Restrictions?" in *Philosophical Studies* 99 (3): 347–372.

Enoch, David (2002). "A Right to Violate One's Duty," in *Law and Philosophy* 21 (4): 355–384.

Fruh, Kyle (2014). "The Power to Promise Oneself," in *The Southern Journal of Philosophy* 52 (1): 61–85.

Habib, Allen (2009). "Promises to the Self," in *Canadian Journal of Philosophy* 39 (4): 537–557.

Hare, Caspar and Hedden, Brian (2016). "Self-Reinforcing and Self-Frustrating Decisions," in *Noûs* 50 (3): 604–628.

Harman, Elizabeth (1999). "Creation Ethics: The Moral Status of Early Fetuses and the Ethics of Abortion," in *Philosophy and Public Affairs* 28 (4): 310–324.

Herstein, Ori J. (2012). "Defending the Right to Do Wrong," in *Law and Philosophy* 31 (3): 343–365.

Hill, Thomas (1973). "Servility and Self-Respect," in *The Monist* 57 (1): 87–104. Reprinted in Hill (1991): 4–18.

-------(1991). *Autonomy and Self-Respect*. New York: Cambridge University Press.

Hills, Alison (2003). "Duties and Duties to the Self," in *American Philosophical Quarterly* 40 (2): 131–142.

Hohfeld, Wesley Newcomb (1919). *Fundamental Legal Conceptions*. New Haven: Yale University Press.

---

kept the project on track; I owe her special thanks.

Hooker, Brad (2000). *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press.

Hurka, Thomas (2014). *British Ethical Theorists from Sidgwick to Ewing*. Oxford: Oxford University Press.

Hurka, Thomas and Shubert, Esther (2012). "Permissions to Do Less Than Best: A Moving Band," *Oxford Studies in Normative Ethics, Volume II*: 1–27.

Hurley, Paul (1995). "Getting Our Options Clear: A Closer Look at Agent-Centered Options," in *Philosophical Studies* 78: 163–188.

Johnson, Robert N. (2010). "Duties to and Regarding Others," in Lara Denis (ed.), *Kant's Metaphysics of Morals: A Critical Guide*. Cambridge: Cambridge University Press.

Kagan, Shelly (1984). "Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation," in *Philosophy and Public Affairs* 13 (3): 239–254.

-------(1989). *The Limits of Morality*. Oxford: Oxford University Press.

Kamm, Frances Myrna (1985). "Supererogation and Obligation," in *The Journal of Philosophy* 82 (3): 118–138.

-------(1992). "Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status," in *Philosophy and Public Affairs* 21 (4): 354–389.

-------(1996). *Morality, Mortality Volume II: Rights, Duties, and Status.* Oxford: Oxford University Press.

-------(2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm.* Oxford: Oxford University Press.

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

Lewis, David (1997). "Finkish Dispositions," in *The Philosophical Quarterly* 47 (187): 143–158.

Mack, Erik (2005). "Prerogatives, Restrictions, and Rights," in *Social Philosophy and Policy* 22 (1): 357–393.

Marcus, Ruth Barcan (1980). "Moral Dilemmas and Consistency," in *The Journal of Philosophy* 77 (3): 121–136.

Martin, C. (1994). "Dispositions and Conditionals," in *The Philosophical Quarterly* 44 (174): 1–8.

McMahan, Jeff (1981). "Problems of Population Theory," in *Ethics* 92: 96–127.

Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

O'Neill, Onora (2001). "Duty and Obligation," in Lawrence C. Becker and Charlotte B. Becker (eds.), *Encyclopedia of Ethics, Second Edition*. Routledge Press: 423–428.

Owens, David (2012). *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Oxford University Press.

-------(2011). *On What Matters, Volume One*. Oxford: Oxford University Press.

Portmore, Douglas (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.

Quinn, Warren (1989). "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," in *The Philosophical Review* 98 (3): 287–312.

Reath, Andrews (2006). "Self-Legislation and Duties to Self," in *Agency and Autonomy in Kant's Moral Thoery: Selected Essays*. Oxford: Oxford University Press. 231–249.

Rosati, Connie (2011). "The Importance of Self-Promises," in Hanoch Sheinman (ed.), *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press. 124–155.

Ross, William David (1930). *The Right and the Good*. Oxford: Oxford University Press.

-------(1939). *Foundations of Ethics*. Oxford: Oxford University Press.

Scanlon, Thomas (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.

Scheffler, Samuel (1992). "Prerogatives Without Restrictions," in *Philosophical Perspectives* 6: 377–397. Reprinted in Scheffler 1994: 167–192.

-------(1994). *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying*

*Rival Moral Conceptions*. Revised Edition. Oxford: Oxford University Press.

Schmidtz, David (1990). "Review of *The Rejection of Consequentialism*," in *Noûs* 24: 622–627.

Schofield, Paul (2015). "On the Existence of Duties to the Self," in *Philosophy and Phenomenological Research* 90 (3): 505–528.

Setiya, Kieran (2014). "What is a Reason to Act?" in *Philosophical Studies* 167 (2): 221–235.

-------(2018). "Must Consequentialists Kill?" in *The Journal of Philosophy* 115 (2): 92–105.

Sider, Theodore (1993). "Asymmetry and Self-Sacrifice," in *Philosophical Studies* 70 (2): 117–132.

Singer, Marcus G. (1959). "On Duties to Oneself," in *Ethics* 69 (3): 202–205.

Slote, Michael (1984). "Morality and Self-Other Asymmetry," in *The Journal of Philosophy* 81 (4): 179–192.

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

Waldron, Jeremy (1981). "A Right to Do Wrong," in *Ethics* 92 (1): 21–39.

Woollard, Fiona (2015). *Doing and Allowing Harm*. Oxford: Oxford University Press.

# CHAPTER FOUR

## Why Isn't Supererogation Wrong?

## Why Isn't Supererogation Wrong?

Supererogatory acts, like donating a kidney to save a life, are optional and yet better than other permissible options. How is that combination possible? The standard line is that supererogation arises from a Self-Other Asymmetric view of benefits: we may treat our own welfare as uniquely (un)important. After presenting counterexamples, I give a new account: supererogation arises from a Self-Other Symmetric view of rights and duties: the limits of what we owe to others arise from what we owe to ourselves. The promise of this account is (1) to accommodate a fuller range of cases, (2) to reduce supererogatory permissions to something we already understand, viz. moral rights; and (3) to explain a curious duality: the supererogator makes a sacrifice that, if imposed on others, would be wrong, as a violation of their rights.

### 1. Introduction

It's a classic drama (Taurek 1977, Parfit 1978). A man owns the last dose of a vital medicine and is choosing whether to save someone he knows and likes—David, who needs the whole dose to survive—or a group of five strangers, who need just one-fifth each. Other things equal, you might think, the man is obligated to do the most good, to help the five instead of poor David.[97]

But there's a twist: the man *is* David. Since the agent himself pays the cost, helping five seems to go beyond the call of duty. In ethics lingo, we would call it *supererogatory:* optional and yet superior to a permissible alternative (McNamara 1996: 429). Now a puzzle arises. The supererogatory act imposes a sacrifice on the agent that would be obligatory to impose on anyone else. Disinterested onlookers must save the five, and may not save David instead—and so we wonder: why should he be allowed to save himself? Why isn't "supererogation" just obligatory?

The standard answer is that David is morally allowed to care more about his own interests than the greater good. On what I call the *Cost View*, supererogation is optional, rather than required,

[97] Taurek (1977) denies that it's better to save five. I disagree, but to accommodate Taurek, we could just pick a new case, e.g. choosing between a benefit to David vs. a bigger benefit to another. (Note that Taurek himself has no account of supererogation, though his case is a classic example.)

because agents have special permissions when it comes to costs and boosts to *their* wellbeing—in this case, David has an "agent-favoring" permission to benefit himself, even at others' expense.

The Cost View is dominant—but known to be itself a bit costly (§2). One problem is that, by inflating the weight of self-interest, the view threatens the optionality of supererogation from the obverse direction. If being selfish matters more, why doesn't that make "supererogatory" sacrifices *wrong?* The Cost View doesn't have a natural answer to this second puzzle, I argue. But my main objection is that there are counterexamples, where permissions and costs come apart (§3).

If costs can't explain supererogation, what can? I propose a new answer: moral *rights*. By basing supererogation in rights, we can give more elegant explanations for a wider range of cases (§§4–6). We will also face the second puzzle with renewed force (§7). The supererogator imposes a sacrifice on herself that she may not force on unwilling others: removing a kidney, throwing a body on a grenade, sending cash off to Oxfam. So why isn't her act *wrong?* I think the Rights View can answer this using nothing more than truisms from Rights 101, along with one extraordinary assumption: that our rights aren't just restrictions on others, but also relevant to how we treat ourselves. With the unfamiliar machinery of self-regarding rights, we can dispel the paradoxes around our everyday concept of the supererogatory.

## 2. The Cost View

Let's not get ahead of ourselves. The standard view is that supererogation has no essential link to rights—it is instead supposed to emerge from the specialness of self-interest. The Cost View is gospel. So how does the view work, and what does it have going for it?

In its pure original form, the Cost View is meant to explain why we may favor ourselves over others. It is not obligatory for me to give my kidney, even if you need it more. Nor does David have to give up his medicine, even though it could be split up to save five others. But why? I don't

objectively matter more than you. David doesn't objectively matter more than the five. So why does morality allow agents to *act* as if they were the center of the universe? Scheffler (1994) says: in a sense, we each *are* the center. From my "personal point of view," projects and pains matter more when they are *mine*, and morality acknowledges this by letting me give my interests extra weight. The same goes for David and his point of view. Any agent's interests can have extra weight for them.

Credit where it's due. Scheffler's Cost View is pure and unified, in part because it is just meant to treat one specific kind of case. Scheffler posits one permission:

> *Agent-Favoring Permissions*
> Agents may sometimes do what's in their own interests even if it is worse for others.

And he gives it a principled rationale: the "personal point of view," which is supposed to be a deep fact of human nature: we are naturally partial creatures, who give our own interests special weight.

But what kind of weight? As Kagan (1994: 338–39) argues, it can't be that we have *moral reasons* to self-benefit. Moral reasons tend to require whatever the count in favor of. But we aren't *required* to favor ourselves; it isn't morally wrong to sacrifice for others. Clearly the Cost View needs to respect this fact. But how? Why exactly, on the Cost View, isn't supererogation wrong?

Rather than canvassing all possible answers, let me focus on what I think is the simplest, most flexible, and most promising way to develop the Cost View here.[98] If there are counterexamples to this version, there are counterexamples to all.

Thomas Hurka and Esther Shubert argue that the cost to the agent isn't a moral *reason* to avoid supererogating; it is instead a *prerogative:* it tends to make actions permissible without tending

---

[98] Here are some of the main alternatives. Slote (1991: 917) and Portmore (2003, 2011) suggest we have only non-moral reasons against supererogating (classic objection: these reasons threaten to make supererogation *irrational*; see Kagan 1991: 927–28, Hurka and Shubert 2012: fn. 7); Bader (forthcoming) invokes dominance principles (but as he admits, they entail that imprudence is morally wrong); Dancy (1993: Chapters 8 and 12) argues that we may "discount" the reasons against supererogating; and Raz (1975) thinks that we have only second-order "exclusionary reasons" facing off against our reasons to supererogate (see Postow 2005 and Whiting 2017 for criticism).

to require them.[99] So the permissibility of favoring oneself doesn't come from egoistic reasons but from an agent-favoring prerogative, which allows us to forsake the pursuit of others' greater good even if that is what we have most reason to be doing. I have more reason to give you my kidney, but I may keep it for myself, since I have a prerogative that lets me self-preserve.

So the Cost View should include:

> *The Prerogatives Principle*
> An option *x* is wrong *iff* for some other option *y*, the reasons to do *y* outweigh the *combined* reasons and prerogatives to instead do *x*.

This says that we have to follow the reasons unless we have a (sufficiently weighty) prerogative not to. Prerogatives put a gap between "most reason" and "must"—and for Scheffler, there is just one gap, allowing agents to self-favor. That is the essence of the original version of the Cost View: we have an agent-favoring prerogative grounded in the personal point of view.

This view has some potential drawbacks. One issue is that the view posits a massive asymmetry between self and other, whereas we might have hoped that morality would be more impartial and impersonal. But Scheffler has a principled response: the personal point of view. If prerogatives emerge from our selfish default perspective, then it is no surprise that morality would end up Self-Other Asymmetric; that just reflects the asymmetry in who matters most from our points of view. A more pressing problem, however, is that we haven't seen any reason why the personal point of view should ground *prerogatives*, rather than moral reasons to be partial to oneself. "The standard problem here" as Dancy puts it, "is that if we succeed in justifying the agent's failure to do the heroic act, we can no longer approve his choosing to be a hero" (1993: 140; see also Nagel 1986: 204, Kagan 1994: 348, Hurley 1995: 168). To say "we have prerogatives instead of moral

[99] Hurka and Shubert say "prima facie duty" and "prima facie permission." I use "(moral) reason" and "prerogative" because they are more common. (Since non-moral reasons won't play any role in later discussion, I will start using "reason" to mean "moral reason.") Also, note that some writers say "option" instead of "prerogative" (Kagan 1989, Dancy 1993). I think "option" is a bit confusing, because it also refers to the things we are choosing between, but that's just me.

reasons" delivers the right results, but it does leave us wondering *why*.

I think these challenges are evocative but not decisive. The Cost View isn't really in danger until we apply it to a wider range of cases: examples where we have prerogatives beyond agent-favoring, which seem inscrutable from the personal point of view.

Let's get dangerous.

## 3. Costly counterexamples

Scheffler's Cost View can decently explain supererogation when it's costly to the agent. But there are other cases we need to account for, where the impartially best act is optional but not costly at all.

There are three key cases here. The first, and most famous, involves the opposite of agent-favoring: "agent-sacrificing," which is bad for the agent, impartially suboptimal, and yet sometimes morally optional instead of wrong (Slote 1984; see also Stocker 1976, Hurka and Shubert 2012).

Suppose David has a pain pill and is choosing whether to give it to Elise or take it himself. Elise has a nasty headache. David's is even worse. An impartial spectator, pining for less global woe, would prefer that David take the pill himself, curing the bigger ache. But surely David isn't morally required to take the pill; he may also give it to Elise, sacrificing his greater weal for her lesser good. He is permitted to do something impartially suboptimal even though it is also worse for him.

Scheffler's Cost View, with its agent-favoring prerogatives, can't allow for this. But some versions of the Cost View can. Hurka and Shubert (2012) propose that agent-sacrifice is permitted by a second kind of prerogative:

> *Agent-Sacrificing Prerogative*
> Agents have a prerogative to do what goes against their own interests.[100]

The basic idea behind these dual prerogatives is that people ought to get a moral say when their

[100] More precisely: an agent has a prerogative to do *x* rather than *y* if *x* is worse for the agent than *y*. What matters is relative costliness.

interests are at stake. David may give his own interests extra weight, as in the choice between his life or five others'. He may also give his interests *less* weight, as in the choice between his greater relief and Elise's lesser relief. Agents don't always have to do what's impartially best, if doing good has an impact on their own welfare. There are prerogatives to avoid costs *or* to incur them.

Again, credit where due. This souped-up Cost View can allow for optional agent-sacrifice. It can even handle our second hard case, which I call "self-benefitting supererogation."[101] Suppose that David is considering whether to cure his pains with his pill before going camping with Elise (who is pain-free). Clearly it's better for David if he takes the pill. Moreover, it's a nice gift for Elise, since she will have a less grumpy and more enjoyable companion. Taking the pill is therefore better for everyone—and yet it's supererogatory, not required.[102] The agent-sacrificing prerogative can allow for this. Not taking the pill is bad for David, and so he has a prerogative not to take the pill.

But wait. How could *that* explain David's permission? We don't normally think that, if an act hurts you, that helps justify it. Mugging is no less wrong when the mugger breaks a hand. Leaving someone to die isn't more permissible if you need their help next week. There is something deeply fishy, I think, about a prerogative to harm and neglect oneself. It's one thing to say that we aren't required to self-favor. It's another to say that we have a permission to impose costs on ourselves at everyone else's expense. Being costly shouldn't justify anything.

So while agent-sacrificing prerogatives can account for some intuitions, they don't give good

---

[101] Michael Ferry's example is a rewarding favor. "Consider the example of buying a book for a friend. Suppose it's on sale, and you think your friend will enjoy it. It may be morally better to buy it than not, but it is the nature of such favors that they are not obligatory. Such acts do not involve significant sacrifice and may even be in the interest of the acting agent if the joys of giving outweigh the costs" (2013: 579). See also Archer 2016: §4 (the "Free Help Guy" case), 2017: 134.

[102] There is a method for constructing this kind of self-benefitting supererogation: start with a clearly optional benefit to oneself (like pain relief), then stipulate that it has a nice effect for others. Often the act remains optional. Another method is to start with a supererogatory sacrifice (like giving a kidney) and then stipulate that the hero will be compensated so that the act is just barely rewarding overall. Again, the extra reasons aren't sure to tip the scales.

explanations. Nor do they fit with Scheffler's rationale. The whole idea of the "personal point of view" is that we may treat our own interests as *more* pressing, not less; this perspective is the opposite of what we would need to justify self-sacrifice.[103] It is hard to reconcile agent-sacrificing and -favoring prerogatives, and even harder to imagine a single basis for them both.[104]

But suppose we found the perfect basis and came to love agent-sacrificing prerogatives. Then our Cost View would be able to explain the optionality of agent-sacrifice and self-benefitting supererogation. But we would still face a very serious counterexample: *cost-neutral supererogation*.[105] Suppose David is in the hospital, and he has a chance to sign up for a medical study that will contribute to curing an annoying disease. Signing up isn't costly; the procedures are ones he would have had to undergo anyway, and he doesn't have anything better to do. The study really does benefit others. Still, that does not entitle anyone to David's signing up; he is still well within his rights to refuse. Participating in the study would be supererogatory. We might also consider a more face-to-face kind of case. Suppose you learn that a stranger on the bus would enjoy watching you strike a silly pose, which wouldn't affect your interests (no one will catch it on camera). I doubt that you are required to amuse them. A shot at harmless fun doesn't entitle anyone to a say over what you do with your body—just as it doesn't entitle them to the use of your property or disclosure of personal secrets. More generally, there doesn't seem to be anything fishy about cost-neutral, supererogatory favors. If so, we have counterexamples to the Cost View in any form.

How could friends of the Cost View resist these three cases? The best move, I think, would

---

[103] This point is due to Michael Slote (1984: 190), who makes the same complaint against Williams's (1973) appeal to "integrity," which can justify agent-favoring but not agent-sacrificing.

[104] To explain why agent-sacrifice is optional, rather than wrong, some writers say that we have only *non-moral* reason to benefit ourselves (e.g. Portmore 2011). (This would also explain why it's not wrong to be a bit imprudent—for example, by not brushing one's teeth for a night.) This view doesn't need agent-sacrificing prerogatives. But it can't explain self-benefitting supererogation, and it seems to be in tension with the idea of duties to oneself.

[105] For related cases, see Horgan and Timmons 2010, Ferry 2013: 579, and Archer 2016: §4.

be to point out that I have been using a narrow conception of what counts as a "cost to one's interests." As I have used the term, "costs" are hits to welfare, drops in wellbeing. But what if we include other things as costs, like expended efforts and frustrated desires? Even "cost-neutral" favors require me to move my body, after all, and that takes at least a modicum of agential *oomph*. And when David declines to participate in the study, presumably that is because he is inclined not to do it, and so he is avoiding some exertion.

This might help, but not by much. For one thing, it doesn't seem true in general that, when something is effortful or unpleasant, we have a prerogative not to do it. "It takes effort" is at best an *excuse* for giving in to violent or vicious urges (see Yetter Chappell forthcoming: 13, fn.4). Second, the wicked, who struggle most to do good, shouldn't perversely enjoy more freedom (Archer 2016: §4). But here's the key point: a broad notion of cost can't fix all the problems. Even if we can show that "cost-neutral" supererogation really has hidden costs, we will still be stuck with rotten explanations for agent-sacrifice and self-benefitting supererogation. That is problematic enough.

So here is where the Cost View ends up, once we have pelted it with a few hard cases. We have prerogatives, but they are a hodgepodge with no coherent unified source, and they give strange explanations for agent-sacrifice and self-benefitting supererogation ("it costs me," perversely, counts as justification). Moreover, these prerogatives can't allow for any truly cost-neutral supererogation. When we stick to agent-favoring, the Cost View might seem fine, but the moment we go beyond, the view has to stretch and contort to capture our intuitions—and even then, some slip free. Costs and benefits to us aren't the source of our prerogatives.

## 4. Toward a rights-based account: two intriguing cases

Where do prerogatives come from, if not costs? Our counterexamples don't point to any answers, unfortunately. But there are two other cases that might. I won't rely on them as objections, but they

might help light up some hidden factors in the supererogator's situation.

Case one: the permissibility of "other-favoring." Recall the choice we started with. Someone is deciding whether to use his drug to save five strangers or poor David. This time we subtract the plot twist: the agent is you, not David himself. We will stipulate that the case is properly boring, with no special reasons why you might owe David the pill. You haven't pledged it to him; you two aren't best friends, lovers, cousins, bodyguard and client—you just know each other. According to Taurek and Anscombe, you may nonetheless permissibly save David rather than the five. I agree (even though, unlike Taurek, I think you have more reason to save five instead).[106] It is wrong to assume, as I did at the start of this paper (following Parfit 1978), that we may not be partial to acquaintances. The lesson here is that your prerogatives aren't just geared to protecting what's in *your* interests. They aren't even confined to the interests of loved ones in your egoistic penumbra. Instead your prerogatives let you help whomever you choose—even relative strangers. What matters, from the point of view of permissibility, isn't so much *who stands to benefit*. What matters is *who owns that drug*.

For suppose—and this is our last hard case—that David owns the drug and tries to save himself, but before he can, the five desperately attempt to rob him of it. What they do is wrong, I think, precisely because the drug is his and not theirs. Taurek agrees:

> Moreover, and this I would like to stress, in not giving his drug to these five people [David] does not wrong any of them. He violates no one's rights. None of these five has a legitimate claim on David's drug in this situation, and so the five together have no such claim. Were they to attack David and to take his drug, they would be murderers. Both you and David would be wholly within your rights to defend against any such attempt to deprive him of his drug. (1977: 300–01)

This is the case of "rights vs. costs," where cost to self tells in favor of violating rights. The striking thing about the faceoff is that costs lose. The five may *not* rob David, even to save themselves. Their

---

[106] Taurek (1977: 300) suggests that it's not just false but meaningless to say that five lives are "more valuable, period" than one. Anscombe (1967), by contrast, seems to take the commonsense view that we do have reason to save the greater number. For other writers who endorse other-favoring permissions, see Setiya 2015, Worsnip 2018, and Bader forthcoming.

"agent-favoring prerogatives" seem to vanish when they confront David's rights—a counterexample to every version of the Cost View we have seen so far. Hurka and Shubert (2012: 9) are forthright about this challenge; they say that any Cost View must grant that agent-favoring prerogatives have zero weight against rights, even though there is no explaining this "complication." But surely this cries out for explanation. There has to be some link between prerogatives and rights.

That, I think, is the ray of light thrown by these last two cases. The Cost View struggles because it tries to squeeze all of supererogation from just a sliver of moral reality—viz. payoffs, facts about who benefits how much. But we also have to ask: *who owns the goods?* The crucial factor in our cases is ownership, or more generally, moral *rights*. When you have rights over a pill, you may use it as you like, even in some ways that aren't best for you or best impartially. (You may give it to David.) You also enjoy a kind of protection from covetous others, who are forbidden from taking your stuff and stuffing even if they need it just as much as you—or even more.[107] Supererogatory permissions, whatever they are, belong inside the realm of rights.

## 5. The Rights View

How do we build up an account of supererogation from rights? It's all about where we start. The Cost View comes from thinking about cases like David and the drug. We wonder: why may he save himself, given that usually one has to save the bigger number? The Rights View comes about from a different paradigm.

It's another classic (Foot 1967, Thomson 1985). Five patients are about to die in a remote clinic, each needing different organs, when along comes a healthy loner. The doctor in charge is sure: this person is the only match within miles for a transplant. So she tells her team to skip the

[107] If you think it is permissible to take someone else's pill, so long as the thief needs it more, you might really be intuiting that the "owner" loses rights in conditions of scarcity. That is consistent with the idea that *if* we have rights over something, *then* sacrificing it is supererogatory.

consent forms, bring out the anesthesia, and use the loner's viscera to save the five. Sounds like a textbook rights violation. The effects are excellent (less death, no witnesses), and still the killing feels monstrously wrong.

Again, a twist: the loner *is* the doctor. She comes in for her shift, and then freely gives her life to save her beloved patients. Not much of a monster, now, is she? Instead of violating rights, she goes beyond the call, supererogating for the greater good.

This is a strange, remarkable duality. One and the same action, depending on who the agent is, is either a sordid violation or a soaring act of heroism. And it's not just this case. Nearly any classic example of supererogation, I'll soon argue, has its own shadowy counterpart. Sending cash to Oxfam, giving a pint of blood, throwing someone on a grenade, committing to adopting a child—these are all lovely deeds when it's *your* belongings, body, and household. But that kind of sacrifice is plain wrong to force on others.

That gives us the essential core of the Rights View, a principle that I call:

> *Prerogatives Reflect Rights*
> You have a prerogative not to $\varphi$ yourself *just if* your rights forbid others from $\varphi$-ing you.

This is rough, but I hope you can see how it is supposed to fit our cases. David has a right that the five not take his drug; he also has the prerogative not to hand it over. The five have no prerogative to seize the drug; David violates no rights by keeping it. *Our prerogatives let us omit precisely those acts that our rights won't let others do for us*—moving our bodies, "donating" a kidney, changing possession of a pill. Prerogatives pattern along with rights.

That's the gist. Now some qualifications. First, what kinds of things do we have prerogatives to do? I have stated the principle with '$\varphi$' ranging over transitive verbs (like 'kick', unlike 'run') that take you, the prerogative-holder, as direct object. But we often use other constructions to express rights and prerogatives. I have a prerogative to retain possession of my kidney. If you promise not to run for the rest of the day, I have a right that you not run. This is all fine and good. I worded the

principle as I did to make it readable, not to insist on just one way to spell out rights. That said, when we invoke "Prerogatives Reflect Rights," either to develop it or give objections, it is important that we be able to describe the relevant acts so that they can in principle be done to the agent by anyone, self or other. Instead of "kicking oneself," we say "kicking David" (or whoever); we can "throwing David's body on a grenade," and "putting the five's pill in David's mouth." Otherwise the principle doesn't make sense. If there are systems of rights that cannot possibly be expressed in this way, that is a problem for the Rights View.[108]

Second, what is a right? We don't need a fancy reduction, just a sense of what we are talking about. I am talking about *claim rights*, what Hohfeld calls rights "in the strictest sense" (1919; see also Thomson 1990: Chapter 1). If I have a (claim) right against you that you not harm me, two crucial things follow, other things equal. (1) Harming me is *wrong* (since it would wrong me), (2) but it is not wrong if you act with my *consent*. Rights are discretionary restrictions, constraints that apply except insofar as the right-holder grants release by consenting. That is a pretty minimal definition. But that's just what we should want. The Rights View shouldn't have to take a stand on whether rights are institutionally defined, natural or social, grounded in long-run interests or exercises of the will. Whatever *else* they are, rights are at *least* consent-sensitive constraints, and that is the only aspect we need to develop the Rights View.

Finally, let me note that there will be principled exceptions, where one has prerogatives without rights or vice versa. This kind of thing is nothing exotic in the realm of rights, since on any view, rights can be shuffled by agreements, legislation, and intimate relationships. You have different claims against friends and foes, fellow citizens and foreigners, promisors and perfect strangers. Here

---

[108] What about privacy rights and rights of autonomy? Aren't they essentially interpersonal? As I say in fn. 54 of Chapter Two (RAO), above, the Symmetric view would have to conceive rights of autonomy and privacy as rights against humdrum acts like perusing files and confining someone in a box, which one can do to oneself, or to the contents of one's own hard-drive.

is the key. By waiving or gaining rights against *all* others at once, you might create a mismatch with your prerogatives. For example: you have no prerogative to vote for a manifestly awful candidate in a legitimate democracy—it's wrong—but you do have the right that others not stop you from so voting. You get these rights from a political arrangement.[109] For a case of prerogatives without rights, suppose that I'm the owner of some nifty loot, and I announce that whoever finds its hiding spot may take it. I still have the prerogative to keep the loot, but I have waived the right that others let me. These sorts of cases aren't any threat to "Prerogatives Reflect Rights." That principle is just supposed to hold other things equal. It says that *unless* some rights are waived or gained, prerogatives will be the intrapersonal flipside of interpersonal rights. That is a weaker claim, but still strong enough to help us with the hard cases, as I'll now argue.

## 6. Rights and the cases

The main objection to the Cost View is that it faces certain counterexamples. Can the Rights View handle them?

Let's start from the top.

| | |
|---|---|
| *Agent-favoring* | David is considering whether to use his lifesaving drug to save the five or save himself. Either choice is optional. |

The Rights View gives us a guide here. We look for reflections: does David have a right against the five that they not take his pill? I think *yes*—it's his pill. And since prerogatives reflect rights, David also has a prerogative to keep the pill for himself.

Now we get to the three hard cases. The very hardest:

| | |
|---|---|
| *Cost-neutral* | David is considering whether to sign up for a medical study. It wouldn't hurt him to enroll, but doing so is still morally optional. |

Does David have a prerogative not to enroll? Again, we can look for reflections in rights. Would you

[109] This is "the right to do wrong" (Waldron 1981; see also "From Rights to Prerogatives.")

wrong David by signing him up against his will, or without his knowledge? Again, I think *yes*. And not just because of arbitrary legal convention. David has rights over his body. He is (with limits) "sovereign" when it comes to what happens in his skin, which means that he has an "authority to choose and make decisions" about surgical procedures and medical tests when veins pricked and pulses felt are *his* (Feinberg 1989: 53).[110] Because he has rights against bodily invasion, we can conclude that he has a corresponding prerogative not to elect to have his body fiddled with by the study's authors.

We will want a similar treatment for:

*Self-benefiting* — David owns a pill. Taking it will cure his headache, and it will also be nice for Elise, who is about to be stuck with him on a camping trip. Still, taking the pill is morally optional.

David isn't obligated to self-medicate for other people's modest benefit, since he has rights over his body. We would wrong David by slipping the pill into his coffee or shoving it down his throat, however gently. And so we would be overstepping our bounds if we said that David is morally required to take it himself; we would be ignoring his bodily prerogatives. The prerogative not to self-benefit reflects our rights against paternalistic meddling.

And much the same goes for:

*Agent-sacrifice* — David has the bigger headache, but it is still morally optional for him to give his pain pill to Elise, who has a smaller headache.

Again, we can appeal to David's body rights: he doesn't have to take the pill for his own good, and he would be wronged if we forced him to take it.[111]

---

[110] There are exceptions. Consider compulsory vaccinations, which are in the overwhelming public interest, as well as being healthy for the person vaccinated. Or consider patients who are incompetent to make decisions, misinformed, or requesting lines of fentanyl.

[111] If we want agent-sacrifice to be *supererogatory*, not just optional, we need to say more. Perhaps Elise's right to aid is an extra reason to help him, on top of the utility of doing so. Some evidence for this view is that agent-sacrifice does *not* seem supererogatory when Elise tries to waive all rights to aid, insisting that David should keep his pill, and he transfers the pill to her anyway.

Things get more interesting in the bonus cases. Recall:

| | |
|---|---|
| *Other-favoring* | You are considering whether to use your lifesaving drug to save David (whom you know and like) or to save the five. Either choice is optional, even though you have no special duties to David. |

To explain why you may favor David over the five, it isn't enough to appeal to your "personal point of view." David is not your intimate. Nor is it enough to point to your body rights, since the choice isn't about what goes into your body. Now it starts to matter again, as with agent-favoring, that we are talking about *the agent's pill*. It's yours, and so you have a say over what will happen to it. The ground of your prerogative isn't that you're helping *David*—he isn't anyone special to you, beyond someone you know and like. The key is that the pill, and so the choice, is yours. You have a prerogative to help David. If the five steal your pill—or David's, if you've already given it—they violate a right.

This brings us to our final case:

| | |
|---|---|
| *Costs vs. Rights* | David has a lifesaving drug, which he intends to use on himself. The five may not steal his drug even to save their own lives. |

This case is trouble for the Cost View, because we have to make a brute exception to the five's agent-favoring prerogatives, so that they don't outweigh David's rights. The Rights View gets things right immediately. The five have no prerogative to take David's drug because *they have no rights over his drug*. They aren't wronged if David takes it, and they have no prerogative to interrupt his taking.

And that is how the Rights View treats the cases. What matters isn't who stands to benefit, but who has rights over what.

### 7. Why is supererogation optional?

If prerogatives reflect rights, that is enough to capture our judgments about some hard cases. But if prerogatives do reflect rights, surely that can't be a coincidence. It is not as if our prerogatives just *happen* to be the flipside of interpersonal rights, shadowing them by sheer fluke. There must be a

link—but what? We need to know more, I think, about the ultimate ground of prerogatives; something that explains *why* they do their thing, justifying courses of action without tending to make them obligatory. How could prerogatives emerge from a system of rights?[112]

The place to start is with the shadow puzzle of supererogation. When David gives his pill to the five, he imposes a sacrifice on himself that would be wrong to impose on an unwilling other. So why isn't his "supererogatory" action wrong?

The tempting gut response is: David has *no rights against himself*. There is a David-sized gap in his rights' moral coverage, a singular exception in the list of people who have the usual duties to him. This seems intuitive enough, and it respects the widespread worry that rights against oneself are paradoxical (see e.g. Singer 1959). But we have to be careful. If we just carve a self-regarding gap in David's rights, what will be left to reflect his rights against others? We need *some* kind of right to be there. But it must—somehow—play a permissive rather than a restrictive role.

How could a right be merely permissive? It is a truism that rights, or at least claim rights, are typically restrictions. David's rights make it *wrong* for the five to take his pill, other things equal. So you might think that, to get prerogatives from rights, we would need another kind of right entirely. Not so! We just need our second truism, which is that rights can be *waived by consent*. If David consents, that is a major moral plot twist: the five now wrong no one if they take his pill. But of course *David consents to his own actions*. When he swallows his pill, he isn't an unwilling benefactor. He freely chooses, in light of the facts, who will get to take his pill; he is a consenting party to his own intentional plans. Supererogation isn't *wrong* for the same reason that is isn't wrong for the five to save themselves with David's permission. Consent waives rights—and we are our own most reliable consenting patients.

---

[112] In this section I am laying out the derivation of prerogatives I defend in Chapter 3 (FRP), where I also raise objections to competing derivations.

We have arrived at a picture of rights that is *Self-Other Symmetric*. We have rights against ourselves that apply to the same actions, and obey the same principles, as our rights against others. Rights against oneself are ordinary claim rights, but they aren't restrictive like the rights of unwilling others. They are instead like the rights of a consenting other—which makes sense, given that we consent to our own choices. This is why supererogatory self-sacrifice isn't wrong.

Now we turn to the original puzzle. Why isn't supererogation *required?* If David's rights are like those of a consenting other, they don't restrict him, so they don't count against self-sacrifice. But there are mighty moral reasons in favor of self-sacrifice here—viz. the needs of the five. Why don't they generate a moral obligation? How does David's right block them?

Here the key fact is that, so long as David *doesn't* waive his right, it can serve a valuable purpose. He can use it in the course of morally justifying his actions, even though his right isn't a moral reason. Suppose we confront David. "How *dare* you! Five people are dying, and you're saving yourself? That's wrong!" He could reply: "But it's *my* pill." He isn't saying that he needs it more than the five, or that his interests are supreme from where he's standing. David isn't giving a reason why he should keep the pill. If anything, he is saying that he doesn't *owe* us a reason. He is leaning on his rights, citing a prerogative, explaining why he may defensibly save himself even if it's better overall to do something else.

That is how to develop the Rights View. We have rights against ourselves, symmetrically reflecting our rights against others, but because we consent to our choices, our own rights matter to us as prerogatives instead of restrictions. David has a right that he not give up the pill, just as he has a right that others not take it. This doesn't make giving the pill to the five wrong, because David doesn't have to worry about violating his right, since he can't give his own pill away unwillingly. But David's right is still potent for the purposes of moral defense. If the moral community demands that he hand it over, he can remind them that the pill is *his*, thus leaning on his right that it remain in his

possession. That is how rights against oneself could make supererogation optional.

## 8. Conclusion

Supererogation is haunted by two puzzles. If the greater good is really so great, why aren't we just *required* to promote it—to give our kidneys, drugs, and best efforts? Presumably the answer is that something weighs against the greater good, like self-interest or duties to oneself. But—puzzle two—if these are so weighty, why don't they make "supererogation" flatly *wrong?*

The Cost View, I have argued, struggles with both paradoxes. The view can explain why supererogation isn't obligatory only in a slice of cases where self-interest clashes with the greater good; there we have an "agent-favoring" prerogative reflecting our tendency towards egoism. When we turn to other examples—sacrifices for others' lesser good, rewarding gifts, cost-neutral favors, partiality towards acquaintances, selfish violations of rights—the Cost View can capture at best a slice of our intuitions, and then only by snipping its aortal link between prerogatives and partiality to self.

As for the other conundrum—why supererogation isn't wrong—the Cost View has a shallow answer. The view says that self-interest, though potentially very weighty, has only the permissive weight of a prerogative rather than the obligating force of moral reason. There is no explaining *why* it has one kind of weight rather than the other, no derivation of prerogatives from independent insights about how interests matter.

That is where the Rights View shines. Instead of saying that there is a brute asymmetry between self and other, where self-interest has extra weight of an extraordinary kind, the Rights View says that rights are Self-Other Symmetric. We have the same basic rights against ourselves as against others: they have the same content and work by the same principles. Because our rights *count against* the imposition of sacrifices—they tend to forbid the transfer of our pills and kidneys—they

give us someplace to stand when moral busybodies demand that we promote the greater good. We can say: "I don't have to give this, even if you need it more, because it's *mine*." This is symmetric to the case where someone demands that we impose sacrifices on unwilling others for the sake of the impartial good; we can refuse on the grounds that others' kidneys, drugs, etc. belong to *them*. But surely we don't have to treat our own kidneys like the organs of unwilling others. If we choose to donate them, that isn't wrong—it's supererogatory. Why? Because of another principle of rights—that they are waived by consent—combined with the psychological fact that we are consenting parties to our own decisions. Generously giving my kidney isn't wrong; it's symmetric with the case where I optimally redistribute the kidney of a consenting other.

In the end, supererogatory prerogatives don't arise from a kaleidoscopic clash of interests, refracted and magnified from divergent "points of view." The real source is a simple picture of rights. We all have the same rights against everyone by default; they are wrong to disobey unless they are waived; and when we are *both* the holder of a right *and* the one bound, that coincidence results in our having a kind of control over our obligations—a moral prerogative. What we owe to others is limited not by our egoistic nature, but by what we owe to ourselves.[113]

[113] I am delighted to thank my audiences at MIT's Work in Progress Seminar, the Australian National University, Monash University, and the Rocky Mountain Ethics Congress at CU Boulder, where I received fantastic comments from Aleksy Tarasenko-Struc. For yet more discussion, my thanks go to Cheryl Abbate, David Balcarras, Nathaniel Baron-Schmitt, Dan Bonevac, Thomas Byrne, Jonathan Dancy, Brendan de Kenessey, Sam Dishaw, Sinan Dogramaci, Julia Driver, Elliot Goodine, Tom Hurka, Doug Portmore, Agustín Rayo, Tamar Schapiro, Kieran Setiya, Judy Thomson, and Quinn White. For comments on drafts, I am grateful to Caspar Hare, Ryan Preston-Roedder, Tamar Schapiro, Kieran Setiya, and the participants in MIT's Dissertation Seminar.

**REFERENCES**


Anscombe, G.E.M. (1967). "Who is Wronged? Philippa Foot on Double-Effect: One Point," in *Oxford Review* 5: 16–17.

Archer, Alfred (2015). "Saints, Heroes, and Moral Necessity," in *Royal Institute of Philosophy Supplementary Volume* 77: 105–124.

-------(2016). "Supererogation, Sacrifice, and the Limits of Duty." *The Southern Journal of Philosophy*, 54 (3): 333–354.

-------(2017). "Supererogation." *Philosophy Compass* 13.

Bader, Ralf (forthcoming). "Agent-Relative Prerogatives and Suboptimal Beneficence," in Mark Timmons (ed.), *Oxford Studies in Normative Ethics*.

Chang, Ruth (2002). "The Possibility of Parity," *Ethics*, 112 (4): 659–688.

Dancy, Jonathan (1993). *Moral Reasons*. Oxford: Oxford University Press.

Feinberg, Joel (1989). *The Moral Limits of Criminal Law Volume 3: Harm to Self*. Oxford: Oxford University Press.

Ferry, Michael (2013). "Does Morality Demand Our Very Best? Moral Prescriptions and the Line of Duty." *Philosophical Studies*, 165 (2): 573–589.

Foot, Philippa (1967). "Abortion and the Doctrine of Double-Effect," in *Oxford Review* 5: 5–15.

Heyd, David (1982). *Supererogation*. Cambridge: Cambridge University Press.

-------(2016). "Supererogation," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), URL = <https://plato.stanford.edu/archives/spr2016/entries/supererogation/>.

Hohfeld, Wesley Newcomb (1919). *Fundamental Legal Conceptions*. New Haven: Yale University Press.

Horgan, Terry and Timmons, Mark (2010). "Untying a Knot from the Inside Out: Reflections on the 'Paradox' of Supererogation," in *Social Philosophy and Policy* 27: 29–63.

Hurka, Thomas and Shubert, Esther (2012). "Permissions to Do Less Than Best: A Moving Band," in *Oxford Studies in Normative Ethics, Volume II*: 1–27.

Hurley, Paul (1995). "Getting Our Options Clear: A Closer Look at Agent-Centered Options," in *Philosophical Studies* 78: 163–188.

Kagan, Shelly (1989). *The Limits of Morality*. Oxford: Oxford University Press.

-------(1991). "Replies to My Critics," in *Philosophy and Phenomenological Research* 4: 919–928.

-------(1994). "Defending Options," in *Ethics* 104 (2): 333–351.

Kamm, Frances (1996). *Morality, Mortality, Volume II: Rights, Duties, and Status*. New York: Oxford University Press.

McNamara, Paul (1996). "Making Room for Going Beyond the Call," in *Mind* 105 (419): 415–450.

Nagel, Thomas (1986). *The View from Nowhere*. New York: Oxford University Press.

Parfit, Derek (1978). "Innumerate Ethics," in *Philosophy and Public Affairs* 7 (4): 285–301.

-------(2011). *On What Matters, Volume One*. Oxford: Oxford University Press.

Portmore, Douglas (2003). "Position-Relative Consequentialism, Agent-Centered Options, and Supererogation," *Ethics* 113: 303–332.

-------(2008). "Are Moral Reasons Morally Overriding?" *Ethical Theory and Moral Practice*, 11: 369–388.

-------(2011). *Commonsense Consequentialism*. Oxford: Oxford University Press.

Postow, Betsy (2005). "Supererogation Again," in *Journal of Value Inquiry* 39: 245–253.

Raz, Joseph (1975). "Permissions and Supererogation," in *American Philosophical Quarterly* 12 (2): 161–168.

Scheffler, Samuel (1982). *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Oxford University Press.

Setiya, Kieran (2015). "Selfish Reasons," in *Ergo* 2 (18).

Singer, Marcus (1959). "On Duties to Oneself," in *Ethics* 69 (3): 202–205.

Slote, Michael (1984). "Morality and Self-Other Asymmetry," in *The Journal of Philosophy* 81 (4): 179–192.

-------(1991). "Shelly Kagan's *The Limits of Morality*," in *Philosophy and Phenomenological Research* 51: 915–917.

Taurek, John. M. (1977). "Should the Numbers Count?" in *Philosophy and Public Affairs* 6 (4): 293–316.

Thomson, Judith Jarvis (1985). "The Trolley Problem," in *The Yale Law Journal* 94 (6): 1395–1415.

-------(1990). *The Realm of Rights*. Cambridge: Harvard University Press.

Waldron, Jeremy (1981). "A Right to Do Wrong," in *Ethics* 92 (1): 21–39.

Whiting, Daniel (2017). "Against Second-Order Reasons," in *Noûs* 51 (2): 398–420.

Williams, Bernard (1973). "A Critique of Utilitarianism," in J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against*. New York: Cambridge University Press.

Worsnip, Alex (2018). "Eliminating Prudential Reasons," in Mark Timmons (ed.), *Oxford Studies in Normative Ethics Volume 8*: 236–257.

Yetter Chappell, Richard (2017). "Willpower Satisficing," to appear in *Noûs*.

# CHAPTER FIVE

## Better to Do Wrong

## Better to Do Wrong

May we give money to ineffective charities, if doing so is better than permissibly keeping it for ourselves? May firms operate sweatshops, if that is better than permissibly going out of business? Some answer "Yes," arguing that wrong choices cannot be better than permissible ones. This "Worse to Do Wrong" principle makes sense if we think of wrong acts as those below a certain cutoff on the scale of bad-to-good. But the principle is false, I argue, since there is no one ultimate scale of moral evaluation, and so there is nowhere to draw a cutoff. On any view that allows for supererogation, wrongness is due to ("incommensurable") factors on two dimensions; in my view, these are reasons (which determine an option's goodness) and prerogatives (which let us choose suboptimally). I defend the intuitive verdicts about suboptimal beneficence and lay out the conditions when Worse to Do Wrong fails. When prerogatives protect only the worse of two suboptimal options—e.g. they protect doing nothing, but not running sweatshops—it can be better to do wrong.

### 1. All or Nothing Problems

A building is about to collapse onto two children, who are strangers to me. At the cost of two crushed arms, I can pry open a small escape route to save the first one—or open a bigger route, saving both. Either way the cost to me is the same. It would seem:

> **Wrong to Save One**
> It is morally wrong for me to save just one child.

After all, I can save the second with no extra cost or effort: letting this child die would be senseless. Such gratuitously bad acts are intuitively wrong. But there is a fascinating argument against this intuition, an argument that saving only one is at least as permissible as saving nobody. Here is how it goes.

The first premise is common sense:

> **May Save None**
> It is morally permissible for me to save no one.

After all, they're my arms: I don't have to give them up for the greater good. That said, while my arms might matter quite a lot to me, they aren't objectively more important than a child's whole life. That leads to premise two:

**Better to Save One**
I have more reason, all things considered, to save the one child than to save no one.

This strikes me as plausible. Letting both kids die is a worse option than saving only one: saving the child is more choiceworthy, more favored by reasons, more strongly recommended. "Surely," Horton (2017: 94) says, "the best moral view would not discourage you from saving the one child" rather than nobody. I think he is surely right. Morality's advice can't be sanctimonious to the point of self-defeat. When rights aren't in the way, better to save more lives. (Really, the argument only needs the weaker claim that saving one isn't *worse* than saving zero. In §3, I will consider a view that swaps in this claim for Better to Save One.)

The kicker is an apparent truism about wrongness and reasons:

**Worse to Do Wrong**
If A is wrong and B is permitted, then I have more reason to do B.

And this sounds irresistible. Shouldn't we always avoid wrongdoing? Isn't it the *last* thing one should ever do? But then we could prove that saving one isn't wrong. If it is wrong, and saving zero isn't, then saving one must be worse—but it *isn't* worse to save the child than to keep my arms.

This is Joe Horton's (2017) marvelous *All or Nothing Problem*: three fine intuitions, one tantalizing principle—and they're inconsistent.[114] Something has to give. If we are tempted to reject Wrong to Save One, we will read Horton's Problem as an argument that ineffective altruism is permissible. If loafing around is good enough, the idea goes, so is being a mediocre hero.

Does this argument work? I argue *no*. It really is wrong to save one, in Horton's case, which is a straightforward counterexample to Worse to Do Wrong. The All or Nothing Problem arises from an over-simple link between reasons and wrongness. In my view, a wrong act can be better

[114] The All or Nothing Problem has lots of precursors (Parfit 1982, 2011b: 225; Kagan 1989: 16; Tadros 2011: 161–62; Portmore 2011: 147; Snedegar 2015: 379; Pummer 2016: 83). Sinclair (2018) and Pummer (forthcoming) respond. (In setting up the Problem, Horton uses 'ought rather', not 'more reason'. He also introduces it as an argument that saving one must be worse than saving zero.)

than a permissible one because we might have a prerogative to do our worst option (save zero) without any prerogative to do the next-worst (save one).

Before I develop this solution, however, let me emphasize that Horton's Problem is quite general, with implications across moral philosophy. One major application, which Horton (2017: §4) explores, is to the ethics of philanthropy (Woodruff 2018). Suppose that I can give my paycheck to either of two charities. Oxfam will use the money effectively, curing 200 children of a dreadful disease. Shmoxfam will bungle the job, curing only 100—not great, but still better than spending the money on myself. Since I have no special reason or hankering to give to Shmoxfam, it seems:

> **May Give Nothing**
> It is morally permissible for me to spend the money on myself.
>
> **Better to Give to Shmoxfam**
> I have more reason, all things considered, to give to Shmoxfam than to spend on myself.

And so, given Worse to Do Wrong, it cannot really be wrong to give to Shmoxfam. We cannot say:

> **Wrong to Give to Shmoxfam**
> It is morally wrong for me to give to Shmoxfam.

This means we can't condemn as wrong an act that involves letting 100 kids suffer for no reason.

We find the same structure in a defense of sweatshops (a simplified version of Zwolinski 2007: 699–700; see also Rulli and Worsnip 2016: 213–14). Suppose that M&H's employees consent to, and benefit from, their miserable wages, but there is enough slack in the labor market that M&H could easily afford to pay much more. Even so, one might argue that the low wage is defensible:

> **May Employ No One**
> It is morally permissible for M&H to do no business at all.
>
> **Better to Exploit**
> M&H has more reason, all things considered, to pay sweatshop wages than to do nothing.

And so we cannot say:

> **Wrong to Exploit**
> It is morally wrong for M&H to pay sweatshop wages.

But it does seem wrong to run sweatshops, assuming (maybe unrealistically) that employers could easily pay more.

Finally, All or Nothing Problems arise when we are allowed to infringe rights, as in just war and self-defense. One nice case involves innocent threats (on which, see Thomson 1990: Chapter 14). Suppose that Threat has gone temporarily mad and is going to injure me. If necessary, I would be allowed to slap him more than a few times to defend myself; indeed, I would have more reason to slap him twice than to let him to hurt me. As it turns out, I'm sure that even one slap would do the trick. It seems:

> **May Slap No One**
> It is morally permissible for me to do nothing.
>
> **Better to Slap Twice**
> I have more reason, all things considered, to slap Threat twice than to do nothing.

And so Worse to Do Wrong will not let us say:

> **Wrong to Slap Twice**
> It is morally wrong for me to slap Threat twice.

But it *does* seem wrong to do unnecessary harm—even when under threat.[115]

The stakes, then, are high. If our All-or-Nothing argument can justify saving the one in Horton's case, the same argument might justify exploitation, ineffective altruism, and gratuitous harm. These further arguments, of course, might be resisted on the ground that they involve unrealistic assumptions. There is no such thing as Shmoxfam; labor markets can be awfully tight; threats in war can't be slapped away. Realistic cases involve uncertain agents and tradeoffs among beneficiaries. These details matter. But if my objection works, it will work regardless of details, since the argument will in each case invoke a false principle: Worse to Do Wrong. To debunk this principle, we won't need to leave Horton's building.

---

[115] This is the classic "necessity" constraint on self-defense (Lazar 2012).

After laying out the case for Worse to Do Wrong (§2), I will show that it is false given a simple theory of supererogation using reasons and prerogatives (§3). Throughout I will assume that some acts (like saving two) are indeed *supererogatory:* they are better than some other permissible option, but not required. Some theories reject this assumption, but it is a presupposition of the debate we are entering. The All or Nothing Problem never arises if doing "Nothing" is forbidden.[116] That said, the other intuitions about Horton's case—that saving one is wrong, and that it is better than doing nothing—are not safe to presuppose. I will defend them (§§4–5) once I am through with objections to Worse to Do Wrong.[117] I conclude (§6) with some reflections on what wrongness could be, if not the mark of worseness, and how we might respond to better-yet-wrong actions.

## 2. Worse to Do Wrong and the Permissibility Bar

The All or Nothing Problem turns on Worse to Do Wrong, which says: if option A is wrong and option B isn't, then there is more reason to do B than A. Any permissible act is better than any wrong act. By "better," I don't mean that the results are nicer, or that the agent deserves more praise. "Better" and "worse," for us, are just a snappy way to describe the balance of reasons.[118]

So, why believe Worse to Do Wrong? Horton says it seems "intuitively correct" (Horton 2017: 96). But so does the idea that saving one is better than permissibly saving zero—and we can't believe both of these at once. Hence the All or Nothing *Problem*.

---

[116] For example, maximizing act utilitarians believe that we must always do what best promotes happiness, and Rossian (1930) deontologists say we must do what we have most moral reason to do.

[117] In a groundbreaking paper, Theron Pummer (2016) has argued for the same intuitions regarding a similar case, though he does not try to debunk Worse to Do Wrong. One of my aims here is to give Pummer's intuitions a deeper rationale. In reading his paper, I was helped by a 2017 discussion on the PEA Soup blog, especially the critical precis by Johann Frick; URL = <http://peasoup.us/2017/04/philosophy-public-affairs-discussion-pea-soup-theron-pummers-whether-give-critical-precis-johann-frick/>.

[118] In light of this, we can't cite as counterexamples familiar cases where wrong acts make things go best, such as Thomson's (1985) *Footbridge:* push a man to his death, and you save five from a trolley—optimal results, wrong act. Here the *act* isn't optimal; you have more reason not to kill.

Is there anything deeper we can say in favor of Worse to Do Wrong? Horton notes that

> …there are countless cases that seem to verify it. Suppose, for example, that it is permissible to say something nice, permissible to say nothing, and wrong to say something nasty. [Worse to Do Wrong] implies that you ought to say nothing rather than say something nasty. And that seems the right result. (Horton 2017: 96)

But lots of false views are right *most* of the time. Most people are under 7' tall, but any basketball fan knows there are exceptions to the claim that everybody is. For all we have seen, it remains open to us, as fans of the intuitions in the All or Nothing Problem, to say the same about Worse to Do Wrong: it holds most of the time, but the case of the two kids in the building is an exception. Saving one is wrong, given that I could just as easily save two, and yet one is better than none.

Another thought is that wrongness *itself* is a "conclusive reason" to abstain, as Darwall (2013c) argues. If so, Worse to Do Wrong follows for free; for any wrong act, we have a conclusive reason to do any permissible act instead. But Darwall's own argument here presupposes Worse to Do Wrong. He assumes that there is a conclusive reason to avoid wrongdoing, and then infers to the best explanation: wrongness must *be* the reason. A nice inference, but not if we are trying to establish Worse to Do Wrong rather than explain it.

So what really underlies the principle's appeal? I think the answer lies in a diagram:

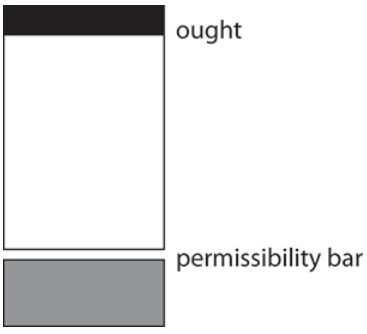


*Fig. 1: Snedegar's Permissibility Bar*

Here, Justin Snedegar (2016: 161) illustrates a simple and seductive idea about the nature of permission: to be permissible is to rank *high enough* on the scale from worst to best. At the scale's peak sit the options that we should choose, or ought to choose. Somewhere below is a cutoff—the

"permissibility bar"—separating the wrong from the permissible. This is already enough to entail Worse to Do Wrong. Wrong acts are below the bar; permissible acts are above it; and "above" just means "better." That is the crux of the All or Nothing Problem. If doing nothing is over the bar, so is anything better—exploiting laborers, saving one life, slapping twice, supporting Shmoxfam. Even if these acts are gratuitously worse than our best option, they are over the bar, and therefore kosher.

But what could possibly be wrong with the idea of a permissibility bar? Would anybody seriously deny that "permissible" implies "good enough," that to be wrong is to be below a cutoff on the scale from bad to good?

I think anyone who believes in supererogation should reject the very idea of a permissibility bar—and with it, Worse to Do Wrong. The problem is that, given supererogation, *right and wrong can't be reduced to positions along any single scale*. The grounds of permission are *incommensurable*, in the sense that an act's moral status can't be fully described unless we say how it ranks along multiple, independent scales. There is no "permissibility bar" because there isn't anywhere to put it.

Now our task is clear. We need to know: why does supererogation require multiple moral dimensions? What are those dimensions? How and when do they make wrong acts better than permissible ones?

### 3. Reasons and Prerogatives

Supererogatory acts—like enduring harm to save lives—are optional and yet better than other permissible options (McNamara 1996). This conjunction is sometimes said to lead to a "Paradox" (Raz 1975, Heyd 1982, 2016, Archer 2018: §4). If "supererogatory" deeds are really so *good*, the idea goes, why aren't they just *obligatory?* Solving the puzzle means finding a coherent picture of the supererogatory, an account of the flexible link between reasons and requirements.

Now isn't the time to argue for any one solution. Instead, let me just note that the best

familiar options *share* a commitment to incommensurability. The lesson of the Paradox is that one dimension isn't enough. Why is that? Well, we know that the deontic status of an act—whether it's forbidden, optional, or required—depends on the balance of reasons, because better acts tend to be required. That's why we are never required to do our worst option, and always permitted to do our very best (ignoring dilemmas). But there is only one salient, simple way to define right and wrong from the dimension of better and worse: say we are required to do our *best*. And this flatly rules out the supererogatory.

What about more complex one-dimensional definitions? Could they make room for supererogation? Yes—but not quite where we want it. We could say:

**The "Lots More" Principle**
An act A is wrong *iff*, for some option B, there is lots more reason to do B than to do A.

This allows us to say that doing my friend a little favor is supererogatory: the favor is better than nothing, though not by too much. But the principle can't allow for slight wrongs (like pinching someone's nose) or splendid-yet-optional deeds (like saving two lives at great personal cost). Another definition:

**The Baseline Principle**
An act A is wrong *iff* A is below the "baseline" of goodness.

Where the "baseline" is either an absolute value or something relative to the menu, e.g. an average (see Hurka 1990 on "comparative" vs. "absolute" satisficing). But this principle has the following further problem. Acts that are *above the line* but *gratuitously bad* seem to be wrong. For example, suppose I am choosing between staying outside the building, saving a kid at the cost of my arms, or saving the kid (at the same cost) while yelling rude insults (Bradley 2006, Yetter Chappell forthcoming). Both acts of heroism, we can suppose, count as "above the baseline." Even so there is no justifying the choice that involves insults; it is gratuitously bad, and so wrong, and yet the "Baseline" Principle permits it. (As does the "Lots More" Principle, since it's just a little naughty; see

Snedegar 2016: 163.)[119] The moral of all this is that the complicated definitions don't work.

Thankfully, supererogation is a breeze when we add another dimension. What do we add? The simplest view says that, in addition to reasons, we have *prerogatives*, which justify actions without counting in favor of them, permitting us to do our worse options.[120] (After we develop this view to debunk Worse to Do Wrong, we'll consider whether we could do the same with other 2D views.)

How do prerogatives work? Think back to the building case. Intuitively, I don't have to save the kids, even though that would be best, because it would cost me my arms. We can explain this with a prerogative to avoid self-injury; a permission, other things equal, not to crush my arms. Crucially, the prerogative isn't just a special *reason*. The force of my prerogative is supposed to be that I may act *against* my reasons, which on balance favor crushing the arms and saving the kids. With prerogatives, I can justify my actions even if we all know I could have done better. Sometimes I don't *owe* people a reason for doing less than best.

Now we can easily link up wrongness to reasons:

> **The Prerogatives Principle**
> An act A is wrong *iff*, for some option B, the reasons to do B outweigh the combination of reasons and prerogatives to do A.

The Prerogatives Principle is essentially *two-dimensional*. The dimension of reasons tells us how good or bad an option is (Dancy 2004, Parfit 2011a, Chang 2014); the dimension of prerogatives tells us how much worse an option can be while remaining permissible.[121]

---

[119] Another 1D trick is to assign each option an *interval* along the real line, representing its value, such that an option $x$ is permissible just in case there is no option $y$ such that $y$'s interval is wholly to the right of $x$'s. See Chapter 6 (SRC) for a counterexample involving an extension of Horton's case.

[120] Hurka and Shubert (2012) argue for prerogatives, which they call "prima facie permissions." For an important precursor, see Scheffler 1982 on agent-centered prerogatives.

[121] I assume that prerogatives, like reasons, have weights. In our cases, we can represent the weight of a prerogative to do an option with a single number (same for reasons), but there are cases where this isn't possible. For example, I might have a prerogative to do $x$ rather than $y$, but no prerogative to do $x$ rather than $z$. (In Chapter 6 (SRC), I argue that we find such a *contrastive* prerogative in Kamm's (1996: Chapter 12) Intransitivity Paradox.)

Instantly, we have an easier time with the cases. The simple two-option case of supererogation is one where the reasons favor one thing, but we have a prerogative to do worse—e.g. I have a prerogative not to give up my two arms to save a life. We can explain slight wrongs and splendid supererogation: in the first case, we have a tiny reason with no prerogative opposed; in the other, we have a mighty reason with a mighty prerogative opposed.

And at last, we can treat our main case of gratuitous badness: The All or Nothing Problem. With reasons and prerogatives, we have a clean account of how wrongly saving the one child could be better than permissibly saving zero. Here it goes.

Intuitively, I have some reason for staying put outside the building—my arms. I also have a stronger reason to save the one, and double that reason to save both. But I do not have to save anyone, thanks to a prerogative to save my skin. (For simplicity, I will assume that this is the only relevant prerogative in play.)

So here is the breakdown.

| | Save 0 | Save 1 | Save 2 |
|---|---|---|---|
| Reasons | 1 | 2 | 4 |
| Prerogatives | 5 | 0 | 0 |

There are two crucial facts here. First: even though I have more reason to save more lives, not even my reasons to save both kids can outweigh the combination of my reasons and prerogatives to save neither: $4 < 6$. This is what makes the saving supererogatory—more reason, not required. Second: notice that there is no comparable prerogative to save one. That is why, given the extra reasons to save two, there can be no justification for saving one instead—saving one has got to be wrong.[122]

[122] It's not essential that I lack *any* prerogative to save one—just that I lack a sufficiently strong

So the case is an exception to Worse to Do Wrong, and we can finally say what *makes* the case exceptional. The only reason why saving none is permissible, and saving one isn't, is that *saving none enjoys more protection from prerogatives*. My worst option is protected while my second-worst isn't. This gives us a necessary and sufficient condition for when doing wrong isn't worse:

> **When Wrong is Better**
> A wrong option A is better than a permissible option B *iff* (1) A is better than B; (2) both are worse than some option C; and (3) B is protected by a prerogative, while A isn't.

In other words: we find counterexamples whenever we have two suboptimal options, and only the worse one is made permissible by prerogatives, allowing it to deontically outperform one of its betters.

Since this is a necessary condition for when wrong acts are better, we can also explain why doing wrong *isn't* better in Horton's unexceptional case from before:

| | Say Something Nasty | Say Nothing | Say Something Nice |
|---|---|---|---|
| Reasons | -1 | 0 | 1 |
| Prerogatives | 0 | 2 | 0 |

Here, we have two suboptimal options—Nasty and Nothing—and only the *better* one is protected by prerogatives, rather than the worse. That is why this case, unlike the building case, obeys Worse to Do Wrong.

That's my solution to the All or Nothing Problem. It's better to do wrong—to save just one—than to permissibly do nothing, and this is possible because of prerogatives, which allow me to suboptimally self-preserve. We can keep the intuitions and give up Worse to Do Wrong.

---

one. Also, note that we can give essentially the same treatment to the case of slightly gratuitous badness (the insult case) by reducing the gap in reasons between best and second-best.

All that is left is to argue for two of the key intuitions—that saving one is wrong, and that it's better than saving zero. But first, let me emphasize that the basic idea behind my two-dimensional account of wrongness doesn't require that we use reasons and prerogatives, as I've been understanding them.

For illustration, let's see how Worse to Do Wrong fails on a view with two dimensions of reasons. Derek Parfit's (2011a: 137–41) view is that supererogation involves a conflict between "partial" reasons (like "it would hurt *my* arms") and "impartial" reasons (like "it would save two lives," which has no essential indexical reference to *me*). Parfit thinks supererogation happens when our impartial reasons favor sacrifice but are offset by our partial reasons to self-preserve.

Now, consider what Parfit might say about Horton's case. Saving two is better overall than saving one; it's impartially better and in no way worse. But even though both savings are impartially better than doing nothing, they are partially worse, and so both are on a par with doing nothing.

We find a similar breakdown as we did before:

| | Save 0 | Save 1 | Save 2 |
|---|---|---|---|
| Impartial Reasons | 1 | 2 | 4 |
| Partial Reasons | 5 | 0 | 0 |

The key fact: partial reasons make my impartially *worst* option fine, but not my impartially *second-worst*. Here it is not worse, all things considered, to wrongly save one than permissibly save none.

Why not go in for Parfit's view? The key problem is that Parfit must say that *no* supererogatory act can be overall better than the moral minimum; saving two can't be better than saving zero. I think it supererogation clearly can be better (as do Hurka and Shubert 2012: 8, Snedegar 2016: 162, and Harman 2016: 383). In response, Parfit might say that saving two is

*impartially* better, but that just raises the question of why impartial value should matter more than partial value (Hurka and Shubert 2012: 9).[123]

To capture supererogation's higher value, we need prerogatives, but my case against Worse to Do Wrong does not essentially require them—only incommensurability.[124]

## 4. Against Ineffective Altruism

Supererogation is possible because our best options aren't the only ones we can justify. Besides reasons, which make options better, we also have prerogatives, which let us choose for the worse. This two-dimensional picture helps us see how a wrong act could be better than a permissible one—our prerogatives might protect only the worse of two suboptimal choices.

What does this mean for the All of Nothing Problem? We began with a three-premise argument about Horton's building case, in which I can save one or two lives at the same cost:

**May Save None**
It is morally permissible for me to save no one.

---

[123] A second issue is evident from the table above. If there are strong partial reasons to self-preserve, why isn't supererogating irrational? Won't it be disfavored by the overall balance of reasons? (Except in the exceedingly rare case where our reasons are exactly equipollent?) Here Parfit appeals to the idea that when two types of reasons conflict, two options might be "on a par" rather than precisely equal in strength—each better in a respect, neither better overall. (See Chang 2002, Hurka 2012; and note that Parfit prefers talk of "rough equality" to "parity.") This allows for an interesting parallel to my treatment of Horton's case, since the paradigm cases of parity can be used to construct counterexamples to Worse to Do Wrong. Start with two options on a par, like a tea and a coffee; then added a mildly "sweetened" version of one option—the same coffee at a discount—that is still on a par with the other option. The original coffee is clearly the wrong choice, since it's beaten by the discount coffee, but it doesn't follow that the original coffee is *worse* than all *permissible* options. Tea is on a par with discount coffee and so is still permissible, and by stipulation, tea is also on a par with the unsweetened coffee option. This is exactly analogous to how Parfit's view would handle Horton's case. The option to save two is the "sweetened" version of saving one, but doing nothing is permissible since it's on a par with both savings. That means that saving one is wrong (because it's worse than saving two) but not overall worse than a permissible option (since it is on a par with doing nothing).

[124] We have many other 2D options besides mine and Parfit's. We could use two further kinds of reasons (e.g. Gert's (2007) justifying and requiring reasons, or moral vs. non-moral). Or we might ground prerogatives in something deeper, like rights (Benn 2017) or self-interest (Scheffler 1982).

**Better to Save One**
I have more reason, all things considered, to save the one child than to save no one.

**Worse to Do Wrong**
If A is wrong and B is permitted, then I have more reason to do B.

And the conclusion was that saving one must be permissible. Whereas we wanted to say:

**Wrong to Save One**
It is morally wrong for me to save just one child.

By developing a picture of reasons and prerogatives, I have been attacking a premise: Worse to Do Wrong. I have also been assuming a premise: May Save None. (We brought in prerogatives to *explain* this kind of permission.)

But I have not directly argued for the other two claims: Better to Save One and Wrong to Save One. At best, I've given an indirect defense, by showing how they could coherently be true. Let me now argue for them directly; afterwards I will consider how to extend the argument to more complex cases.

Sticking to the building case for now: why think that saving one is *better* than saving zero? Well, in a two-option choice—{*Do Nothing*, *Save 1*}—is it clearly better to save one.[125] I claim that adding the option to *Save 2* shouldn't change that fact, unless there is a special reason why *Save 2* would make *Save 1* worse or *Do Nothing* better. By default, we should assume that the facts about betterness are not menu-relative.[126] Reasons don't just change for no reason; exceptions call for explanations.

---

[125] Unless we deny supererogation or accept something like Parfit's view of it (see above). My argument in this section could be recast as an argument that saving one *isn't worse*.

[126] To be sure, we are forced to accept menu-relativity in certain cases where wrongness is itself menu-sensitive—viz. cases that violate Property α ("The Independence of Irrelevant Alternatives"), which says that subtracting an option can't make another option wrong, or Property γ, which says that an option permitted in all sets in a collection is also permitted in their union. But Horton's case obeys both α and γ; it violates only Property β, which says that if two options are permissible, adding another can't make one wrong without also making the other wrong. Pure β-violations—"β blockers!"—are associated with intransitivity and incommensurability, not menu-relativity. See Chapter 6 (SRC) on this association, and for the Greek properties, see Sen 1993, 2017: Chapter 1*.

Could there be a good explanation? Let's consider the two main options. Lazar and Barry (ms.) argue that *Save 1* becomes worse, in *Save 2*'s presence, because we have reasons not to *disrespect* the second child; leaving someone to die, when saving is no extra effort, is said to be an offense to their humanity. The key challenge for this view, as I see it, has to do with advice. Suppose I tell you that I won't be saving two, and I ask whether I should save zero lives or one. Surely the right advice is that I should at least save the one. (Though you might note, for the record, that I should really be saving both!) As it turns out, Lazar and Barry have a way to make room for this intuition. Their view is that reasons of disrespect are agent-relative (they are specially bound up with the relation between the agent and those disrespected), and so they don't guide the advice of third parties. But this implies that third-party advice should *always* ignore reasons of respect (i.e. deontological restrictions), which seems to me a bit extreme. A good advisor wouldn't recommend murdering people for their organs, even when the effects are wholesome agent-neutrally.

Another way to explain menu-relativity would be to say that *Save 1* becomes worse (than *Do Nothing*) when it becomes *wrong*, and that adding *Save 2* makes it wrong. But I can't think of any rationale for this view except Worse to Do Wrong, which we have already rejected. There doesn't seem to be any persuasive reason to posit menu-relative goodness in Horton's case. If it's better to save one than zero in a pairwise choice, it remains better even if we add in *Save 2*.

Next, why think that saving one is genuinely *wrong*? I think everyone will admit that there's something awful about refusing to save the second child if I'm going to be losing my arms anyway. But you might think that what's on display here is just bad character, that we are dealing with vice on parade, not bona fide wrongdoing. Horton (2017) gives two arguments here: (1) I can't *justify* the act of saving only one, given that I could easily have saved more, and (2) saving only one in the three-option case is *like* doing nothing in a pairwise choice between saving zero and costlessly saving one (which, in effect, is the choice we face after ruling out the option to do nothing). Both of these

points are powerful, and we could try refining them.[127] Instead, let me offer a simple complementary argument, much like Horton's (2), but again drawing on the idea of menu-relativity.

Premise 1: it's wrong to save one in a pairwise choice between saving one and saving two (where both options involve me losing my arms); if these are the only options, clearly I have to save the second child—it's extra goodness at no extra cost. But then how could saving one become permissible by adding in the inferior option to save no one? It couldn't, according to Premise 2: merely adding the option to save zero cannot make saving one permissible. Adding options, in general, doesn't remove wrongness. The only exceptions involve menu-relative goodness, where a new option might create new reasons to pick something previously wrong. To be clear, I am open to the idea that we might find such relativity in outré cases.[128] But not this case. There is no reason to think that adding the chance to do nothing—a worse option than saving one!—should make saving one *permissible* instead of wrong. Conclusion: saving one is wrong in a choice between saving zero, one, or two.

Now we have arguments for Better to Save One and for Wrong to Save One. Can we extend them to less artificial cases? Yes; here's how. Notice the arguments start from judgments about pairs: *Save 1* is the *better* option in a choice with *Do Nothing*, and *Save 1* is *wrong* in a choice with *Save 2*. Step two is to argue that these judgments shouldn't change when we add a third option. Whenever we have a three-option case with the same pairwise judgements—e.g. Shmoxfam is better than nothing, but wrong to choose over Oxfam—we can run this two-step argument. We just have to make sure that there is no plausible way for the third option to change the relative goodness of the others.

---

[127] In Chapter 6 (SRC), I more formally define and develop the link between justifiability and permission. Bader (forthcoming) rigorously recasts the argument from dynamic consistency.

[128] An example from Sen (1993: 501) involves *positionality:* suppose I must pick the second-biggest slice of cake available. It's wrong to pick *Medium Slice* from {*Small Slice*, *Medium Slice*}, but mandatory to pick it from {*Small Slice*, *Medium Slice*, *Big Slice*}. This is nothing like my choice from {*Do Nothing*, *Save 1*, *Save 2*}. (Whereas Horton's case violates β, Sen's case violates α and γ; see fn. 126, above.)

## 5. Horton's Own View

I have argued for my take on the All or Nothing Problem. Assuming that supererogation is possible, the weak link is Worse to Do Wrong, and there is a powerful case for thinking that saving one is wrong and yet better than nothing. Now I want to consider an alternative take, Joe Horton's own solution, which is in some ways similar to mine—he appears to have something like the notion of a prerogative—but his view is in a way more complex, and more skeptical of supererogation.[129]

Horton's view is that my reasons and permissions depend on what I am willing to give up (2017: 97). He thinks it is permissible to save zero just if I have a justification for doing so to which I could "reasonably appeal." The cost of my arms is a fine justification, he thinks, but I can't reasonably appeal to it if I'm *willing* to give them up (even if I'm only willing to give them up to save the first kid). However large a sacrifice I'm willing to make, I not only have to make it; I have to make the most of it. So Horton rejects May Save None—except in the case where I am unwilling to save anyone. This means that in the original case, if I *am* willing to sacrifice, Horton can reconcile Wrong to Save One, Better to Save One, and Worse to Do Wrong. The key is the conditional: if I'm willing to save anyone, I have to save both children.

I have three quick moral objections. First, obligations shouldn't get stricter the more willing one is to make sacrifices (see Kamm 1996: 315). On Horton's view, the reward for an open heart is less moral freedom. Second, the view entails that if I'm unwilling to save anyone, it's better to save zero children than one. (For Horton, saving one is wrong no matter what I'm willing to do, and it's worse to do wrong.) Horton's defense is that his view won't "discourage anyone who is willing to save one child from doing so" (Horton 2017: 97, fn. 8). That may be true, but the view still oddly

[129] McMahan ultimately accepts a similar position, though he at one point suggests that saving one is "both wrong and permissible," adding, "I offer this suggestion without being confident that it is coherent" (2018: 100). It is remarkable that McMahan would sooner allow for permissible wrongs than he would give up Worse to Do Wrong—a testament to the allure of the permissibility bar!

entails that the relative values of saving one and zero *change* depending on one's willingness. Again, I think changes in value cry out for explanation: saving zero shouldn't become the better option just because I've lost my willingness to pay the cost of heroism. Third, supposing that I am willing to pay the cost, I don't see why appealing to that cost has to be "unreasonable." Suppose you know that I'm willing to donate my kidney to a needy stranger, and you say: "Well, given your willingness, it's wrong not to donate. So I hereby demand, on behalf of the moral community, that you surrender that organ." This seems to me not just nosy but outrageous. My willingness to give the kidney doesn't entitle anyone to it—not unless I have expressed my will in a binding agreement. So long as the kidney is still mine, I can say so, and thereby justify keeping it.

Finally, a semantic point. The conditional that expresses Horton's view—"If I'm willing to save anyone, I have to save both"—might seem truer than it is thanks to an ambiguity. As Frank Jackson (1985: 183) points out in his discussion of 'given' clauses, conditionals serve two functions in statements of obligation. They can *restrict the options we're evaluating*, as in this famous sentence from the Gentle Murder Paradox: "If you're going to murder someone, you have to murder them gently" (Forrester 1984). What this means is that, in a choice just among the options in which you murder someone, you are obliged to pick the best option available—the gentlest murder on the menu. So understood the claim is true. But there's a "paradox" because of another tempting reading, on which the conditional invites us to *hold fixed the antecedent* as we evaluate our (unrestricted) options. For instance: "If that aspirin is poisonous, giving it to your patient is wrong." On this reading, I can validly infer from the conditional and the non-normative antecedent to the truth of the normative consequent; deontic logicians call this inference "factual detachment."[130] But clearly I can't do factual detachment on the gentle murder sentence; even if I am in fact going to murder someone, it doesn't

[130] As opposed to *deontic detachment*: inferring from Obligatory(if *p*, *q*) and Obligatory(*p*) to Obligatory(*q*). See e.g. the supplement on Chisholm's Paradox in McNamara 2018.

follow that, among all my actual options, what I have to do is murder gently—I still have the option of murdering no one!

So what should we make of Horton's conditional? I think it's like "If you're going to murder someone, you have to do it gently." It's true *only if* it's restricting our options—in this case, to saving one or two—and asking what's required in a choice between them. But the conditional is false if read in the other way, which would allow us to factually detach: Horton infers from "I'm willing to save someone" to "I must save both, *even though costlessly saving zero is still an option*." We can reject this inference even if we accept the conditional on its "restriction" reading.

## 6. Conclusion

Being an ineffective altruist can be wrong, and yet no worse than permissibly doing nothing. I've argued that this view is intuitive and that it flows from a natural theory of supererogation. That theory uses reasons, which make acts better, and prerogatives, which justify without favoring, thus allowing us to do less than best.

The link between prerogatives and ineffective do-gooders is simple. Suppose our prerogatives protect one suboptimal act (like keeping my arms) but don't protect a second, better suboptimal act (like giving my arms to save one life instead of two): then a wrong act will be better than a permissible one. Similar points hold for our prerogatives of self-defense, the firm's prerogative not to strike a deal, and the altruist's prerogative to keep what she owns. Gratuitously bad defensive moves, employment schemes, and philanthropic gifts can all be better than nothing and still seriously wrong.

In the end, we are left with a meta-normative question. What is the *point* of wrongness, if it doesn't track what's worse? If it can be better to do wrong, what's the problem with doing wrong?

My answer is that wrongness isn't a cutoff between worse and better, meant to guide us toward the good; it is instead a line in the sand. By doing wrong, one does what can't be justified to the moral community (Darwall 2013a, 2013b), which opens one up to sanctions and enforcement: blame, coercive prevention, perhaps even punishment.[131] The crucial point is that *wrong* choices in particular, not just *bad* choices in general, entail a loss of moral standing. If I am about to do wrong, people can legitimately demand that I refrain, whereas they may not demand that I go beyond the moral minimum. ("Don't murder!" is fine; "Go be a hero!" is bossy.) And this distinction holds even when the minimum is worse than the wrong act, as in an All or Nothing Problem. People can't expect me to give up my limbs to save a life, but they do have the standing to demand that I not save only one child, when I could just as easily save two—though actually *pressing* this demand might be a bad move, if it will dissuade me from saving anyone!

This point—that blaming might backfire in "better to do wrong" cases—is worth reflecting on. It tells us something about the complex moral texture of better-but-wrong acts, speaking to the normative question of how we ought to *react* to the wrong choice in All or Nothing Problems. Consider a real-world case: the garment industry in India.

> India is the world's second largest manufacturer and exporter of fashion garments after China, with some 13 million people working in factories within its supply chain alone. But millions more are employed in less formal settings and, according to [a new report from the University of California, Berkeley] — titled "Tainted Garments" and written by Siddharth Kara, an expert on contemporary slavery — many are women and girls from historically oppressed ethnic communities or Muslims who work from home, the majority for long hours and in hazardous conditions, earning as little as 15 cents per hour.[132] (Paton 2019)

These garment workers are exploited: they have few other options for employment, and so they

---

[131] Contrast this view of moral obligation with Derek Parfit's: "If…we often had most reason to act wrongly, morality would be undermined. Like other normative requirements, *moral requirements matter only when they give us reasons*" (2011: 7, emphasis added). Later he continues: "If we had no reasons to care about morality, or to avoid acting wrongly, morality would…have no importance" (2011: 147).

[132] See also Kara 2019.

tolerate dismal pay, exhausting work, and petty cruelty from their managers. The companies who contract with these managers, I think, are morally obligated to be more transparent about their supply chains and more fair to their workers, if they are going to be doing business in India. The companies who rely on exploitative labor practices are doing wrong; their actions and unjustifiable and deserve blame.

But Kara (2019), the researcher behind "Tainted Garments," didn't use his research to blame anyone:

> Foreign brands found to be involved — "largely household names," said Mr. Kara — were not named in the report in an effort to discourage them from pulling out of contracts or from limiting economic opportunity. "We could name and shame them, but it could be more successful to try and take a more constructive avenue here," Mr. Kara said. "These women and girls may only earn pennies but they are crucial ones. If the brands simply pulled out and they lost their home work, it could be disastrous for them and their families." (Paton 2019)

It is wrong to pay someone mere pennies when you can afford to pay more—but even pennies can be crucial. Blaming isn't always worth it. More generally, it's a bad idea to wag the finger at wrongdoers when that will only cause them to do something (permissible but) worse. Blaming Big Garment might vanish crucial pennies; shaming Shmoxfam donors might just redirect their cash to first-world luxuries. The right reaction to wrongdoers isn't always to throw the moral book, even if they deserve it. Sometimes we ought to feel a tense restraint.

If "better to do wrong" cases can teach us anything about wrongness, they teach that it is complex and multidimensional—sensitive to both reasons and prerogatives, linking individual choices to communal sanctions, suboptimal but not always a good idea to blame. We cannot just plop down a permissibility bar, with the bad verboten acts below and good permissible acts above, and call it a meta-normative day. For better or for worse, wrongness is more than a bar on a line.[133]

---

[133] I would first like to thank Joe Horton, both for his fascinating paper and his amazingly prompt comments on this paper's first draft. My warmest thanks also to the University of Vermont's Ethics Reading Group: Terence Cuneo, Tyler Doggett, Kate Nolfi, and Justin Zylstra, and to audiences at the Australasian Association of Philosophy Conference 2018 and at MIT's Work in

**REFERENCES**

Archer, Alfred (2016). "Moral Obligation, Self-Interest, and the Transitivity Problem," *Utilitas*, XXVIII, 4: 441–464.

-------(2018). "Supererogation," in *Philosophy Compass* XIII, 3.

Bader, Ralf (forthcoming). "Agent-Relative Prerogatives and Suboptimal Beneficence," in *Oxford Studies in Normative Ethics*.

Benn, Claire (2017). "Supererogatory Spandrels," in *Ethics & Politics* XIX, 1: 269–290.

Bradley, Ben (2016). "Against Satisficing Consequentialism," in *Utilitas* XVIII: 97–108.

Chang, Ruth (2002). "The Possibility of Parity," *Ethics*, CXII, 4: 659–688.

-------(2014). "Practical Reasons: The Problem of Gridlock," in Barry Dainton and Howard Robinson (Eds.), *The Bloomsbury Companion to Analytic Philosophy*. Continuum Publishing Corporation. 474–499.

Dancy, Jonathan (2004). *Ethics Without Principles*. Oxford: Oxford University Press.

Darwall, Stephen (2013a). "Morality's Distinctiveness," in *Morality, Authority, and Law: Essays in Second-Personal Ethics I*. Oxford: Oxford University Press: 3–19.

-------(2013b). "Bipolar Obligation," in *Morality, Authority, and Law: Essays in Second-Personal Ethics I*. Oxford: Oxford University Press: 20–39.

-------(2013c). "'But It Would Be Wrong'," in *Morality, Authority, and Law: Essays in Second-Personal*

---

Progress Seminar. For discussion, I am grateful to David Balcarras, Nathaniel Baron-Schmitt, Dan Bonevac, Joe Bowen, Paul Butterfield, Jonathan Dancy, Sinan Dogramaci, Tomi Francis, Nick Geiser, Brian Hedden, Mirjam Müller, Anni Räty, Kirun Sankaran, Mark Schroeder (who turned me on to the All or Nothing Problem), Katie Steele, Anna Waldman-Brown, Quinn White, and Paul Woodruff. For detailed comments on some or other version(s) of the paper, I thank the editors of *Philosophy and Public Affairs*, Arif Ahmed (who gave me several rounds of comments), Renee Jorgensen Bolinger, Thomas Byrne, Brendan de Kenessey, Elliot Goodine, Joe Horton, Seth Lazar, Rose Lenehan, Katy Meadows, Theron Pummer, Tamar Schapiro, Ginger Schultheis, Kieran Setiya, Brad Skow, Jack Spencer, Judy Thomson, and Steve Yablo. Special thanks to Caspar Hare for encouragement early on and Kieran Setiya for insightful advice throughout.

*Ethics I*. Oxford: Oxford University Press: 52–71.

Gert, Joshua (2007). "Normative Strength and the Balance of Reasons," in *The Philosophical Review*, CXVI, 4: 533–562.

Harman, Elizabeth (2016). "Morally Permissible Moral Mistakes," *Ethics*, CXXVI, 2: 366–393.

Heyd, David (1982). *Supererogation*. Cambridge: Cambridge University Press.

-------(2016). "Supererogation," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), URL = <https://plato.stanford.edu/archives/spr2016/entries/supererogation/>.

Horton, Joe (2017). "The All or Nothing Problem," in *The Journal of Philosophy* CXIV, 2: 94–104.

Hurka, Thomas and Shubert, Esther (2012). "Permissions to Do Less Than Best: A Moving Band," in *Oxford Studies in Normative Ethics Volume 2*: 1–27.

Jackson, Frank (1985). "On the Semantics and Logic of Obligation," *Mind*, XCIV, 374: 177–195.

Kagan, Shelly (1989). *The Limits of Morality*. Oxford: Oxford University Press.

Kamm, Frances (1996). *Morality, Mortality, Volume II: Rights, Duties, and Status*. New York: Oxford University Press.

Kara, Siddharth (2019). *Tainted Garments: The Exploitation of Women and Girls in India's Home-Based Garment Sector*. Blum Center for Developing Economies: University of California, Berkeley.

Lazar, Seth (2012). "Necessity in Self-Defense and War," in *Philosophy and Public Affairs*, XL, 1: 3–44.

Lazar, Seth and Barry, Christian (ms.). "Acting Beyond the Call of Duty: Supererogation and Optimization."

McMahan, Jeff (2018). "Doing Good and Doing the Best," in *The Ethics of Giving: Philosophers' Perspectives on Philanthropy*, Paul Woodruff (Ed.). New York: Oxford University Press: 78–102.

McNamara, Paul (1996). "Making Room for Going Beyond the Call" in *Mind* CV, 419: 415–450.

-------(2018). "Deontic Logic," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall

2018 Edition). URL = <https://plato.stanford.edu/archives/fall2018/entries/logic-deontic/>.

Parfit, Derek (1982). "Future Generations: Further Problems," *Philosophy and Public Affairs*, XI, 2: 113–172.

-------(2011a) *On What Matters: Volume One*. Oxford: Oxford University Press.

-------(2011b) *On What Matters: Volume Two*. Oxford: Oxford University Press.

Paton, Elizabeth (2019). "Made for Next to Nothing. Worn by You?" in *The New York Times*, February 6, 2019.

Portmore, Douglas (2003). "Position-Relative Consequentialism, Agent-Centered Options, and Supererogation," *Ethics*, CXIII: 303–332.

-------(2008). "Are Moral Reasons Morally Overriding?" *Ethical Theory and Moral Practice*, XI: 369–388.

-------(2011). *Commonsense Consequentialism*. Oxford: Oxford University Press.

Pummer, Theron (2016). "Whether and Where to Give," *Philosophy and Public Affairs*, XLIV, 1: 77–95.

-------(forthcoming). "All or Nothing, but if Not All, Next Best or Nothing," in *Journal of Philosophy*.

Raz, Joseph (1975). "Permissions and Supererogation," in *American Philosophical Quarterly* XII, 2: 161–168.

Ross, W.D. (1930). *The Right and the Good*. Oxford: Oxford University Press.

Rulli, Tina and Worsnip, Alex (2016). "IIA, Rationality, and the Individuation of Options," in *Philosophical Studies* CXXIII: 205–221.

Scheffler, Samuel (1982). *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Oxford University Press.

Sen, Amartya (1993). "Internal Consistency of Choice," in *Econometrica*, LXI, 3: 495–521.

-------(2017). *Collective Welfare and Social Choice: Expanded Edition*. Cambridge: Harvard University

Press.

Sinclair, Thomas (2018). "Are We Conditionally Obligated to be Effective Altruists?" in *Philosophy and Public Affairs*, XLVI, 1: 36–59.

Snedegar, Justin (2015). "Contrastivism About Reasons and Oughts," in *Philosophy Compass*, X: 379–388.

-------(2016). "Reasons, Oughts, and Requirements," in R. Shafer-Landau (Ed.) *Oxford Studies in Metaethics*, XI. Oxford: Oxford University Press: 183–211.

Tadros, Victor (2011). *The Ends of Harm: The Moral Foundations of Criminal Law*. Oxford: Oxford University Press.

Thomson, Judith Jarvis (1985). "The Trolley Problem," in *The Yale Law Journal*, XCIV, 6: 1385–1415.

-------(1990). *The Realm of Rights*. Cambridge: Harvard University Press.

Woodruff, Paul (2018). "Afterword: Justice and Charitable Giving," in *The Ethics of Giving: Philosophers' Perspectives on Philanthropy*, Paul Woodruff (Ed.). New York: Oxford University Press: 204–220.

Yetter Chappell, Richard (forthcoming). "Willpower Satisficing," in *Noûs*.

Zwolinski, Matt (2007). "Sweatshops, Choice, and Exploitation," in *Business Ethics Quarterly* XVII, 4: 689–727.

# CHAPTER SIX

## Supererogation and Rational Choice:

## Incommensurability, Intransitivity, Independence

**Supererogation and Rational Choice:**

**Incommensurability, Intransitivity, Independence**

## 1. Introduction

We don't always have to do what's best. Some good deeds are *supererogatory:* they are optional and yet better than other permissible options. Examples include enduring injury to save a life, doing a favor, and granting forgiveness.[134] Very lovely; hardly required.

But as innocent as the examples may sound, the theory of supererogation is fraught with paradoxes. Three in particular are fairly wide open: the classic Paradox of Supererogation, Frances Kamm's (1985, 1996) Intransitivity Paradox, and Joe Horton's (2017) All or Nothing Problem. The classic Paradox asks: if "supererogatory" acts are really better, why aren't they just required? This question has a bite even in the simplest cases, like the choice between doing nothing and saving a life at great personal cost. The other paradoxes center on the strange effects of adding a third option. Kamm adds the option to do a "duty;" like keeping a promise to meet a friend for lunch. Horton adds the option to supererogate even more splendidly, saving two lives instead of one, at no extra cost. The strangeness comes out once we try to rank the options in a coherent way.

These three puzzles, though rarely mentioned together, are deeply linked, and they need a unified treatment.[135] The problem is that there are two internally coherent solutions to Kamm and Horton's paradoxes: we can either make moral rankings *menu-relative* or *intransitive*. How to decide? The key fact, I think, is that only intransitivity follows from a solution to the classic Paradox of Supererogation, involving *incommensurability* between reasons (which make acts better) and

[134] When I say that supererogation is "better" than other permissible action, I mean that there is more reason overall to do it, not that it nicer effects or merits more praise. (On praise and supererogation, see Massoud 2016.) Nothing hangs on this; later, I will show that my arguments can be run for other conceptions of supererogation. They all involve incommensurability, in the relevantly broad sense.

[135] An exception is Rulli and Worsnip (2016: fn. 15), who link a Kamm-style case to a Horton-style case.

prerogatives (which allow us to be suboptimal). Though intransitivity may seem strange, it is exactly what we should have expected given the right view about what supererogation is.

After laying out the paradoxes (§2), I show that Kamm's and Horton's have the profile of a transitivity failure: both violate a principle called Property β without violating its cousin Property α, also called the *Independence of Irrelevant Alternatives* (§3). I then explain why some writers nonetheless pin Horton's paradox on menu-relativity, which is usually linked to violations of Independence (§4). From there, I turn to incommensurability: why it's essential to supererogation, and why it entails intransitivity alone (§5). (In Appendix 1, I extend the paradoxes to concoct yet more intransitivities. Appendix 2 takes up other writers' solutions.)

The goal of all this is to better understand our interlocking paradoxes: to explain Kamm's and Horton's discoveries using only basic facts about what makes supererogation possible. My conclusion (§6) is that supererogation entails incommensurability and intransitivity, also allowing for Independence. Supererogation isn't really paradoxical—it's just more interesting than we expected.

## 2. Three Paradoxes of Supererogation

The original "Paradox of Supererogation" stars an agent who may choose either of two options, one of which is better than the other. For example, suppose it's better to save a person from a falling building, at the cost of injuring oneself, than to safely wait outside. The puzzle is to explain why it isn't wrong to pick the worse option. If being the hero is so great, why isn't it obligatory?[136]

The other paradoxes add a third option. In Joe Horton's "All or Nothing Problem," an agent must choose between safely doing nothing, rushing into a building to save one person, and rushing in to save the one person *and* a second—supposing that the injury and effort are the same

[136] See e.g. Raz 1975, Heyd 1982, 2016: §3, Archer 2017, and Chapter 4 (WSW), above. The paradox can be put in terms of two principles: (1) if an option *x* is better than an option *y*, then there is more reason to do *x* than *y;* (2) if there is more reason to do *x* than *y*, then *y* is wrong.

whether the agent saves one or two. Intuitively, saving only one person is wrong, since it's unjustifiable to gratuitously let someone die; and yet, wrongly saving one still seems better than permissibly saving zero. How could it be better to do wrong? Presumably, being permissible is a matter of being "good enough." Permissible options are the ones that, on the ranking from worst to best, manage to make it over a certain cutoff, the "permissibility bar" (to borrow a phrase from Snedegar 2016: 161). But then surely anything that outranks, or even ties with, a permissible act must also be over the bar, and therefore must also be permissible.[137] There's the rub: saving only the one person, in Horton's case, seems wrong even though it's better than permissibly doing nothing.

In Kamm's "Intransitivity Paradox," meanwhile, the options are intriguingly similar. An agent must choose between safely doing nothing, running into a building to save one (again, at personal cost), and doing a "duty"—Kamm's example is keeping a promise to meet someone for lunch. Intuitively, life-saving and promise-keeping are both optional here, whereas it's wrong to do nothing. But there is a curious intransitivity. In a choice between doing nothing and saving the life, I may do nothing; in this sense, doing nothing may "take precedence over" supererogation. And supererogation may take precedence over keeping the promise. But doing nothing may not take precedence over promise-keeping; if those are my only options, I have to keep my word. So "may take precedence over" is intransitive. It is natural to think, however, that whether one option may take precedence over another is a matter of how the two compare on a ranking from worst to best—and surely if one outranks the other, it must also outrank whatever the other outranks. Again, think of the permissibility bar. If an option is over (or tied with) an option that is over the cutoff, how could the first option *fail* to be over the cutoff, too? How could precedence be intransitive?[138]

---

[137] Horton 2017, McMahan 2018, Sinclair 2018. I defend the intuitions in Chapter 5 (BDW). Horton's case is also considered in Parfit 1982: 131, Kagan 1989: 16, Tadros 2011: 161–62, Portmore 2012: 146–47, and Pummer 2016.

[138] See Kamm 1985, 1996: Chapter 12, 2007: 30–1. Dorsey (2013) raises the same basic problem, drawing replies from Archer 2016 and Portmore 2017; see Appendix 2, below.

The similarity between Horton's case and Kamm's is striking.[139] Both can be constructed by starting with the classic Paradox and adding a third option. But is the third option playing a similar role in both cases? Are the resulting paradoxes alike in any deep way?

### 3. α Without β

Kamm's and Horton's cases are united by a distinctive kind of *option-sensitivity*. To say that a moral judgment is "option-sensitive" is to say that it depends on the presence of options other than the ones being judged—e.g. we might say that *x* is better than *y* only when *z* is on the menu.

What's distinctive about our cases? Notice, first, that they are in a way option-insensitive: taking away an option won't make any of the remaining ones wrong. (For example: if I lose the option to be the hero in Kamm's case, it's still fine to meet you for lunch.) Both cases thus obey the *Independence of Irrelevant Alternatives*, which Sen (1969, 2017) calls:

> *Property a*
> If *x* is permissible to choose from a set of options O, then if *x* is in a subset O* of O, *x* is permissible to choose from O*.

Property α says that you can't make an option wrong just by removing other options; wrongness is "contraction consistent." Sen considers α a "very basic requirement on rational choice" (2017: 63). It certainly sounds nice a priori. If *x* wins a permissibility contest against *y* and *z*, "why would it not also be a winner against *y* alone?" (Tugodden and Vallentyne 2005: 143) Presumably, the rules of the game would have to be changed by *z*'s presence. That would raise the question of what makes *z* so special.[140]

---

[139] Except that Kamm's (1985) case features a kidney donation rather than a rescue mission.

[140] There is also something odd about α-flouting preference, as in Sidney Morgenbesser's joke:

> BARTENDER: Would you like red wine, white wine, or beer?
> PATRON: White wine, please.
> BARTENDER: My apologies—we're actually out of beer.
> PATRON: In that case, I'll take the red.

Instead of eschewing α, our two paradoxes violate its less famous cousin, a principle of "expansion consistency" known as:

> *Property β*
> If *x* and *y* are permissible to choose from a set of options O, then if *x* is permissible to choose from a superset O* of O, so is *y*. (Sen 1969, 2017)

The idea behind β is simple: if two options are both permissible, adding an option can't make *only one* of them wrong. Equally permissible acts are made wrong by the same things. Property β is "perhaps somewhat less intuitive than property α," Sen admits, but is it "also appealing" (2017: 64).[141]

Why is β supposed to be appealing? Well, to its credit, it certainly holds in cases where the permissible options are more or less the same. Suppose I'm allowed to give you my pain pill using either my left hand or my right. If I then get a vastly better third option—giving the pill to someone else who needs it even more—presumably I won't be allowed to give you the pill by either means. If I get a worse new option—throwing the pill away—presumably I will still be allowed to hand it to you however I would like. Since the two ways of helping you are basically similar, they are made wrong by the very same things. Property β gets obeyed.

And yet, β is flouted by our paradoxical cases. We start with the classic Paradox's options: *Do Nothing* and *Save 1*. Only one of these becomes wrong when we add a third option. Kamm adds *Keep Promise*, which makes *Do Nothing* wrong. Horton adds *Save 2*, which makes *Save 1* wrong. Both cases are β-violations; the difference is just a matter of which of the original options is beaten by the newcomer.

That is the link between our two paradoxes: they violate β while respecting α. Removing options won't make anything wrong, but adding options can rule out one previously kosher choice without ruling out the others.[142]

---

[141] Sen calls these "properties" because they are meant to characterize choice functions; see fn. 145, below.

[142] You might wonder: isn't there an easy way to rule out option-sensitivity? What if we

Now, what is the upshot? What follows from this kind of option-sensitivity? In a nutshell: we need α to capture the facts about wrongness with a binary relation, and β makes the relation transitive.[143]

What is this relation? It's what Kamm has in mind when she talks about "precedence." I call it "weak defeat," and express it with '≥'. The idea behind this relation is that it tells us what is justifiable to choose over what, which in turn can tell us what is permissible.

*Weak Defeat*
$x \geq y$ *iff* one can justify choosing $x$ over $y$.

*Permissions and Defeat*
$x$ is permissible to choose from a set of options O *iff* for all $y \in O$, $x \geq y$.

In other words: permissible options are ones we can justify over anything on the menu, and "weak defeat" expresses justifiability.[144]

From there, we can define up the notions of a tie and of "strict" defeat, which I'll just call "defeat."

*Tie*
$x = y$ *iff* $x \geq y$ and $y \geq x$.

*(Strict) Defeat*
$x > y$ *iff* $x \geq y$ but not $y \geq x$.

---

redescribe each option so that it belongs *essentially* to a single menu? (Instead of *Save 1*, we have *Save 1 (Rather Than Do Nothing)*.) But this move just obscures the influence of context, which can't be stipulated away. Clearly, there is some interesting relation between saving one in the case where I could also save two and in the case where I can't. We want to know: what difference does the additional option make? For more, see Dietrich and List 2017: Appendix 2.B., Broome 1993: 100–02, Sen 1993: 501, fn. 7, Neumann 2007, and Rulli and Worsnip 2016.

[143] In an illuminating paper on Horton-style cases, Bader (forthcoming) mentions that Horton's case violates β and not α, as do certain cases of incommensurable values. (He does not discuss Kamm's case, nor does he link pure β-failures to intransitivity; his main interest in β-failures is that they might lead to dynamic inconsistencies.)

[144] In rational choice theory, defeat relations express what is *preferred* to what, rather than facts about justifiability, and the key link is to rational choice rather than moral permission (see e.g. Sen 1969). I am giving this framework a moral spin (following Dietrich and List 2017).

Now we have names for the two specific ways for *x* to weakly defeat *y*. They won't let us say anything new, but they will let us express some things more quickly, like the link between defeat and permission. We can say: a wrong act is one that is defeated by something. Mighty convenient.[145]

Where there is option-sensitivity, unusual things happen with defeat. In particular, α-failures ought to raise our eyebrows, since they make it impossible to express the facts about what is permissible in terms of weak defeat (unless we relativize it to what's on the menu). Consider an example adapted from Sen. I am choosing between slices of cake, and the rule is that I may not pick the biggest one available (1993: 501). When my options are {*Small*, *Medium*, *Large*}, I may pick *Medium*. If *Large* is taken off the menu, *Medium* becomes wrong, because it is now the biggest.[146] This violates α. Moreover, we can see that there is no way to record these shifty permissions into static relations. *Medium* must weakly defeat *Small* to be permissible in the big set, but it must fail to defeat it to be wrong in the smaller set. Instead of a binary relation between a pair, we would need a relation that is *menu-relative*. The *Medium* slice loses to the *Small* relative to the pair, but not relative to {*Small*, *Medium*, *Large*}.

Without α, then, we can't say whether *x* defeats *y* even if we know everything about their

---

[145] For more convenience, I assume that (1) every option ties itself; and (2) every set of options O is finite with (3) a choice function *f* defined over it—which means for any nonempty subset O* of O, *f*(O*) is the nonempty set of permissible options to choose from O*. When is a choice function defined over a finite set using a binary relation ≥? When ≥ is complete, reflexive, and acyclic. Acyclicity means: there are no options *x*, *y*, … such that $x > y > \ldots > x$; see Sen 2017: 62.

[146] In Sen's original example, the agent *must* pick the second-largest slice of cake. This is still an α-violation (without a β-violation), but it is not a minimal example, since it also runs afoul of:

> *Property γ*
> Suppose we have some sets of options $O_i$. If *x* is permissible to choose from any $O_i$, then *x* is permissible to choose from the union of all $O_i$.

Sen's example violates γ because *Small* is permissible from {*Small*, *Medium*} and from {*Small*, *Large*} but not from their union, {*Small*, *Medium*, *Large*}. Property γ doesn't feature in my arguments, but it is important for another reason: γ and α are severally necessary and jointly sufficient to express the permissibility facts with binary weak defeat relations; see Sen 1993: 500.

contents; we have to also know the context—the rest of the options on offer. The ranking that determines right and wrong, which is generated by the weak defeat relation, can change depending on which options are available.

What about β? Interestingly, even when it fails, we can still sometimes distill the permission facts into a binary relation. The real upshot of β-failures is that—supposing that we have already got such a relation—they make it *intransitive*. Consider Horton's case. *Do Nothing* and *Save 1* are both permissible in a pairwise choice, but only *Do Nothing* remains fine when we add the option to *Save 2*. We can easily capture this with binary relations: *Save 2* > *Save 1* = *Do Nothing* = *Save 2*. But the relation violates a kind of transitivity, which I call:

> *Transmission Over Ties*
> If $x > y$ and $y = z$, then $x > z$.[147]

And so I think our paradoxes violate exactly this principle. In Horton's case, *Save 2* defeats *Save 1*, which ties *Do Nothing*, but *Save 2* doesn't defeat *Do Nothing*. In Kamm's paradox, as she observes, *Keep Promise* defeats *Do Nothing*, which ties *Save 1*, but *Keep Promise* doesn't defeat *Save 1*. And this is what gives these cases the air of paradox. If $x$ beats $y$, and $y$ ties $z$, how could $x$ fail to beat $z$? If I beat you in a height contest, I beat anyone you tie with. If your dive score beats mine, it beats any score tied with mine. One starts to wonder why moral defeat should be different.

Let me emphasize right away that this kind of intransitivity—however odd it might seem at first—is less radical than you might think. If β fails, and defeat fails to transmit over ties, we do have to accept a kind of menu-relativity. But it isn't the relativity of defeat, as with α-failures. Instead, β-failures make it menu-relative whether two options have the same deontic status, i.e. whether they are *equally permissible*. To illustrate, suppose $y$ and $z$ are tied, and $x$ beats only $y$. ($x > y = z = x$.) Are $y$

[147] Sen (2017: Chapter 1*) calls this "PI-intransitivity;" 'P' means '>' (strict preference) and 'I' means '=' (indifference). Transmission Over Ties is equivalent to "IP-intransitivity," which says that if $x = y$ and $y > z$, then $x > z$. It is also logically weaker than the transitivity of weak defeat, stronger than the transitivity of tying, and independent of the transitivity of strict defeat ("quasi-transitivity").

and $z$ equally permissible? It depends. Are we choosing from the pair of them? If so, yes. But if we are choosing from $\{x, y, z\}$, only $z$ is permissible. This is a mild kind of relativity because permissibility is already, by nature, is a function of how an option compares to the rest of what is on the menu.[148] Permissible options are ones justifiable over all others, ones that weakly defeat all challengers. Defeat, by contrast, is a relation between pairs; indeed, it's natural to think it should be an internal relation. It is more radical, then, for defeat to be menu-relative than to have relativity for *being equally permissible*. Transmission failures only commit us to the modest kind of relativity.

Moreover, we can give up Transmission Over Ties without being forced towards the more infamous forms of intransitivity, such as cyclicity. If there are *cycles* of defeat, where $x > y > \ldots > x$, then there can be choices where no option is undefeated. Everything gets ruled out by something—which sounds like rational doom. No such threat is implied by violations of Transmission Over Ties. In Horton's case, nothing rules out *Do Nothing* or *Save 2*. In Kamm's, nothing rules out *Keep Promise* or *Save 1*. The cases are unusual, maybe, but they aren't dooming anyone.[149]

That is my take on Kamm's and Horton's cases. Since they violate Property β and respect Property α, we can understand them using a binary relation of weak defeat that, while intransitive, does not generate any problematic cycles or shocking relativities. Kamm already saw her case as involving intransitivity; now we are uniting her case with Horton's.

Some philosophers, however, do not tolerate even this mild intransitivity in Horton's case. They have formidable reasons for thinking that the case must really menu-relative defeat (the only kind of relativity I'll discuss from here on). Let's see what they have in mind.

---

[148] On a radical "non-comparative" view of permissibility, we can tell that certain options are wrong just given their intrinsic features. The only remotely plausible examples are unspeakable acts like torture that are supposedly verboten no matter the alternatives. (It's wrong to torture, the idea goes, even the alternative is Armageddon.) Since none of my cases involve such awful acts, I think it is safe to assume that permission is comparative, i.e. that permissible acts are ones that weakly defeat every alternative (where weak defeat may or may not be relativized to menus).

[149] See also Temkin 2012: 196 on the acyclicity of Kamm's intransitivity.

## 4. Intransitivity or Relativity?

I have suggested that Horton's case violates a transitivity principle called Transmission Over Ties. *Save 2* defeats *Save 1* but doesn't defeat what it ties—viz. *Do Nothing*. Why don't other writers see the case as an intransitivity?

One reason is that some writers don't distinguish the different flavors and upshots of option-sensitivity. Property β isn't kept separate from Property α, nor intransitivity from menu-relativity.[150] So we have Theron Pummer, for example, saying this about a Horton-style case:

> it is a familiar feature of nonconsequentialist ethics that the moral status of an act can depend on which alternative acts are available. In this case, the presence of [*Save 2*] alters the moral status of [*Save 1*], thereby altering the way that [*Save 1*] and [*Do Nothing*] compare morally. I believe the intuition that [*Save 1*] is wrong only if [*Do Nothing*] is wrong too has force only when considering these acts in isolation from the full choice situation. But with the full choice situation in view, it is clear that there is something to be said against [*Save 1*] that cannot be said against [*Do Nothing*] or [*Save 2*]: the performance of [*Save 1*] constitutes a deliberate refusal to do something much better at no extra cost. This is a serious moral failing. (2016: 86–7)[151]

I think Pummer is deeply right. The presence of *Save 2* means that we can say something against *Save 1* that can't be said against *Do Nothing*. But he leaves open how *Save 2* is supposed to have this effect. Does it change the tie between *Do Nothing* and *Save 1* into a defeat? That's menu-relativity. Or does *Save 2* simply defeat *Save 1* without defeating that which it ties, viz. *Do Nothing*? That's intransitivity. In this passage, Pummer doesn't commit to either of these over the other.

Others, however, endorse menu-relativity explicitly. Now, I think this kind of relativity calls out for explanation. Adding a new option doesn't intrinsically change the old ones. Why should it change which beats which? Seth Lazar and Christian Barry (ms., emphasis original) have a direct

---

[150] For example, Rulli and Worsnip (2016: 206) define the "Independence of Irrelevant Alternatives" in a way that entails both α and β. Traditionally, "Independence" refers just to α, as in Sen 1969: 384 and Rubenstein 1998: 11. (Confusingly, "The Independence of Irrelevant Alternatives" is also the name for a principle in Arrow's Impossibility Theorem, a principle that has little to do with α. See Arrow 1951 [1963], Sen 1969: 386, 390–91, and Morreau 2014: 4.5. These two principles are, understandably, sometimes mixed up by ethicists; see e.g. Kamm 1996: 344.)

[151] For simplicity, I have switched the options in Pummer's case to match those in Horton's.

answer:

> we must explain why choosing [*Save 1*] from [{*Do Nothing*, *Save 1*, *Save 2*}] is different from choosing it from [{*Do Nothing*, *Save 1*}]. But how do we do this? We think the most plausible route is to argue that, although [*Save 1*] is better than [*Do Nothing*] in *agent-neutral* terms, the availability of [*Save 2*] as an alternative makes [*Save 1*] worse than [*Do Nothing*] in *agent-relative* terms, by enough to make it worse all-things-considered).[152]

More specifically, Lazar and Barry say it is *disrespectful* to gratuitously allow somebody to die. By "so grossly and gratuitously allowing others to suffer grievous harm," the choice to *Save 1* rather than both "conveys an egregious disrespect to those whom the agent does not help" (Lazar and Barry ms.). Adding the option to save two thus adds agent-relative reasons against saving only one—and these reasons favor doing *anything else* over saving one, including doing nothing.

This view faces a challenge. Suppose that a third party is watching as I am about to wrongly *Save 1*, and while they can't make me *Save 2* (nor can they save anyone themselves), they do have the option to prevent me from going into the building. Surely that would be wrong. But if it's really worse to save one than none, why isn't it obligatory to prevent me from saving one? Lazar and Barry have a nice answer: reasons of respect are *agent-relative*, so they guide the deliberation of the agent, but not the interventions of third parties. Saving one is worse all-things-considered, but best impartially, and so onlookers shouldn't intervene when it leads to more badness on the whole.

So I think that Lazar and Barry have defended their view well, and that it is the best version of the view that Horton-style cases force us toward menu-relativity. But I want to note two things. First, nothing about the permissibility facts *themselves* entails menu-relativity. The case violates Property β, not Property α, which means that we could in principle capture the permission facts with a binary, intransitive relation of weak defeat. To rule this out, Lazar and Barry would need a further principle, based in substantive moral views.

---

[152] Again, I have changed the option's names for simplicity: Lazar and Barry use "option 1," "option 2", and "option 3." Also, while they are discussing a Horton-style case, their view works the same in a Kamm-style case: *Keep Promise* makes *Do Nothing* worse than *Save 1*.

But, second, they have exactly such a principle. In particular, they assume a (very tempting) link between reasons and permissions:

*Worse to Do Wrong*
If A is wrong and B is permissible, then there is more reason to do B than to do A.[153]

And Lazar and Barry also think that, in the pairwise choice, there is more reason in favor of *Save 1* than *Do Nothing*. So the ranking must invert when we add in *Save 2*, and so they naturally conclude that the new option brings with it new reasons against saving only one. These new reasons convert the tie between *Do Nothing* and *Save 1* into a defeat. When there are reasons against disrespecting the second victim, *Do Nothing* > *Save 1*.

The result is a filled-out solution using relativity instead of intransitivity. What beats what depends on the menu, but hold the menu fixed, and weak defeat is transitive. While this view has some costs—accepting menu-relativity, positing extra reasons against disrespect—it also has a clear advantage, in that it explains how *Save 1* could be excellent to choose over *Do Nothing* but wrong when the menu also features *Save 2*. The view is coherent and principled, being motivated by Worse to Do Wrong. It is also quite different from the view that Horton's case violates transitivity.[154]

So which do we choose: intransitivity or menu-relativity? Both have their advantages. To break the tie, I suggest we return to the question of what supererogation has to be like to be possible—the Paradox of Supererogation. Our solution will naturally lead to intransitivity, but not menu-relativity, while also debunking Worse to Do Wrong.

## 5. From Incommensurability to Intransitivity

### 5.1 *Transmission Failures and Threshold Tricks*

[153] For endorsements, see Darwall 2013: 59, Horton 2017: 96, and McMahan 2018.

[154] If defeat is menu-relative, we should expect violations of Independence (α). My view doesn't predict this.

How could there be moral intransitivity? How could defeat fail to transmit over ties? The relevant tie: *Save 1* = *Do Nothing*. Each is justifiable to choose over the other, in the classic Paradox, and yet they are defeated by different things in Kamm's and Horton's. How can *Keep Promise* defeat only *Do Nothing*? How can *Save 2* defeat only *Save 1*?

The fundamental fact here, crucial to all three paradoxes, is that *options can be tied even though they are morally different*. Tied options don't have to be indistinguishable—or even close. Only for a certain special class of (binary weak defeat) relations will ties always be relevantly alike. Consider "is at least as tall as." If two people are at least as tall as each other, they must be the same height. Similarly, if two cities are at least as populous as each other, they must have the exact same number of people; they must be exactly relevantly alike when it comes to population size. But plenty of other "defeat" relations don't require homogeneity amongst tied options. Consider the relation of being as tall *or* almost as tall as, which we'll call '$\geq_{AT}$' (for "Almost as Tall"). To make it precise, suppose $x \geq_{AT} y$ just if the height of $x$ is greater than or equal to the height of $y$ minus a foot. Now let's say Amanda is 5', Bert is 6', and Caroline is 7'. Clearly Amanda and Bert are tied with respect to '$\geq_{AT}$'. Amanda $\geq_{AT}$ Bert (since she is within a foot), and Bert $\geq_{AT}$ Amanda (since he is taller). But there's a difference in how the two relate to Caroline. Bert $\geq_{AT}$ Caroline (and vice versa), but it's not true that Amanda $\geq_{AT}$ Caroline. Caroline "defeats" Amanda, who ties Bert, and yet the "defeat" doesn't transmit, precisely because Bert is *taller* than Amanda even though they tie.

Something similar, I think, goes on in the classic Paradox of Supererogation. The supererogatory act and moral minimum are "tied" yet relevantly different; the moral minimum seems harder to justify. But what's the key difference? I think there is more *reason* to supererogate; being the hero is the worthier choice.[155] But I don't think extra reasons can be the only difference. If

[155] See Hurka and Shubert 2012, Harman 2016: 370, fn. 8, Snedegar 2016: 162. Also, let me emphasize again that not much hangs on this view; we will extend the arguments to some other accounts of supererogation in §5.3.

they were, then we would need our weak defeat relation to be somehow brutely forgiving of inferior options, which leads to familiar problems.

To illustrate, let's consider a simple "threshold trick." Suppose that weak defeat is a matter of having at least as much reason *or* being close enough (along the lines of $\geq_{AT}$). To make the threshold of closeness precise, suppose that the weight of reasons to do an option $x$—written R($x$)—can be measured with a real number, and suppose that $x \geq y$ just if R($x$) is greater than or equal to R($y$) – 10. Now we can coherently describe the classic Paradox: let's say R(*Do Nothing*) = 5 and *R*(*Save 1*) = 10. Less reason to do the minimum, but it's close enough. We can also capture the possibility that *Do Nothing* = *Save 2* in Horton's case; suppose R(*Save 2*) = 10. Even so, *Do Nothing* is still close enough. But here's the problem. We can't capture the wrongness of *Save 1*. If *Do Nothing* is good enough, and *Save 1* is even better, then surely *Save 1* is good enough, too. A similar point is true in Kamm's paradox, where (let's suppose) the agent has least reason to do nothing, more reason to keep the promise, and most reason to save the one. If *Do Nothing* is close enough to be permissible given *Save 1*, and *Keep Promise* is no better than *Save 1*, how could *Do Nothing* be made unjustifiable by *Keep Promise*? If R(*Do Nothing*) is within 10 of R(*Save 1*), and *R*(*Keep Promise*) is even closer, there is no way that the prospect of promise-keeping could make it wrong to loaf around; it cannot have more prohibitive power than the supererogatory *Save 1*. So a simple threshold isn't enough.

### 5.2 *Reasons and Prerogatives*

We need a further way for options to tie, something more complex than one's being "close enough" to a fixed threshold along the lone relevant dimension. Here we should borrow an insight from Kamm.[156] *Tied options might differ along multiple relevant dimensions*. Reasons, which favor acts and ground

[156] Kamm (1996: 336–37) writes: "The exchange between [*Keep Promise*] and [*Save 1*] represents playoffs of one component of objective morality, rights and duties (representing the objective

requirements, are not the be-all and end-all of ethics. We also have *prerogatives* to act against the balance of reasons. Prerogatives don't count in favor of anything; they are pure justification, tending to make permissible without tending to make choiceworthy or required—the kind of thing we mean when we say "I have a right not to get harmed (so I may decline to save someone from the falling building)," and "The kidney is yours (so they can't demand that you give it up)."

With reasons and prerogatives, we can say what it takes to justify an action—we can morally define weak defeat—in a way that makes sense of supererogation. The idea is simple:

> *The Prerogatives Principle*
> $x \geq y$ *iff* the reasons to do $y$ (rather than $x$) cannot outweigh the combined reasons and prerogatives to do $x$ (rather than $y$).[157]

Or to put it a bit more compactly, with P($x$) representing the weight of the prerogative to do $x$:

> *The Prerogatives Principle*
> $x \geq y$ *iff* $R(x) + P(y) \geq R(y)$.[158]

The idea this: we are always justified in doing what we have more reason to do over what we have less reason to do; we are justified in doing less if, and only if, we have the right to do so—i.e. a strong enough prerogative.

On this view, two options tie can tie even if the reasons prefer one over the other. This is what happens, I think, in the classic Paradox. There is more reason to *Save 1*, which raises the question: why don't we have to do it? The answer is that one has a prerogative, in this case, to save

---

characteristics that make a person an end-in-itself) against another component of objective morality, which is more concerned with welfare considerations and related well-being values, but may also be derived from concern that persons who are ends-in-themselves succeed in pursuing their conceptions of the good." I agree that there are two components, but not the ones Kamm thinks.

[157] I give some defense of this principle, and try to develop a theory of reasons, prerogatives, and wrongness, in the earlier chapters of this dissertation, especially Chapter 3 (FRP), Chapter 4 (WSW), and Chapter 5 (BDW). For a defense of prerogatives ("prima facie permissions"), see Hurka and Shubert 2012. An alternative notion is the idea of reasons with "justifying strength" (Gert 2007).

[158] (Apologies for using '≥' in two ways here.) This will need to be slightly amended to deal with essentially contrastive prerogatives (see below): instead of "P(B)," we need "prerogative to do B *rather than A*." (Same for contrastive reasons.) I signal this in the informal Prerogatives Principle.

one's skin instead. In other cases, prerogatives protect the non-optimific use of one's property (I don't have to lend my pencil), body parts (you don't have to give your kidney), etc.

But unlike the threshold trick, the Prerogatives Principle naturally extends into a solution for Kamm's and Horton's paradoxes. Start with Horton. How could *Save 1* be better than the permissible *Do Nothing* and yet still wrong in a choice from {*Do Nothing*, *Save 1*, *Save 2*}? The answer is that, in this case, prerogatives protect the worst option, *Do Nothing*, more strongly than they protect the second-worst, *Save 1*. We have a prerogative not to injure ourselves even for noble ends, but no comparable prerogative to gratuitously let die. We can represent the case like so:

| | Do Nothing | Save 1 | Save 2 |
|---|---|---|---|
| Reasons | 0 | 5 | 10 |
| Prerogatives | 10 | 0 | 0 |

In this case, thanks to prerogatives, a wrong act is better than a permissible one. It isn't always worse to do wrong.

Crucially, to get this result, we need reasons and prerogatives to be *incommensurable*, in the sense that the reasons and prerogatives to do one action over another can't be expressed by those acts' positions along a single scale. The balance of reasons doesn't fix the balance of prerogatives; "more reason" doesn't entail "more prerogative." This is essential to treating Horton's case, where prerogatives protect the worst option more than its betters, so that it isn't always worse to do wrong.

To solve Kamm's paradox, however, incommensurability by itself is not enough. We are supposing that there is more reason behind *Keep Promise* than *Do Nothing*, and that there is more reason still in support of *Save 1*. But then how could *Do Nothing* be wrong, if it is supported by a prerogative that outweighs the biggest reasons around? It would seem that we have to represent the

case like so:

| | Do Nothing | Keep Promise | Save 1 |
|---|---|---|---|
| Reasons | 0 | 2 | 5 |
| Prerogatives | 10 | 10 | 0 |

And this is clearly inadequate, because it implies that *Do Nothing* is permissible. The question, then, is this: how could my prerogative to *Do Nothing* outweigh the reasons behind *Save 1* but not the *weaker* reasons behind *Keep Promise?*

The answer is that prerogatives are *contrastive*.[159] I can have a prerogative to do nothing *rather than make a sacrifice*, and yet no prerogative to do nothing *rather than keep my promise*. In general, there is no single fixed number we can assign to each option to represent the weight of one's prerogative to do it. (Something similar may be true of reasons, too.) This is just what we should expect from a prerogative not to harm oneself; whether I have a prerogative to do nothing rather than $x$ doesn't depend on the absolute harmfulness of doing nothing; it depends on whether the alternative $x$ is *more* harmful. Some options are, some aren't. So that is what explains the special weirdness of Kamm's paradox: *Do Nothing* is defeated by *Keep Promise* but ties with something even better, viz. *Save 1*—which is possible because the prerogative here is contrastive.

Horton's and Kamm's paradoxes turn out to be importantly different when we treat them using reasons and prerogatives. In Horton's case, it's better to do wrong; in Kamm's case, an option ties with something better than its defeater; and Kamm's case alone calls for contrastive prerogatives. In Horton's case, our prerogatives protect only the worst of three options. In Kamm's, they protect the worst option *only from being defeated by the best*, not by the second-best. But there is still

[159] For an introduction to contrastivism about normative notions, see Snedegar 2015.

a deep essential core to the paradoxes, which is that defeat fails to transmit over ties. This is possible because of the fundamental fact we started with: tied options can differ. One option might be better even though the other is backed by a stronger prerogative, or a prerogative that pops out in different contrasts.

Moreover, nothing we have said so far at all puts pressure on us to accept menu-relativity. Prerogatives can be incommensurable with reasons, and can even be contrastive, and yet be expressible in terms of binary defeat relations between pairs of options. A theory with reasons and prerogatives very naturally leads to Transmission failures—and solutions to our three paradoxes—without forcing us to give up binary defeat.

*5.3 Incommensurability in General*

Let me now say a bit about how the argument can be extended to some other accounts of supererogation.

Consider Derek Parfit's (2011: 137–41) view that supererogation is a matter of *parity* between partial and impartial reasons. The idea is that *Save 1* is "impartially" better than *Do Nothing*, which is "partially" better than *Save 1*, but neither option is better overall—nor are they precisely equal in value, like precise duplicates. They are "on a par" (the term is from Chang 2003). On Parfit's view, supererogation is permissible because, although it is better in a way, it is not better all things considered. Partial reasons are like prerogatives to be impartially suboptimal (except that they can also ground obligations, when partiality weighs more). Instead of pure prerogatives weighing against reasons, we have reasons weighing against each other, generating parity.

Parity, interestingly, leads to intransitivities, because the "on a par" relation is *insensitive to mild sweetening* (Hare 2009). Example: tea and coffee are on a par, and so both are permissible in a pairwise choice. Now add a third option—coffee at a $.10 discount. The tea might stay (rationally)

permissible even as the original-price coffee is ruled out. The "sweetened" version of the coffee option beats the original, which is on a par with the tea, and yet the tea and discount coffee are on a par, too, neither beating the other. This is a failure of Transmission Over Ties: *Discount Coffee* > *Coffee* = *Tea* = *Discount Coffee*. It's possible because the tied options are different, which allows for them to be beaten by different things.

For Parfit, the same will hold in the case of supererogation. In Horton's case, *Save 2* is the "sweetened" version of *Save 1*, and it remains on a par with *Do Nothing*. *Save 2* > *Save 1* = *Do Nothing* = *Save 2*. In Kamm's, *Keep Promise* is the sugary sibling of *Do Nothing*, and we have *Keep Promise* > *Do Nothing* = *Save 1* = *Keep Promise*. These intransitivities flow naturally from Parfit's view of supererogation, even though he doesn't have my notion of a prerogative.

Many views of supererogation share their structure with Parfit's, though they might use other flavors of reasons (like Portmore's (2012) "moral" and "non-moral" instead of "impartial" and "partial"), and other relations between them (like Raz's (1975) first- and second-level reasons, or Bader's (forthcoming) non-dominance, instead of Parfit's parity). It is trivial to extend from Parfit's view to these others. What unites them all is a commitment to incommensurable dimensions, and so on any of these views, it's possible for two permissible acts—like *Save 1* and *Do Nothing*—to be morally different. If tied options can differ, we can see how defeat might not transmit over ties. And tied options must differ if there is going to be supererogation, which is by definition somehow better than something it ties.

## 6. Conclusion

I have argued that supererogation arises from an incommensurability between reasons and prerogatives; that this incommensurability leads to the intransitivity of moral justification ("weak defeat"); and that this intransitivity does not itself threaten The Independence of Irrelevant

Alternatives or binariness. Now I'm out of arguments. So let's take a moment to reflect. What can we learn from all the supererogatory paradoxes? Why did we need to think about them?

For decades, moral philosophers have felt some tectonic link between option-sensitivity and supererogation, between permissions and the principles of rational choice. We have been slow to appreciate the varieties of option-sensitivity, because we have not always distinguished menu-relativity from intransitivity. But the distinction here is crucial and intuitive. In cases of menu-relativity, adding an option can change whether one of the original options *itself* makes another original option wrong. In a failure of Transmission Over Ties (a transitivity principle), adding an option has a different effect on two previously equally permissible options, making only one wrong.

How could there be this intransitivity? How could *x* defeat *y* without defeating what *y* ties? The key fact is that tied options don't have to be indistinguishable. Supererogating "ties" the moral minimum, in that either can be justified over the other, but in my view, one has more reason to supererogate, plus a prerogative not to.

The distinction between reasons and prerogatives also unlocks the Paradox of Supererogation. Why may we go against the balance of reasons? Because we have prerogatives, which free us from the demand to do what there is more reason to do.

Since reasons and prerogatives are incommensurable, in the sense that they are separable factors, they also solve our paradoxes of intransitivity. When our prerogatives protect our worst option (*Do Nothing*) but not our second-worst (*Save 1* instead of *Save 2*), it can be better to do wrong. When our prerogatives can protect our worst option against our best (*Save 1*) but not against our second-best (*Keep Promise*), supererogation can override what is otherwise a duty. In both cases, a third option defeats an option but fails to defeat what that option ties with—a Transmission failure. In neither case do we have to relinquish Independence or endorse menu-relative defeat.

There may be further cases of genuine menu-relativity, where a tie between two options is

broken by "irrelevant alternatives." But we are not forced to say this about our cases. Although our supererogatory paradoxes rule out a whole menagerie of properties clustered around transitivity (see Appendix 1), the supererogationist can take comfort amidst this maelstrom of violations in the fact that her options are menu-sensitive in just one way—and only as the predictable consequence of a much-needed dose of incommensurability. Tracing the line from incommensurability to intransitivity is, I hope, some progress. Still, there are plenty of tweaked and weakened kinds of transitivity left to explore—and the battle for Independence is just getting started.[160]

[160] My thanks to Kieran Setiya, Yael Loewenstein, and an anonymous referee at *Noûs* for much-needed comments on a sprawling early draft. Thanks also to Theron Pummer, Jack Spencer, Justin Khoo, Paul McNamara, Mirjam Müller, Kelly Gaus, Seth Lazar, Tomi Frances, Caspar Hare, and Jocelyn Wang.

## Appendix 1: Extending the Paradoxes

To see just how closely Kamm's and Horton's paradoxes are linked—and to push them further—we may combine our four options to generate an even nastier bonus puzzle. Recall the defeat relations from before: *Save 2* > *Save 1* = *Keep Promise* > *Do Nothing* = *Save 2*. We just need to add *Keep Promise* = *Save 2*, which seems reasonable: we aren't required to save two at great injury to ourselves in order to keep our promise, and we aren't required to keep our promise rather than heroically saving two.

Now we have a smorgasbord of Transmission failure, but still no need to give up Independence (Property α). We do, however, have a counterexample to the weak and quaternary cousin of Transmission:

> *Interval Order Property*
> If A > B and C > D, then A > D or C > B.[161]

This is clearly violated by our mixed case, where *Save 2* > *Save 1* and *Keep Promise* > *Do Nothing*, and yet *Do Nothing* = *Save 2* and *Keep Promise* = *Save 1*.

What is interesting about this result is that, if sound, it entails that moral defeat can't be modeled by a simple operation on *intervals*. (A weaker version of Gert's "Range Rule," endorsed in Gert 2004: 505–08, rejected by Chang 2005.) The model goes like this: we assign each option an (closed or open) interval on the real number line, and we say that *x* defeats *y iff x*'s interval is wholly to the left of *y*'s. This picture simply can't handle violations of the Interval Order Property. On the interval model, if D is wholly to the right of A, then A > D; if D is not wholly to the right of A, then C must be at least as far left as A, and so C > B. The combined paradox shows that the intransitivity of defeat can't be measured on a single dimension even if we use intervals instead of points.[162]

---

[161] Sen 2017: 295; notably, this property can be used to generate Arrow-style impossibility theorems.

[162] This entails that moral defeat can't be a *semi-order*. (If it were, it would be representable using unit intervals.) Our simple four-option case violates Luce's (1956: 181) third axiom of semi-orders: if A > B = C > D, then A > D. (For a counterexample to the fourth axiom, see below. The first two axioms are the reflexivity of '=' and completeness of '≥'.)

Let me note three more ways to extend the paradox still further—new varieties of intransitivity without any loss of Independence. First, we can extend "sideways" by adding more pairs of supererogatory options, with different levels of sacrifice and benefit (e.g. adding the opportunities to save 10 or 11 lives at the cost of immense agony); this leads to violations of weaker and weaker transmission principles. In a four-option case, *Save 2* not only fails to defeat what *Save 1* ties (*Keep Promise*); it fails to defeat what is defeated by what *Save 1* ties (*Keep Promise* > *Do Nothing*). In a six-option case, *Save 11* > *Save 10* = *Save 2* > *Save 1* = *Keep Promise* > *Do Nothing*, and yet *Do Nothing* = *Save 11*. If this can go on without limit, then for any *n*, we have counterexamples to:

*Transmission Over* n-*Ties*
If $A_1 > B_1 = A_2 > B_2 = \ldots = A_{n+1} > B_{n+1}$, then $A_1 > B_{n+1}$.

For as large an *n* as we like. Intuitively, the larger the *n*, the greater the sacrifice and benefits.

Second, we can extend "vertically" by using triples, quadruples, etc. instead of pairs. For instance, instead of *Keep Promise* > *Do Nothing*, we could have *Fulfill Blood Oath* > *Keep Promise* > *Do Nothing*, where it's still the case that *Fulfill Blood Oath* = *Save 1*. This will get us violations of:

*Semi-Transitivity*
If A > B > C, then A > D or D > C.[163]

Crudely but intuitively: if A defeats B, and B defeats C, then for any option D, D either defeats C (the biggest loser among the three), or is defeated by A (the biggest victor among the three). This is false in our case because *Fulfill Blood Oath* > *Keep Promise* > *Do Nothing* = *Save 1* = *Fulfill Blood Oath*. Weaker still is the generalization:

*Semi-Transitivity*$_m$
If $A_1 > A_2 > \ldots > A_n$, then $A_1 >$ B or B $> A_n$.

For any *m* we like. (An example might help. Suppose I've got a choice between supererogating or

[163] See Sen 2017: 295, but note that this is not the same as Houthakker's (1950) "semi-transitivity," which is more like acyclicity. Note also that the case violates Luce's fourth axiom of semi-orders, which says: if A > B > C and B = D, then it can't be both that A = D and C = D. But *Fulfill Blood Oath* > *Keep Promise* > *Do Nothing*, and each of these ties *Save 1*.

keeping my promise to relieve my friend from as much of his headache as possible. My options are: {*Full Relief*, *One-Half Relief*, *One-Third Relief*, …, *One* n-*th Relief*, *Do Nothing*, *Save 1*}. It's wrong to relieve my friend's headache any less than fully, unless I supererogate and save the one. Here, *One* n-*th Relief* = *Save 1* = *Full Relief*, despite the chain of defeat from full to 1/*n*.)

We can also extend in both ways at once. Instead of a tuple of tuples of ranked options, <<*Keep Promise*, *Do Nothing*>, <*Save 2*, *Save 1*>>, we have an *m*-tuple with *n* options each, for a total of $m \times n$ options. The bigger the *m*, the longer the chains of defeat (if $m = 3$, e.g., we have <*Keep Blood Oath*, *Keep Promise*, *Do Nothing*>). The bigger the *n*, the more chains we have, each better than the last (if $n = 3$, e.g., we end with the chain featuring *Save 10*).

Third and finally, we can aggravate the paradox by forcing a certain wacky species of *β*-failure, in which the third option is itself defeated. Start with *Do Nothing* and *Save 1*. Now suppose that there is another option, *x*, such that *x* > *Do Nothing* and *Save 1* > *x*. (Perhaps doing *x* is better than nothing, but I have no right to do *x* rather than save people.) This is a *β*-failure: the new third option defeats one of the originals—and then is itself defeated by the other, making this case a special kind of *β*-violation. These special cases violate a weakening of Property *β* that Sen (1971: 315) calls:

> *Property δ*
> For any finite set of options S and subset S*, if A, B ∈ *f*(S*), then {A} ≠ *f*(S).

Where '*f*' denotes a deontic choice function (a function from subsets of a set to the permissible options inside). Put more simply: if two options are permissible in a set of options, then it can't be that one of them is the *only* permissible choice from a superset of those options. This is just what happens when we add *x* to {*Do Nothing*, *Save 1*} in the above example.[164]

[164] Interestingly, Kamm's (1996: 343–44) "Charities Case" involves a failure of Property δ, but as she notes, the case also violates the Independence of Irrelevant Alternatives (so it involves menu-relative defeat). My point here is that δ-failures don't need to violate Independence.

I conclude by noting that, intuitively, there cannot be failures of Property δ in a Horton-style case, where the third option defeats a previously supererogatory option. Property δ can only fail in Kamm-style cases, where the third option defeats something that was previously the moral minimum. The intuition here is that we are always allowed to do our best option. Since supererogation is optimal, we can't add an even better option without its being optimal. And so that option, too, will be permissible to choose from the freshly expanded set. Since these results are guaranteed by my Prerogatives Principle, which requires that wrong acts be worse than others (by a margin uncovered by prerogatives), that is a mark in favor of that principle.

## Appendix 2: Other Solutions to Kamm's Paradox

*Archer's Solution*

With incommensurable reasons and prerogatives, we have a powerful solution to the paradoxes. But could we also solve them using other kinds of incommensurability?[165] We'll look at two more kinds.

First, let's consider Alfred Archer's solution to Kamm's paradox, which uses only incommensurable reasons. The idea goes like this. Reasons can play multiple roles. They can *require*, *justify*, or *favor*.[166] The more favored an act, the more choiceworthy it is; and an act is permissible just if the justifying reasons to do it can meet or outweigh the requiring reasons to do otherwise. Moreover, these are three incommensurable dimensions; a reason can favor strongly without justifying strongly, justify without requiring, require urgently without so strongly favoring—etc. (Justifying reasons are basically just prerogatives, except that prerogatives don't count in favor of anything—whereas some writers see "counting in favor" as the essential mark of all normative reasons; see e.g. Parfit 2011, Chang 2014: 485.)

With these three degrees of freedom, Archer captures the intuitions in Kamm's case by assigning weights like the following:

| | Do Nothing | Keep Promise | Save 1 |
|---|---|---|---|
| Requiring | 0 | 8 | 5 |
| Justifying | 5 | 8 | 10 |
| Favoring | 0 | 8 | 10 |

[165] I focus on Kamm's paradox because we have already considered views on Horton's (on which, see Chapter 4 (BDW)). I won't discuss Dorsey's (2013) solution, which implies that supererogation is morally required (see Archer 2016: 452–55). Portmore's (2017) solution, which relies on his "maximalism" about reasons, has some possible problems—it assumes transitivity (his "C4") and does not apply in one-off decisions or choices of life-plan—but it is also very interesting.

[166] For related ideas, see Horgan and Timmons 2010 on the "merit-conferring role" of reasons (and Little and McNamara 2017 on "commendatory" reasons), Gert 2007 on the "justifying" and "requiring" roles, and see also Portmore 2012 on the "moral-justifying" and "moral-requiring" roles.

The key facts here: it's permissible either to keep the promise or break it to supererogate; it's better to keep the promise than do nothing; and it's better still to supererogate and save the life.[167]

But Archer's solution has a problem common to complex theories, anarchic markets, and spoiled children: too much freedom![168] Archer's three undetached dimensions lead to three puzzles. First, if "favoring" comes apart from "requiring" and "justifying," we are left wondering *why* there are so many tight conceptual links between the right and the good. It isn't an open question whether required acts are more favored than wrong ones. But it seems it should be, on Archer's view, since he doesn't put a priori constraints on how these dimensions of strength co-vary. For all we're told, an act could be required without being at all favored—or most favored without being justified. Second, by divorcing favoring from the other roles, Archer raises the question of *how* a consideration could purely favor. How exactly do favorers manage to favor, if not by tending to require or justify? What else could favorers be up to? Third and finally, Archer's view will have serious trouble giving a story for how reasons are related to 'should' and 'must' claims. Does 'should' mean "most reason of a certain kind?" Is it just the favorers that count—or just the requiring reasons? Can Archer explain why 'best' entails 'may', and 'must' entails 'ought', but not vice versa?[169] Taken together, these puzzles suggest that we should try to minimize degrees of freedom, and use fewer dimensions if we can.

So I think my solution, which uses reasons and prerogatives, is less problematic—and plain simpler—than a solution with three different roles for reasons. On my view, reasons ground requirements by making actions better, and prerogatives free us from the default requirement to

---

[167] My table is adapted from Archer's treatment of Dorsey's case; see Archer 2016: 459.

[168] Portmore (2017: 292–93, fn. 11) has an objection along these lines: the (moral-)justifying and (moral-)requiring strength of a reason should not come apart.

[169] On the problems for recovering 'ought' and 'must' from flavors of reasons, see Snedegar 2016. A theory of reasons and prerogatives has an easier time with 'ought': we ought to do what we have most reason to do. (So long as we are talking about maximally specific options.) I say more about this in an unpublished draft, called "Reasons and Prerogatives."

make the better choice. We should do an act just if there is most reason to do it. So, we should always do what we have to do, but we have to do as we should only when we lack the prerogative to do otherwise.

*Kamm's Solution*

Finally, let me consider Frances Kamm's own solution to her paradox, in part to see how my view respects some of her insights while updating her formalism. Her paper—originally published as Kamm 1985, reprinted in Kamm 1996—is brilliant, groundbreaking, and extraordinarily rich, but also challenging. Here I can only touch on a few key aspects.

Kamm's basic idea is that acts can compare differently along several different "standards," so that an obligation (*Keep Promise*) can be tied with a supererogatory act (*Save 1*) but defeat something (*Do Nothing*) that the supererogatory act cannot. Kamm therefore believes in incommensurable moral dimensions, in my sense, and so my treatment of the paradoxes owes a great deal to her work, though I think it calls for some tinkering and elaboration.[170]

We should start with a few interpretive points. First, instead of "weak defeat," Kamm talks about what "may permissibly take precedence" over what, and instead of the "defeat" relation, Kamm talks about "dominance" (see e.g. Kamm 1996: 338). This might confuse some readers. The concept of dominance is usually reserved for something like the notion of being purely better than—somehow better, in no way worse. In rational choice theory, for example, one option is said to dominate another just if the first is better in some circumstances and at least as good in all others (Tversky and Kahneman 1986: S253). But as Kamm intends to use "dominance," it ought to refer to

[170] Kamm herself denies that her solution uses incommensurability. That is because she is using the term to mean what I would call *incomparability between options*, rather than incommensurability between dimensions. Kamm is denying that, to give her solution, she needs weak defeat to be incomplete. See Kamm 1996: fn. 32.

a relation between one option that makes the other impermissible; the "dominant" option in this sense can be worse in some respects. *Do Nothing* might be more pleasant than *Keep Promise*, but promise-keeping defeats loafing around.[171]

Second, it is not obvious what Kamm takes weak defeat to be defined over. In our formalism, following the precedent of rational choice theory, defeats and ties are always between options: things that the agent could do. (We've focused on mutually incompatible, maximally relevantly specific options—ignoring disjunctions like "saving one or two kids.") But Kamm also says that, e.g., keeping a promise "takes precedence over" avoiding effort. But avoiding effort is not an option; the effort required to do something is more naturally seen as a property of an option, just like the level of sacrifice that an option requires, which Kamm treats as a property of the supererogatory choice to save one.

Finally, Kamm presents two versions of the "intransitivity" rather than the one I have focused on. One version compares keeping a promise to the (costly) saving of a life to the avoidance of effort. The other compares supererogation to duty to personal interests. Both of these specific versions raise thorny problems that I don't see as essential to the paradox or intransitivity. The first version only works if levels of effort are morally relevant to permissibility; my view is that in general they are not. When we ward off blame by saying that better options were more effortful, usually we are just offering an excuse, not a justification (see Yetter Chappell 2017 on willpower satisficing). The second version presupposes that prerogatives are peculiarly linked to personal interests, a view pioneered by Samuel Scheffler (1982). But this view faces grave objections. It can't explain why we

[171] Kamm (1996: 339) also gives a very abstract example where "dominance" appears to be intransitive; this would be a failure of the transitivity of defeat, also known as *quasi-transitivity*—which is more than enough to rule out the definition of a choice function. (Quasi-transitivity is even more demanding than acyclicity; see Sen 2017: Chapter 1*.) I think it is implausible that every option can be defeated (in a finite choice), because I think it's always permissible to do what's best, and I don't think there can fail to be a (tie for) best option in a finite set. But see Temkin 2012 for a defense of the intransitivity of "all things considered better than."

have prerogatives to do things that aren't in our interests, such as forgoing a benefit for the slightly *lesser* benefit of someone else (Stocker 1976, Slote 1984, Hurka and Shubert 2012), declining to give personally rewarding gifts (Ferry 2013), or doing costless favors (Horgan and Timmons 2010; see also Chapter 6 (SRC)). The view is also messy when combined with a view of rights, because it needs exceptions so that we don't have prerogatives to harm others willy-nilly (Hurka and Shubert 2012: 9). So I think we should avoid essentially linking the paradox's prerogatives to efforts and interests, staying open-minded about the nature of prerogatives, and focusing instead on the structural properties of defeat.

Here is the bottom line. I agree, in broad outline, with Kamm's solution. Weak defeat is intransitive because defeat can fail to transmit over ties, and Transmission can fail because one tied option might have relevantly different properties than the other. But unlike Kamm, I don't have to take a stand on the special weights of efforts and personal goals. I have a simple, neutral account of how acts get to be permissible: they weakly defeat everything on the menu, by virtue of the reasons and prerogatives in favor weighing at least as much as the reasons against. This account doesn't commit us to any view of the nature of prerogatives or of reasons. The account can handle Kamm- and Horton-style cases without any need to give up Independence, which allows us to explain permissibility using binary weak defeat. And the account's key notions—those of reasons and prerogatives—are just the basic elements of a solution to the Paradox of Supererogation.

**REFERENCES**


Archer, Alfred (2016). "Moral Obligation, Self-Interest, and the Transitivity Problem," in *Utilitas*, 28 (4): 441–464.

-------(2017). "Supererogation," in *Philosophy Compass* 13 (3).

Arrow, Kenneth (1951) [1963]. *Social Choice and Individual Values*. New York: Wiley. 2nd Edition 1963.

Bader, Ralf (forthcoming). "Agent-Relative Prerogatives and Suboptimal Beneficence," in *Oxford Studies in Normative Ethics*.

Chang, Ruth (2002). "The Possibility of Parity," in *Ethics* 112 (4): 659–688.

-------(2005). "Parity, Interval Value, and Choice," in *Ethics* 114: 331–350.

-------(2014). "Practical Reasons: The Problem of Gridlock," in B. Dainton and H. Robinson (eds.), *The Bloomsbury Companion to Analytic Philosophy*. Continuum Publishing Corporation: 474–499.

Darwall, Stephen (2013). "'But It Would Be Wrong'," in *Morality, Authority, and Law: Essays in Second-Personal Ethics I*. Oxford: Oxford University Press: 52–71.

Dietrich, Franz and List, Christian (2017). "What Matters and How it Matters: A Choice-Theoretic Interpretation of Moral Theories," in *The Philosophical Review* 126 (4): 421–479.

Dorsey, Dale (2013). "The Supererogatory, and How to Accommodate It," in *Utilitas* 25 (3): 355–382.

Ferry, Michael (2013). "Does Morality Demand Our Very Best? Moral Prescriptions and the Line of Duty," in *Philosophical Studies* 165 (2): 573–589.

Gert, Joshua (2004). "Value and Parity," in *Ethics* 114 (3): 492–510.

-------(2007). "Normative Strength and the Balance of Reasons," in *Philosophical Review* 116 (4): 533–562.

Heyd, David (1982). *Supererogation*. Cambridge: Cambridge University Press.

-------(2016). "Supererogation," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring

2016 Edition), URL =
<https://plato.stanford.edu/archives/spr2016/entries/supererogation/>.

Horgan, Terence and Timmons, Mark (2010). "Untying a Knot from the Inside Out: Reflections on the 'Paradox' of Supererogation," in *Social Philosophy and Policy* 27(3): 29–63.

Horton, Joe (2017). "The All or Nothing Problem," in *The Journal of Philosophy* 114 (2): 94–104.

Houthakker, H.S. (1950). "Revealed Preference and the Utility Function," in *Economica* 17 (66): 159–74.

Hurka, Thomas and Shubert, Esther (2012). "Permissions to Do Less Than Best: A Moving Band," in *Oxford Studies in Normative Ethics Volume 2*: 1–27.

Kamm, Frances (1985). "Supererogation and Obligation," in *The Journal of Philosophy* 82 (3): 118–138.

-------(1996). *Morality, Mortality, Volume II: Rights, Duties, and Status*. New York: Oxford University Press.

Lazar, Seth and Barry, Christian (ms.). "Acting Beyond the Call of Duty: Supererogation and Optimization."

Little, Margaret and McNamara, Colleen (2017). "For Better or Worse: Commendatory Reasons and Latitude," in Mark Timmons (ed.), *Oxford Studies in Normative Ethics, Vol 7*: 138–160.

Luce, R. Duncan (1956). "Semiorders and a Theory of Utility Discrimination," in *Econometrica* 24: 158–171.

Massoud, Amy (2016). "Moral Worth and Supererogation," in *Ethics* 126 (3): 690–710.

McMahan, Jeff (2018). "Doing Good and Doing the Best," in *The Ethics of Giving: Philosophers' Perspectives on Philanthropy*, Paul Woodruff (Ed.). New York: Oxford University Press: 78–102.

Morreau, Michael (2014). "Arrow's Theorem," in Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). URL =
<https://plato.stanford.edu/archives/win2016/entries/arrows-theorem/>.

Muñoz, Daniel (ms.). "Reasons and Prerogatives."

Parfit, Derek (1982). "Future Generations: Further Problems," in *Philosophy and Public Affairs* 11 (2): 113–172.

-------(2011). *On What Matters, Volume One*. Oxford: Oxford University Press.

Portmore, Douglas (2012). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. New York: Oxford University Press.

-------(2017). "Transitivity, Moral Latitude, and Supererogation," in *Utilitas* 29 (3): 286–298.

Pummer, Theron (2016). "Whether and Where to Give," in *Philosophy and Public Affairs* 44 (1): 77–95.

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

Raz, Joseph (1975). "Permissions and Supererogation," in *The Philosophical Quarterly* 12 (2): 161–168.

Scheffler, Samuel (1982). *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Oxford University Press.

Sen, Amartya (1969). "Quasi-transitivity, Rational Choice and Collective Decisions," in *Review of Economic Studies* 36: 381–393.

-------(1971). "Choice Functions and Revealed Preference," in *The Review of Economic Studies* 38 (3): 307–317.

-------(1993). "Internal Consistency of Choice," in *Econometrica* 61 (3): 495–521.

-------(2017). *Collective Welfare and Social Choice: Expanded Edition*. Cambridge: Harvard University Press.

Slote, Michael (1984). "Morality and Self-Other Asymmetry," in *The Journal of Philosophy* 81 (4): 179–192.

Snedegar, Justin (2015). "Contrastivism About Reasons and Ought," in *Philosophy Compass* 10 (6): 379–388.

-------(2016). "Reasons, Oughts, and Requirements," in R. Shafer-Landau (Ed.) *Oxford Studies in*

*Metaethics*, XI. Oxford: Oxford University Press: 183–211.

Stocker, Michael (1976). "Agent and Other: Against Ethical Universalism," in *Australasian Journal of Philosophy* 54 (3): 206–220.

Tadros, Victor (2011). *The Ends of Harm: The Moral Foundations of Criminal Law*. Oxford: Oxford University Press.

Temkin, Larry (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Tungodden, Bertil and Vallentyne, Peter (2005). "On the Possibility of Paretian Egalitarianism," in *The Journal of Philosophy* 102 (3): 126–154.

Tversky, Amos and Kahneman, Daniel (1986). "Rational Choice and the Framing of Decisions," in *The Journal of Business* 59 (4): S251–S278.

Yetter Chappell, Richard (2017). "Willpower Satisficing," to appear in *Noûs*.

# AFTERWORD

## Freedom, Rights, and Equality

## Afterword:

## Freedom, Rights, and Equality

All right, that was a lot to unload on you. The least you deserve is a recap.

Our guiding idea is the Self-Other Symmetry: we have the same rights against ourselves as against any relevantly similar other. The first worry was that the very idea of owing yourself something is "paradoxical," since you could release yourself at will. But there isn't any logical problem with self-release. The only real truth in the neighborhood is that, if you could waive a right at will, then it wouldn't restrain you in the sense of giving you a *reason* to comply with it. But I don't see this as an objection. On the contrary, it's a crucial insight for developing the Symmetric theory. It is precisely *because* rights against oneself aren't reasons that they don't make supererogation wrong. (I have a right against myself that I not take out my kidney, just as I have against you, but that right doesn't make donation immoral.)

The second threat to Symmetry was from humdrum counterexamples: surely we wrong strangers when we mildly harm them for no reason, but we don't wrong ourselves when we foolishly do a little harm to ourselves. My response was that the self/other difference isn't the real difference-maker here. We wrong the stranger because they do not consent to being harmed. We don't wrong ourselves because we *do* consent to our own actions.

The third worry came from splashier counterexamples: irrational suicide, accidental self-harm, drunken masturbation—more cases where doing something to oneself seems innocent whereas doing it to others would violate a right. Here I had two responses. First, we have to make sure that, when comparing acts done to oneself vs. those done to another, we are really dealing with parallel cases. We have to refine the pair to remove any relevant differences *besides* the self/other difference and bring to light any suppressed factors that might be biasing our intuitions. Once we do so, we may be able to revise our hunches in line with the Self-Other Symmetry. "Refine and revise."

The second point I made was that these cases have a further, crucial aspect that is elegantly explained by Symmetry: sometimes, others may intervene to stop us from doing harm by force. When? Symmetry predicts that enforcement will be more justifiable when people are acting without the valid consent of those they act on. In the case of acts done to oneself, that means: accidents, ignorance, incompetence, and inalienability—general conditions when consent doesn't waive rights. It is a striking fact that this prediction seems to line up with our intuitions, including classic cases (like Mill's agent who unwittingly risks self-harm) devised without the Self-Other Symmetry in mind. Other things equal, when someone self-harms without valid self-authorization, others may enforce that person's rights against himself or herself.

The result is a theory of self-regarding morality that flows from a more general theory of when people wrong each other. Wronging oneself can be understood entirely in terms of the principles of wronging plus the psychological facts that follow from your being identical to you. (The big fact: you consent to your own choices.) Justified paternalism can be understood entirely in terms of those same facts plus the principles that govern the enforcement of rights. Any theory that builds everything on the specialness of the "self" rather than the qualitative and relational factors that matter for interpersonal rights—consent, ignorance, harm—is missing the point, and needlessly mystifying the foundations of morality.

From here, we turn to another source of moral mysteries: supererogation. Why are we permitted to do less than best? I first argued (in Chapter Three (FRP)) for a conditional: *if* rights are Self-Other Symmetric, *then* rights entail prerogatives. Waivable rights against oneself would be "finkish," waived by the choice to do that which they forbid, and so they would not be reasons. But they would still be powerful for the purposes of moral defense. I can't reason like this: "Well, it's my kidney, so I had better not take it and give it away for the greater good." But if someone demands that I donate, I can say: "It's *mine*." If they demand to know why I should keep it, I can say that I

don't *owe* them a reason. My right carries weight in defense and can outweigh a considerable improvement in global happiness—just like the right of an unwilling other, except that others' rights are also reasons for me to comply.

If prerogatives come from rights, moreover, we can understand why they protect the actions that they do. If I own a pill, I may use it to save myself rather than a needier other (agent-favoring). I may also give the pill to another *even if they need it less than me* (agent-sacrifice). I may give it to relative strangers, even if other strangers need it more (other-favoring). And it can be merely optional to give it to a needy other even if I would overall benefit from the gift (self-benefitting supererogation) or break even (costless supererogation). The standard account of prerogatives, on which they are essentially a matter of protecting and neglecting the agent's interests, can't give a unified or principled account of these cases—least of all costless supererogation. The Self-Other Symmetric view handles the cases with ease. You may use the pill as you like because it is *yours*. (And others may not take it, even to help themselves, because the pill is *not theirs*.) The real source of prerogatives appears to be rights, not self-interest.

With a theory of prerogatives in place, we can take on various new puzzles of supererogation, using the fundamental insight that supererogation emerges from the interplay of reasons and prerogatives. (My main arguments here can be run without rights-based prerogatives—indeed, without prerogatives at all—so long as there are two analogous dimensions that factor into right and wrong.)

The first puzzle is the All or Nothing Problem: you choose between the moral minimum (*Do Nothing*), a supererogatory feat of heroism (*Save 2*), and a gratuitously worse version of being the hero (*Save 1*). The gratuitously bad act seems wrong. But it also seems better than doing nothing. How could a wrong act be better than a permissible one? Answer: we might have a prerogative to do our worst option, but no prerogative to do our second-worst. The key is that justifiability is sensitive

to two incommensurable dimensions ("more reason" doesn't entail "more prerogative")—and the background idea is that permissible acts *just are* those that can be justified over all alternatives.

Second, Kamm's Intransitivity Paradox: you choose between supererogating (*Save 1*) and either of two non-heroic options: doing a duty (*Keep Promise*) or simply doing nothing (*Do Nothing*). Doing nothing is justifiable over being the hero, given the cost, and being the hero is justifiable over fulfilling the humdrum duty—but doing nothing *isn't* justifiable over fulfilling the duty. How could that be? Answer: this is really the All or Nothing Problem seen from a different angle. (*Save 1* is justifiable over *Do Nothing*, which is justifiable over *Save 2*….) Again, the key is going to be having two incommensurable dimensions.

But there is, coincidentally, a deep and separate moral to be drawn from Kamm's case. The moral is that it's not enough to think of prerogatives and reasons for doing options as two static numbers we slap onto them. The strength of one's prerogative to do an option depends on *what the alternative is*. In Kamm's example, I have a powerful prerogative to do nothing rather than sacrifice myself and be the hero, but that prerogative doesn't justify doing nothing *rather than keeping my promise to meet you for lunch*. The prerogative here is contrastive, in the sense that it only justifies an option over some alternatives and not others. This is pretty intuitive. In justifying one option over another, what matters are the differences between them (like *relative* costliness), not how they compare to yet other options on the menu (like the very costly sacrifice).

Justification, moreover, is still a mercifully simple and formally tractable affair. Even though prerogatives are contrastive, and justification is intransitive, we can still get a fixed ranking of acts that determines what's right and wrong. The ranking is built like this: look at each pair of options, compare their relevant features (the grounds of reasons and prerogatives, even the constrastive ones), and say that one option outranks the other *iff* it makes it unjustifiable. The permissible options are those justifiable over all others, those never outranked. (Assuming that the "justifiability" relation

is complete.) The wrong acts are those outranked by something or other. The ranking might seem unusual, given that wacky intransitivities and constrastive effects crop up in three-option cases like Kamm's and Horton's, but these cases do not require us to posit that the justifiability relation between a pair in any way wobbles or reverses depending on the rest of the menu. Even as options are tacked on and stripped away, the ranking never has to reshuffle the constant options. Incommensurability leads to some wild things—but not that!

And now we are really at the end. To be honest, there are a few things I've been waiting to get off my chest. First of all: 'supererogation'. It's a terrible word. Half technical term, half intuitive notion. I put no stock whatsoever in my intuitions about what is "supererogatory" *except* what is contained in my more ordinary intuitions about which acts are optional, good, and so on. So let me be clear. There are some cases that my view counts as supererogatory that might not strike you as supererogatory. I have in mind pure prudence, like morally optionally brushing one's teeth. "Why call this supererogatory," you might ask, "since we wouldn't praise someone for doing it? Any shmuck can be selfish!" I don't have an answer. There is no philosophical reason why we should use 'supererogatory' in my way rather than any other. I have chosen to focus on one aspect of the traditional supererogatory actions—their being optional-yet-superior. Others focus on praise and motivation. The only conflict between these approaches, so far as I can see, has to do with verbal emphasis. If I have committed myself in the above papers to any view about proper motives, it was by accident.

Second, I want to admit that there are some genuine philosophical challenges for the rights-based view of supererogation. Here I'll mention just two sorts of cases. The first involves competition over unowned things, like natural resources or a forsaken $20. Even if I don't own the $20, should it really be wrong for me to deprive the needy of it by acquiring it? (Their greater need outweighs my lesser need, and I can't lean on my rights to the $20—I don't own it yet!) Another

case is forgiveness—seems supererogatory, but it is hard to see forgiving as the waiving of a right against oneself. In both instances, we wonder: if the prerogative is a right against oneself, what is its interpersonal flipside? Do I have a right against others that I acquire the $20? That I forgive (or not forgive) the person who wronged me?

My view is that these are difficult cases not because of the Self-Other Symmetry, but because acquisition and forgiveness are themselves tough issues. (I will note that other views of supererogation have as much, if not more, trouble explaining the optionality of forgiving.) Maybe they are stubbornly Self-Other Asymmetric in a way that other cases aren't. There is more work to do here.

And that, itself, is the point I want to close with. These papers, though intended as a unified argument, do not comprehensively defend a comprehensive view. I have not said where rights come from; which rights we have;[172] how moral defense relates to deliberation (whether joint or individual); whether Symmetry holds of all possible cases or just the flashy classics; how moral obligation relates to 'should' and advice; or how supererogation in my sense relates to praise and moral worth. I have had to leave out most of my arguments for using reasons and prerogatives rather than moral and non-moral reasons. I have not even tried to count up all the hard cases still looming.

What I have tried to do is to put forward a vision of morality on which we are all equals with a certain degree of personal freedom. This isn't the equality of utilitarianism, where any one of us may be shackled or stripped for parts whenever it's for the collective good. Nor is this the egoistic freedom of Scheffler and Nagel, based in the idea that there is something morally respectable about

[172] Another worm-can: I have been assuming a commonsense stock of rights over property and body parts. But what if we discover, after doing some history and political philosophy, that we don't really have rights over, say, our land? (Perhaps our ancestors stole it.) Should we still believe in the prerogative to use it for our own gain? My answer is *no*. If something isn't rightfully ours, we don't in general have any prerogative to use it selfishly (*contra* the Cost View).

seeing one*self* as *eo ipso* more important than others, no matter how similar we are in the ways that matter. My Self-Other Symmetric view is that, by default, everyone's interests matter equally *and* everyone has the same rights against everyone. When it comes to interpersonal rights, the mechanism of consent allows us some flexibility in determining whether those around us will be subject to moral sanction. Our bodies and things are presumptively—but not irrevocably—off-limits. But this entails something peculiar and serendipitous in the reflexive case. A right against oneself, thanks to the consent mechanism, still allows one to make sacrifices for the greater good. One may also leave the right in play as protection from the moral community's sanctions. If one's body remains "off-limits," refusing to give up one's spare kidney is perfectly justifiable even though others need it more; the choice here is in a respect like the choice not to steal a kidney from an unwilling other. But there is a difference. Our own bodies do not belong to unwilling others, like temples to a demanding deity. Our bodies are fully *ours*, and that is why it is *our choice* what to do with them, at least to a great extent, even when the collective good is on the line. Other people are our equals—but some choices belong to no one else.